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

April 2020

Juanita González-Uribe
London School of Economics

Santiago Reyes
Inter-American Development Bank

Abstract

Why is high-growth entrepreneurship scarce in developing countries? Does this scarcity reflect firm capabilities constraints? We explore these questions using as a laboratory an accelerator in Colombia that selects participants using scores from randomly assigned judges and offers them training, advice, and visibility but no cash. Exploiting exogenous differences in judges' scoring generosity, we show that alleviating constraints to firm capabilities unlocks innovative entrepreneurs' potential but does not transform subpar ideas into high-growth firms. The results demonstrate that some high-potential entrepreneurs in developing economies face firm capabilities constraints and accelerators can help identify these entrepreneurs and boost their growth.

JEL Classification: G24, L26, M13
Keywords: High-Growth Entrepreneurship, Business Accelerators, Young Firms, Firm Capabilities

---

[1]Corresponding authors: Juanita González-Uribe: j.gonzalez-uribe@lse.ac.uk, Houghton Street, WC2A 2AE, London; and Santiago Reyes Ortega: sreyes@iadb.org, 1300 New York Ave NW, Washington DC. We thank Ulf Axelson, Jean Noel Barrot (discussant), Shai Bernstein, Vicente Cunat, Paul Gertler, Dirk Jenter, Borja Larrain, Igor Makarov, David McKenzie, Daniel Paravisini, Julio Riutort, Antoinette Schoar, Betty Simkins (discussant), Consuelo Silva (discussant), Christopher Woodruff, and seminar participants at ESSEC, FMA, IADB, LSE, Nova, Oxford, Universidad de los Andes, International Conference Finance PUC Chile, HEC Workshop on Entrepreneurship and Economic Development, and NBER Entrepreneurship 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.

1. Introduction

A large literature in economics aims to understand the fundamental causes of the growth differences between developed and developing countries. Most of the focus thus far has been on cross-country disparities in small business growth.[2] However, recent work suggests that the development problem can be traced to cross-country differences in "high-growth young firms," also informally known as "gazelles."[3] For example, Eslava, Haltiwanger, and Pinzon (2018) show that a scarcity of Colombian gazelles drives the average life cycle growth gap between the United States and Colombia. This new evidence raises a number of questions regarding high-growth young firms, a subpopulation of businesses hitherto understudied by the academic literature. For instance, why are high-growth young firms scarce in developing countries? Does this scarcity reflect an optimal allocation of the most talented individuals outside of entrepreneurship due to social norms and incentives (e.g., Lerner and Schoar, 2010)? Or are there constrained entrepreneurs in emerging economies who are unable to grow high-potential ideas? If so, what type of constraints are faced by these entrepreneurs? And how can constrained high-potential entrepreneurs be identified in the population of businesses?

In this paper, we take an initial step in exploring these questions using "ValleE," an ecosystem business accelerator in Colombia, as a laboratory.[4] Business accelerators are ideal research laboratories because their aim is to identify and boost high-potential entrepreneurs by alleviating their constraints to growth. It follows that the success of accelerators in achieving these goals could contribute to solving the development problem. Focusing on ValleE is particularly useful because it allows us to address several empirical challenges that are common in exploring these questions, including distinguishing different types of constraints to high growth, identifying high-potential entrepreneurs, and measuring young firm performance.

We distinguish constraints to "firm capabilities," as ValleE provided participants with no cash but instead offered standardized business training, customized business advice, and visibility. By firm capabilities we mean key elements of the growth process that firms cannot readily buy in the market and that impede growth even if ventures are injected with cash, such as identifying the correct market

[2] Empirical studies have examined financial and firm capabilities constraints. The impact of microcredit results shows modest, but not transformative, results (for a summary of this literature, see Banerjee, 2013). Evidence using cash grant experiments shows large average effects driven by a minority of businesses (see de Mel, McKenzie, and Woodruff, 2008; Karlan, Knight, and Udry, 2012). The impact of business training programs is more mixed (for a summary of this literature, see McKenzie and Woodruff, 2014).

[3] The disproportionate contribution of high-growth young firms to economic growth both in developed and developing economies is a well-established fact in economics (see Birch and Medoff, 1994; Henrekson and Johansson, 2008; Schoar, 2010; Haltiwanger, Jarmin, and Miranda, 2013; Ayyagari et al., 2014; Hsieh and Klenow, 2014; Grover, Mesdvedev, and Olafsen, 2019; Eslava, Haltiwanger, and Pinzon, 2018).

[4] 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).

need or gaining market recognition (cf., Sutton, 2012). Little is known about the importance of firm capabilities constraints for high growth, as most interventions focus instead on easing potential financial constrains by providing entrepreneurs with seed capital (see Clarysse, Wright, and Van Hove, 2015).

Our empirical strategy exploits the fact that participants were selected based on scores from three randomly assigned judges who independently evaluated their business plans. While the accelerator provided uniform criteria by which a judge should score applications, we show there is substantial variation in the interpretation of these criteria across judges.[5] As a result, otherwise identical applicants differ in their acceleration probability because they were randomly assigned to judges with different "scoring generosities." Our approach is similar to that found in the "judge leniency literature" (e.g., Doyle, 2007 and the literature thereafter). The main departure is that we exploit the multiple judge assignment per applicant to control for unobserved applicant heterogeneity as perceived by the judges. We show that controlling for this heterogeneity is crucial in our setting given the skewness in young firm potential and the ability of ValleE judges to identify high growers (which we show is orthogonal to scoring generosity). To track performance, we use novel administrative data from the Colombian business registry two years before application and three years after, together with annual surveys.

We find compelling evidence that participation in the accelerator has large positive impacts. Using an instrumental variables (IV) approach, we estimate that over the first three post-application years, participation in the accelerator increases annual revenue by $66.3M COP ($20K USD), which corresponds to a 166% (130%) increase from the rejected applicant (average applicant) revenue.[6]

The IV results are representative for the group of applicants whose selection decision is altered by the judge assignment. This group of applicants includes both type one and type two selection mistakes by the program. By type one selection mistakes, we mean the high-potential applicants who are rejected because of the strictness of their randomly assigned judges but who would have been accepted if the selection process controlled for judge heterogeneity. By type two selection mistakes, we mean the low-potential applicants who are accepted because of the generosity of their judges and who would not have been accepted in the accelerator otherwise.

Examining the heterogeneity of impacts by exploiting the continuity of our instrument, we show that the IV results are driven by type one selection mistakes. We find no apparent performance improvements for type two participants relative to similar projects that were correctly rejected from the program. Instead, we find a remarkable growth of correctly accepted high-potential participants relative to similar type one rejected applicants who were mistakenly rejected by the program.

[5] This variation in interpretation can reflect variation in the subjective meaning of scores; for example, in a scale from zero to one, a mediocre proposal might score 0.7 from "generous" judge A but only 0.4 from "strict" judge B.

[6] There is also evidence that acceleration has large positive impacts on profits and employment.

Inspecting the mechanisms of impact, we find evidence that customized advice and visibility are more impactful aspects of the intervention than the standardized business training. Consistent with entrepreneurs' perceptions of business accelerator programs worldwide (see Roberts et al., 2016, 74% of surveyed ValleE participants said that advice and visibility added the most value, whereas only 8% described the standardized business training as a key impact driver. Furthermore, estimated effects are stronger among entrepreneurs who at the time of application reported needing strategic advice rather than standardized training as well as among applicants who could take the most advantage of strategic advice and visibility, such as entrepreneurs with already existing businesses (rather than business ideas) at application. This evidence substantiates the documented outperformance of customized interventions over standardized ones in the context of microenterprises.[7] As one limitation, we note that acceleration can also trigger additional impact mechanisms such as capabilities-enhancing network, certification, and internal validation effects, which, as is common in the literature on capabilities-building interventions, we cannot rigorously distinguish with our current data (e.g., Chevalier et al., 2004).

Overall, our findings provide an initial step in understanding why high-growth entrepreneurship is missing in developing countries. The results provide compelling evidence that firm capabilities' constraints can hinder the ability of high-potential entrepreneurs in developing economies to reach their full "gazelle potential." But on the other hand, the results also highlight the limits of capabilities-building interventions as policy instruments. In particular, they suggest that accelerators can help solve the development problem by identifying and boosting top growers but not necessarily by transforming subpar ideas into high-growth firms. A back-of-the-envelope calculation based on our findings shows that ValleE accelerated high-potential participants' growth beyond Colombian top growers. These participants reach about three times their initial revenue by the fourth year of the business, on average, which roughly doubles the 90th percentile of four-year life cycle revenue growth of Colombian businesses (cf., Eslava and Haltiwanger, 2018).

As one caveat, we note that, as with all firm capabilities' interventions that have been conducted so far, our empirical exercise is a joint test of two closely related hypotheses: first, that firm capabilities are first-order constraints to high growth and second, that firm capabilities can be conveyed via the intervention in the first place. Therefore, failure to find effects for low-potential ideas does not necessarily prove firm capabilities do not matter for the growth of these projects or that firm capabilities cannot be taught to low-potential entrepreneurs. Instead, another simple explanation could be that the firm capabilities provided by the business accelerator were not enough for these types of projects and more intensive interventions are needed.

[7] Several papers in the business training literature show the importance of accompanying standardized business training with more personalized components such as mentoring and technical assistance; see Bloom et al. (2013); Valdivia (2015); Brooks, Donovan, and Johnson (2018); Bruhn, Karlan, and Schoar (2018); Campos et al. (2018); Ubfal et al.(2019). For reviews, see also McKenzie and Woodruff (2014) and Quinn and Woodruff (2019).

This paper is related to two main bodies of literature. The first addresses the sources of and constraints to high-growth entrepreneurship. An increasing body of work shows that a small number of young firms (variously termed "high-growth young firms," "high-growth entrepreneurs," "gazelles," "transformational entrepreneurs," or "super-start firms," among other monikers) make disproportionate contributions to economic growth. At the same time, a long-established literature suggests that a perceived lack of sufficiently impactful high-growth entrepreneurs in developing economies could be in part due to market failures. For example, limited access to firm capabilities (also called "entrepreneurial capital" or "managerial capital," among other terms) can prevent individuals with high-potential ideas from successfully growing their businesses (e.g., Bruhn, Karlan, and Schoar, 2010). Our findings bring new insight to this literature by showing that a deficiency of high-potential participants and/or of more customized intervention services can help explain the historically lackluster performance of firm capabilities-building programs (cf., de Mel, McKenzie, and Woodruff, 2014). The closest paper to ours is McKenzie (2017), which shows no impact of standardized business training on venture growth but does find large effects from the provision of bundled training and cash in the context of a high-growth business plan competition in Nigeria. Buttressed by the heterogeneous impact patterns and the survey evidence, a compelling explanation for the comparative success of ValleE is its provision of intensive customized advice, which is not commonly included in business plan competitions given their short-term and more "at arm's length" nature.

Our work is also related to the growing literature on business accelerators in which selection issues are a main concern to identify their impacts. Business accelerators are an increasingly popular method by which governments help high-growth firms.[8] While these "business schools for entrepreneurs" have been touted by the popular press as critical to the development of startup ecosystems, the evidentiary support for their impact remains thin, especially outside the innovation hubs of Silicon Valley and Boston. Measuring the impacts of business accelerators is hard because cohorts are typically very small and because accelerators put a lot of emphasis on trying to select the best people for the cohort. This difficulty raises concerns that approaches that use matching/selection on observables (as is common in previous studies) will be severely biased.[9] But on the other hand, it is going to be hard to convince many accelerators to randomize which applicants they take.

We make two main contributions to the literature on accelerators. First, we propose a novel identification strategy that can work even when programs are trying to choose the best firms, precisely

[8] The proliferation of business accelerators is well documented; see, e.g.: Cohen and Hochberg (2014); Grover, Mesdvedev, and Olafsen (2019); Clarysse, Wright, and Van Hove (2015); Roberts et al. (2016). The increasing prevalence of public funds for these programs is also well documented; an estimated 40% of businesses accelerators receive some form of government support (e.g., Bone, Allen, and Haley, 2017).

[9] A nonexhaustive list of papers using matching/selection on observables to assess accelerator/incubator impacts includes Colombo and Delmastro (2002); Schwartz (2009); Yu (2019); Smith and Hannigan (2015); Hallen et al. (2014); Lasrado et al. (2016). See also Bone, Allen, and Haley (2019) for a summary of the literature.

by showing the amount of randomness that comes from human decision-making in judging future success. This approach is common in other academic areas, but to the best of our knowledge, this is the first application of this method for evaluating entrepreneurship programs. Given that these programs often select participants based on random allocation of applicants to judges, it is potentially applicable to many efforts to evaluate these types of programs worldwide. Second, we characterize for the first time the treatment heterogeneity of the nonfinancial services provided by accelerators. Most prior impact estimates cannot be extrapolated beyond a small subsample of entrepreneurs with similar growth potential; for example, in papers exploiting qualifying thresholds to assess impacts using regression discontinuity designs, as is the case in the closest paper to ours in this literature, González-Uribe and Leatherbee (2018a). The patterns of treatment heterogeneity show that selection lies at the heart of accelerators' success, as impacts are visible only for high-potential entrepreneurs.

Our research laboratory is, however, not without limitations, and the main drawback is the small cross-section of applicants in business accelerator programs. We have 675 firm-year observations, corresponding to a five-year panel for the 135 applicants who were evaluated by the expert judges. To address potential small sample biases (namely, the possibility that our effects pick up the 5% chance we would see an effect when no such effect exists), we show the robustness of results to using different measures of scoring generosity, sets of controls, methodologies, and outcome variables. We address potential issues from high serial correlation in outcomes using the approach by Fee, Hadlock, and Pierce (2013).

We also show that the results continue to hold when using local randomization methods (Rosenbaum, 2002) that conduct exact finite sample inference and 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. In terms of external validity, we show that the sample is not special by tracing similarities between the entrepreneurs in our sample and the average applicant to ecosystem accelerators worldwide. We emphasize, however, that the external validity of our findings is likely confined to other ecosystem accelerators in developing countries that attract young businesses with traction and have access to high-quality resources, including staff, mentors, and judges.

The rest of this paper proceeds as follows. In Section 2, we describe the context and data. In Section 3, we detail the empirical strategy and present results. We discuss the interpretation of results and their external validity in Section 4. We present robustness checks in Section 5 and offer concluding remarks in Section 6.

2. Institutional setting

ValleE is a local ecosystem business accelerator that was 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.[10] 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.[11] As is common among ecosystem accelerators, ValleE's main objective is to encourage local growth by identifying and boosting high-growth entrepreneurs (cf. Clarysse, Wright, and Van Hove, 2015).[12] Examples of ValleE participants include "Luces projects," a company offering residential wind energy solutions, and "Contratan.do," an information and communication technologies business-to-business (B2B) hiring platform in Latin America.

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. Cohen and Hochberg, 2014; González-Uribe and Leatherbee, 2018b). As explained in more detail in Section 2.2, participants are selected based on average scores from partially overlapping three-judge panels 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 (zero to three years) business located in the region is, in principle, eligible for the program. However, the program focuses on high-growth entrepreneurs, and many applicants are de facto incompatible and are thus rejected (as explained in more detail in Section 2.2).

Also similar to traditional business accelerators, ValleE provides participants with firm capabilities (which we describe in more detail below). The distinguishing feature of our setting, however, is that the program offers no cash (as is nevertheless common among the subset of ecosystem accelerators worldwide; cf. Clarysse, Wright, and Van Hove, 2015).[13] The perception is that for many young businesses, the foremost constraint to growth is their lack of firm capabilities. By contrast, most other entrepreneurship interventions attempt to alleviate financial constraints by providing seed capital. The narrative in ecosystem business accelerators instead argues that entrepreneurs with access to positive net present value opportunities could have no actual ability or market recognition to successfully execute those opportunities. Consequently, the businesses of these entrepreneurs would not grow even if they received cash injections. For example, financial support will not unleash the

[10] 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%).
[11] 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.
[12] The top two impact objectives among ecosystem accelerators are employment generation (35%) and community development (30%). Source: Entrepreneurship Database at Emory University.
[13] Circa 55% of the ecosystem accelerators in the Entrepreneurship Database at Emory University provide no seed capital.

growth potential of a B2B business that is not “plugged into” the right network to secure a key corporate client. In this case, a personal introduction to the corporation’s upper management by the accelerator’s staff or mentor could improve the outlook of the venture more than seed capital would.

Like traditional business accelerators, ValleE provides participants with firm capabilities through a variety of services, including standardized grouped business training, one-to-one customized advice, and increased visibility. The business training sessions are highly structured and simultaneously attended by all participants in the offices of the Chamber of Commerce. They consist of roughly eight weekly hours of standardized content (totaling 100 hours over a space of three months) delivered by hired local and national experts. Bootcamps combine lecture-based conceptual sessions together with case-based sessions discussing real-life practical examples and cover the topics of business modeling, early stage financing, market validation, prototyping, accounting, and pitching. Two types of one-to-one customized advice sessions are provided. The first type consists of bimonthly 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 can 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 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’s website and monthly publications and are 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).

It is possible that participants benefit through mechanisms other than the services per se. For example, services such as customized advice and visibility can trigger potential network, certification, or internal validation effects that have been shown to have large impacts in other contexts, such as among prestigious business school students (e.g., Lerner and Malmendier, 2013). We return to this point in Section 3.6, in which we discuss the potential channels of impact.

2.1. 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.[14] 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 2.2.

Our sample size is standard for business accelerator programs and exceeds that of similar papers exploring the impact of business training (e.g., 14 participants and 14 control plants: Bloom et al., 2013; 47 participants and 66 control business owners: Mano et al., 2014). However, it 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 3, in which we describe several robustness checks we run to address this issue, and in Section 5, in which we conduct exact finite sample inference using a randomization inference approach that is valid even when the number of observations is small.

Based on the program's records, we constructed several variables to use as controls in our empirical strategy: 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 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 of sector experience, are male (79%), educated (67% have a bachelor's or master's degree), are likely to be serial entrepreneurs (61%), and have a founding team (88%). The average number of founders is three, and the average number of employees is four. The likelihood of positive revenues is 45%, and median (positive) annual revenue is $52M COP, or approximately. $15,000 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 [$1.3K USD] in revenue and 2.7 employees in 2014; see Table 1).

Our sample is comparable to average applicants of ecosystem accelerators worldwide, based on information from the Entrepreneurship Database (ED) program at Emory University. The average

[14] 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 to track different projects evaluated by the same judge.

applicant to ValleE is similarly sized (ED applicants have an average of 3.5 employees, a 43.2% likelihood of positive profits, and median [positive] revenue of $12,000 USD) but 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).[15] 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%), to have between two and three employees, and to be predominantly from services industries such as e-commerce (18%). Finally, our sample is also similar to that in prior work on early stage ventures. Haltiwanger, Jarmin, and Miranda (2013) show that 33% of young firms (less than a year old) in the United States 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.

2.2. 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 three judges who individually score the application.[16] The total number of judges is 50, and thus judges 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. Finally, the staff at the accelerator makes the final decision by picking the top 35 applicants based on average scores. It is impossible for judges and applicants to manipulate the ranking process. Judges are unaware of the weight of each criteria in the final score; they independently score projects, are not aware of the identity of the other judges in the panel, and no judge sees all applications. Applicants do not know who their judges are, nor do they know their position in the ranking. [17] The capacity threshold of 35 participants was determined prior to the launch of the program and is due to budget and space limits.

[15] See https://www.galidata.org/accelerators.

[16] The main reasons behind using judge panels (rather than individual judges) are to minimize the burden on individual judges (given their time constraints) and to 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.

[17] Entrepreneurs were never given their ranking or scores to avoid any negative psychological effects or to avoid creating 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.[18] The maximum length allowed for business plans submitted with the applications and read by the judges was two pages. The average number of projects scored by any given judge was 8, and the minimum (maximum) was 5 (14). The program picked the judges based on the relevance of their backgrounds to help sort applicants. Judges were not compensated for evaluating applicants, and their identities are private to us. The pool of 50 judges included individuals with substantial experience in business and entrepreneurship, such as C-level executives in local businesses, independent business consultants, and industry experts as well as managers in entrepreneurship departments in development agencies and two staff members. This average business and industry expertise of judges is not necessarily common among other business accelerator programs, in which applications are managed by platforms that rely on a wider variety of less "hands-on" experienced judges such as academics (cf., González-Uribe and Leatherbee, 2018a).

Compliance with the selection rule was perfect: the top 35 applicants (based on judges' average scores) were selected, and all selected applicants participated (see Fig. 1, Panel A). Table 2 shows statistically significant differences at the time of application between accelerated and nonaccelerated applicants: participants have bigger founding teams, 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 nontransformational entrepreneurs from the sample.

While the accelerator provided uniform criteria by which a judge should score proposals (see Online Appendix 1), we show in Section 3.1 that there was substantial variation in the interpretation of these criteria across judges in the first cohort of ValleE. This heterogeneity in scoring generosity is reminiscent of the systematic differences in judge leniency reported in other settings, such as in bankruptcy courts in the United States (e.g., Dobbie and Song, 2015). In Sections 3.2 and 3.3, we discuss how we use this heterogeneity in scoring generosity across the randomly assigned judges to estimate the causal impact of participation in the accelerator.

2.3. Outcome data

Table 3 presents summary statistics of the main variables used in the regression analysis, including data on firm outcomes. We use two complementary strategies to collect outcome data. First, we collected novel administrative data from the business registry in Colombia on registration, survival,

[18] 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.

and annual revenues. The Colombian registry includes information on annual revenues and closures because Colombian firms submit annual mandatory business license renewals with the Chambers of Commerce that manage the registry. 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 track annual revenues two years before and three years after application to the program. We also use these data to discern which applicant ideas eventually turned into actual businesses (i.e., startup rates) and to 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). Access to administrative data represents a major advantage relative to most other work in the literature of accelerators and in the literature of business training interventions. With a few exceptions, most prior related work relies on surveys to measure performance (see Woodruff, 2018). We also improve upon prior work in which the median number of follow-up observations per firm is one (cf. McKenzie and Woodruff, 2014). Longer follow-ups are important, as short- and long-term impacts of many policies can differ substantially (cf. King and Berhman, 2009).

Our second strategy to collect performance data was to partner with ValleE to design 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, employment, and profits. Our main objective for the survey was to collect additional performance metrics based on employment, profits, and fundraising, which are not included in the business registry, and which we used to explore other outcomes, and to run robustness checks (see Online Appendix 7). Survey response rates were 77%, 67%, and 60%, respectively, in the 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).

One novel feature of our data collection strategies is that we have information on revenues from both administrative data and surveys. Thus, we can cross-check self-reported revenues in the surveys with those in the registry to gauge the degree of potential selective survey attrition and selective survey responses. We find little evidence of either, which mitigates concerns of data quality from the survey variables and lends credence to the additional analysis using the information on employment, profits, and fundraising (see Online Appendix 7). On average across the survey years, most (74%) survey attrition can be explained by real business closures rather than by 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., nonrespondents 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).[19] There is 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), and discrepancies do not vary across participants and nonparticipants.

3. Empirical strategy and results

In this section, we explain how we exploit the random allocation of projects across judges with different scoring generosities to show causal and heterogeneous impacts of the accelerator on growth. We begin by showing that there is a large variation in scoring generosity across judges (i.e., some judges tend to assign high scores, whereas some judges tend to assign low scores). We then show that scoring generosity strongly predicts selection into the accelerator. Next, we use an IV approach based on scoring generosity as an exogenous predictor of acceleration to assess the program's impact. Finally, we characterize impact heterogeneity by exploiting the continuity of our instrument.

3.1. Scoring generosity

We provide evidence of systematic differences across judges in scoring generosity by exploiting the multiple judge assignment per applicant to run fixed effects models of application scores against judge and applicant fixed effects. Our approach is similar to the methodologies in papers assessing the importance of managers in corporations (cf. Bertrand and Schoar, 2003) and general partners in limited partnerships (Ewens and Rhodes-Kropf, 2015). The idea is that judge fixed effects would be jointly significant if judges systematically vary in their tendency to assign high or low scores to projects.

We begin by decomposing individual scores into company and judge fixed effects using the following regression:

$$(1)\ Score_{ij} = \alpha_i + \gamma_j + \varphi_{ij}\ ,$$

where α_i are project fixed effects and γ_j are judge fixed effects. We normalize scores so that they vary between zero and one, corresponding, respectively, to the "bottom score" project (scored at one by all judges) and the "top score" project (scored at five by all judges). Each judge fixed effect is estimated with eight observations, on average (see Section 2.2). 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; they represent "adjusted scores" after

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

controlling for potential systematic differences in scoring generosity across judges. For conciseness in exposition, we refer to the estimated firm fixed effects in Eq. (1) as adjusted scores hereafter. Also, for ease of exposition, we plot the results in Figs. 2 and 3 rather than report regression estimates. Fig. 2 plots the distribution of fixed effects across judges. Fig. 3 plots the average score against adjusted scores; the correlation between these two scores is high at 0.83 (significant at the 1% level).

There are four main findings from estimating Eq. (1). First, there is statistically significant heterogeneity across judges: 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). To address concerns regarding the validity of *F*-tests in the presence of high serial correlation (Wooldridge, 2002), 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).[20] 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 Fig. 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).

The second finding is that the judge heterogeneity reflects systematic differences across judges, as the 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 Fig. 5 shows, the average standard deviation per judge is 0.003, and the maximum is 0.006. Fig 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 (see Fig. 5 for more estimation details).

[20] 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 could still apply. Regarding the second concern, Heckman (1981) and Greene (2001) discuss the ability of small sample sizes per group to allow for meaningful estimates of fixed effects with a rule of thumb of eight observations per group.

The third finding is the sizable economic significance of the scoring generosity heterogeneity.[21] Fig. 2 shows that the most generous (strict) judge adds (subtracts) an average of 0.26 (0.28) to any given project she scores, roughly a third of the mean average score of 0.7. Relying on a panel of judges rather than on individual judges helps mitigate the effect of judge heterogeneity by averaging out the scores of a strict judge with a lenient one in some cases. However, it does not fully correct it because judge panels are small, with only three individual judges assigned per applicant (see Online Appendix 2).

The fourth finding is that these systematic differences across judges are unrelated to the judges' skill in distinguishing high growers and instead reflect judges' propensities to assign high or low application scores. Fig. 6 shows a nil correlation between judges' generosity and their ability to correctly rank applicants. We measure judges' ranking ability using the correlation between a "judge's ranks" and "actual ranks." To produce this correlation for every judge, we rank the companies she evaluated based on (i) 2017 revenue ("actual rank") and (ii) the judge's score ("judge's rank"). Fig. 6 is a scatterplot of each judge's generosity and ranking ability for the 50 judges in our sample.

We note that the results in Fig. 6 do not imply that judges have no ranking ability. In fact, Table 4 shows that projects' actual scores are predicted by the judges' rankings, even after controlling for judge fixed effects. That is, judges are, on average, good at ordering projects according to their potential, even though they vary in how high or low are the scores they assign. Table 4 shows results from regressions of actual ranks on judges' ranks and includes judge fixed effects (Column 2).

However, the results in Fig. 6 do suggest that, in contrast to judges' ranks, average scores are unlikely to be good predictors of performance because scoring generosity artificially inflates (deflates) the judges' perceived potential of lucky (unlucky) projects, as shown in Fig. 3. Instead, adjusted scores are more likely to reflect the predicting ability of judges, evidenced in Table 4, precisely because adjusted scores "clean" the average scores from these systematic differences in scoring generosity across judges.

Consistent with this intuition, Tables 5 and 6 show that adjusted scores predict revenue in a way that adds to the predictive power of the hard information in application responses, whereas average scores are not correlated with future performance (or high growth). Tables 5 and 6 summarize estimates from variants of the following regression using observations during 2013–2017 for all applicants to the accelerator:

$$(2)\quad k_{it} = \alpha + \beta\, Score_i + \rho\, Score_i \times After_t + \theta Z_{it} + \mu_t + \varepsilon_{it}\ ,$$

where *i* indicates applicants and *t* time, $After_t$ is a dummy that equals one during 2015–2017, Z_{it} is a vector of controls, and $Score_i$ corresponds to either the *Average score or Adjusted score*. We include time (μ_t) fixed effects and several controls at the founder and project levels (i.e., the hard information

[21] In Fig. 2, 34% of judges tend to systematically award individual scores that are one standard deviation above or below the average score of the other judges, whereas only 6.8% of judges do so in the 500 placebo assignments (Panel B in Fig. 4; see Fee, Hadlock, and Pierce, 2013).

described in see Section 2.1) as well as interactions of these characteristics with the variable $After_t$. We bootstrap standard errors clustered at the applicant level to account for any serial correlation across applicants (Young, 2018). In unreported analysis, we show results are robust to computing standard errors based on a bootstrap procedure that also takes into account the fact that the adjusted score is a generated regressor following (Cho, 2019). Table 5 shows ordinary least squares (OLS) estimates using revenue as a dependent variable; Table 6 estimates a probit model using the variable for gazelles as a dependent variable and indicator. We define gazelles as applicants who are in the 90th percentile of revenues by 2017.[22]

Table 5 shows that the top adjusted score project has an additional $86M COP ($26K USD; Column 2, Panel A, Table 5) in annual revenue relative to the bottom adjusted score project, controlling for both secular growth and applicants' characteristics (relative to a baseline level of revenues in 2014 of $25M COP; Table 3).[23] The correlation between adjusted scores and future revenue is strongest for business ideas (see Online Appendix 3) and is not explained by a potential treatment effect. Rather, it reflects the predictive power of judges, as it holds even after controlling for participation (see Column 3, Panel A, Table 5).[24] Table 6 shows that adjusted scores also specifically predict high growth. Applicants among the top quartile of adjusted scores are 20% more likely to become gazelles by the end of the sample period (Column 1, Panel A, Table 6), controlling for covariates and acceleration (Columns 3 and 4, Panel A, Table 6). In contrast to adjusted scores, average scores are uncorrelated with revenue (see Columns 5 and 6, Panel A, Table 5) and explain a much lower fraction of the variation in revenue than entrepreneurs' characteristics (see Column 3, Panel B, Table 5).

[22] So defined, these gazelles surpass top Colombian growers: they expand their initial revenue about five times, roughly doubling the 90th percentile of the four-year life cycle revenue growth of Colombian businesses (cf., Eslava and Haltiwanger, 2018; see Panel B, Table 5). We note, however, that 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 gazelle literature that define gazelles using performance thresholds (Kirchhoff, 1994; Picot and Dupuy, 1998; Schreyer, 2000; Fritsch and Weyh, 2006; McKenzie and Sansone, 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 revenue 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 one year relative to incorporation (77 applicants) and (ii) less than one year since incorporation or not incorporated (58 applicants). We then define as gazelles the top 10% of firms in each group according to revenue in 2017 (nine and seven applicants in each group, respectively; see Online Appendix 4). Applicants classified as gazelles would also classify as high growers under definitions based on growth rates rather than levels: Panel B of Table 5 shows that their average annual revenue growth is 68% in the three years following their application, which exceeds the 20% growth rate requirement for gazelles in most other definitions used in the literature (20% for Birch and Medoff, 1994 and Birch, Haggerty, and Parsons, 1995; 50% for Ahmad and Petersen, 2007; Deschryvere, 2008; Autio, Arenius, and Wallenius, 2000).

[23] Adjusted scores explain 4.4% of variation in revenues, almost a third of the contribution of the entire set of entrepreneurs' characteristics (Column 3, Panel B, Table 5).

[24] 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.

Why are ValleE judges able to predict high growth in a way that adds to the predictive power of the hard information in application responses? The reason is that judges evaluate projects not only based on applicants' characteristics (i.e., the regression controls) but also based on the business plans, which are unobservable to the econometrician and whose information is not easily codified. Results in Tables 5 and 6 thus suggest that business plans have information that helps predict future growth when evaluated by judges. We note, however, that judges' predicting ability may not necessarily replicate well in other settings, as ValleE judges have uncommonly high business and industry experience compared to the judges in other business accelerators (see Section 2.2). We return to this point in Section 4, in which we discuss the external validity of the findings.

In conclusion, the results from this section show that the program made selection mistakes because it selected participants on the basis of average scores rather than on the basis of judges' ranks. We refer to the unlucky applicants who were rejected because of their strict judges in spite of their high potential as type one selection mistakes (or type one applicants). Analogously, type two selection mistakes (or type two participants) correspond to the lucky applicants who were accepted in spite of their low potential because of the generosity of their judges. We now explain how we exploit these selection mistakes to assess the causal impacts of the program.

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 scoring generosity as an IV 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 2.1), 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 Eq. (3) is

$$(4)\ Acceleration_i \times After_t = \alpha + \vartheta_i + \beta f(SG) \times After_t + \theta X_{it} \times After_t + \mu_t + \epsilon_{it},$$

where SG is the "project's scoring generosity," defined as the sum of the fixed effects of the three judges who evaluated each project (see Online Appendix 2). For the sake of space, we refer to a project's scoring generosity simply as "scoring generosity" hereafter. We present the 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 scoring generosity interacted with $After_t$ to instrument for acceleration yields a consistent 2SLS 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 sum of the (average) leave-one-out measures of judge's scoring generosity introduced in Section 3.1 to build our instruments for acceleration. We also estimate "leave-one-out adjusted scores" by subtracting the sum of the (average) leave-one-out judge fixed effects from the average score of each project. The correlation between the adjusted score and the leave-one-out adjusted score is high at 0.98 (significant at the 1% level). In unreported results, we verify that the results are similar using the raw measures of scoring generosity to construct the instrument as well as the raw measure of adjusted scores.[25]

The ρ estimate measures the local average treatment effect of the accelerator for applicants whose participation is altered by scoring generosity. These applicants include both the type one and type two selection mistakes by the program (i.e., applicants who in spite of their potential were, respectively, mistakenly rejected/accepted due to the generosity/strictness of their judges). 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.

The first assumption is empirically testable. Panel A in Fig. 7 shows a positive and nonlinear association between acceleration and scoring generosity. The figure plots average acceleration versus our leave-one-out measure of scoring generosity and shows that the relation between these two variables is monotonic and exhibits a larger slope in the upper tail of scoring generosity. The positive association cannot be explained by applicant heterogeneity because judges were randomly assigned, and indeed Panel B in Fig. 7 shows evidence of a positive association holding constant applicant quality (as measured by adjusted score). To produce Panel B in Fig. 7, we classify applicants into quartiles of scoring generosity and estimate for each quartile the distribution of acceleration over adjusted scores.[26] The figure shows that for a given adjusted score, the acceleration probability is always highest (lowest)

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

[26] 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). 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.

for projects assigned to the top (bottom) quartile of scoring generosity. With this classification, the nonlinearity in the relation between scoring generosity and acceleration becomes apparent, as further shown in Table 7. The table shows that for any quartile of scoring generosity, the probability of acceleration is always below 6% among low-quality applicants (Column 2, Table 7). Instead, for applicants in the 75th percentile of quality, the probability of acceleration decreases from 99.78% to 3.32% when we move from judges in the top to the bottom quartile of scoring generosity (Column 4, Table 7). Given the nonlinear relation of scoring generosity and acceleration, we use quartiles of scoring generosity as our main explanatory variables in Eq. (4).[27] We discuss nonlinearity further in Section 3.4, in which we present estimates of marginal effects using nonparametric approaches.

The first-stage results in Table 8 show a large and precisely estimated relation between quartiles of scoring leniency and the probability of acceleration. The results in the table 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). We formally test the relevance of the instrument and report the *F*-test of joint significance of the quartiles of scoring generosity, showing that the instruments are not weak (Stock and Yogo, 2005).

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 Fig. 3 (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 Online 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. Scoring generosity also does not measure differences in predicting ability across judges, as shown in Fig. 6 (see also Section 3.1). Finally, because applicants are not made aware of their scores, nor of the generosity of their judge panel, psychological reactions are also unlikely (e.g., feelings of grandeur or depression). Ultimately, however, the assumption that scoring generosity only systematically affects applicants'

[27] We estimate Eq. (4) with f(SG) as $\sum_{i=2,3,4} Quartile_i \times After_t$, where $Quartile_i$ is a dummy indicating the *i*th quartile of scoring generosity (the left out category is the bottom quartile).

performance through acceleration is fundamentally untestable, and our estimates should be interpreted with this identification assumption in mind.

For the monotonicity assumption, we summarize the supporting evidence of several tests presented in Online 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. In Online Appendix 6, 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

In this section, we summarize results from the IV regressions. We defer the more detailed interpretation of the results to Section 4.

Table 9 shows compelling evidence of causal impacts of acceleration. Over the first three post-application years, acceleration increases annual revenue by \$66.3M COP (\$20K USD; see Column 3). This effect corresponds to a 166% (130%) increase from the rejected applicant's (average applicant) 2017 revenue of \$51M COP (\$40M COP). Complementary analysis suggests that the local average effect is not driven by a few outliers: Online Appendix 8 shows a shift in the revenue distribution three 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.

The IV estimates the local average treatment effect (cf., Imbens and Angrist, 1994) for the applicants at the margin of selection, which include type one and type two selection mistakes. Intuitively, the IV averages out two types of performance comparisons: first, the performance difference between high-potential participants and similar potential type one applicants who were mistakenly rejected, and second, the performance difference between low-potential rejected applicants and similar potential type two participants who were mistakenly accepted. A natural question asks which performance comparison drives the IV results. We turn to this question in Section 3.4, in which we explore the heterogeneity of impacts.

We contrast the IV estimate with the naïve OLS estimate of Eq. (3) 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). This positive difference suggests that the projects at the margin of acceptance are most sensitive to acceleration (cf. Card, 2001). We come back to this point in Section 3.5, in which we compare results across different methodologies of impact estimation.

A comparison between Columns 6 and 9 in Table 9 reveals that the increase in revenue is driven by established firms (Column 9) and is not significant for business ideas (Column 6). Established firms had an additional annual revenue of $116M COP ($35K USD) during 2015–2017 (or 2.4 times their initial revenue), whereas the estimate for business ideas is negative, albeit not statistically significant. The results in Table 10 provide a possible explanation for this difference in estimated average impact across established firms and business ideas. The table shows that accelerated entrepreneurs who 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 accelerator 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.

In robustness checks reported in Online Appendix 7, we show similar evidence of causal impacts if we use survey-based information on employment and profits. In contrast, we find no evidence that acceleration leads to additional survey-based fundraising. Over the first three post-application years, only 21 applicants secured an average external financing of $25K USD, including 9 participants and 12 rejected applicants. Of those that secured external financing, 14 sourced it from bank loans, 4 from combined bank loans and equity, and 3 relied on equity only. These fundraising results are not surprising given the underdevelopment of private equity markets in Colombia, particularly in Cali, where the first formal network of business angels and the first local private equity were only launched in 2017.

3.4. Impact heterogeneity: Who benefits from acceleration?

The results so far show compelling evidence of average treatment effects among the applicants whose acceleration is altered by the judge assignment. One question that remains regards the drivers of the IV effect (i.e., is it explained by type one and/or type two selection mistakes?). More generally, can the accelerator's impact differ across entrepreneurs of different growth potential? No rigorous evidence exists on this point, as prior accelerator impact estimates cannot be extrapolated beyond small subsamples of entrepreneurs with similar growth potential, such as in studies exploiting qualifying thresholds using regression discontinuity designs (e.g., González-Uribe and Leatherbee, 2018a).

To investigate the drivers of the IV effect, we take advantage of the continuity of our instrument. Our goal is to estimate changes in accelerator impacts as we move from more low-growth-potential applicants to more high-growth-potential applicants by progressing from type two to type one mistakes. We can approximate this goal because the wide variation in the scoring generosity of the judges implies that the selection mistakes span different parts of the growth-potential distribution of applicants. Panel B in Fig. 1 illustrates this wide span, in which the mistakenly accepted (rejected) applicants who make up the type one selection mistakes (type two selection mistakes) correspond to the solid (open) dots at the left (right) of the 35th rank threshold.

We begin by running matching regressions, in which 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 on judge scoring generosity only (relative to the information sets of the econometrician and judges). In detail, at the left of the 35th rank threshold, we match type two participants to correctly rejected applicants. At the right of the 35th rank threshold, we match correctly selected participants with type one rejected applicants. 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 that 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. The results are presented in Online Appendix 9. 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 accelerator impacts by each level of propensity score for acceleration (within the common support) as the average difference in post-application annual revenues between participants and matched applicants, in which the kernel weights are used to weight the outcomes of the matched applicants (cf. Smith and Todd, 2005).[28] Online Appendix 9 shows that the average of the individual treatment 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.

[28] 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.

Finally, we transform our data so that the individual impacts constitute the observed data subject to further modeling. We then apply nonparametric regressions to the transformed data to predict the relation between the individual impact estimates and the propensity score of acceleration within the common support (cf., Xie, Brand, Jann, 2012). The first derivative of this relation is then evaluated at different values of acceleration propensity using the coefficients from the nonparametric regression. We calculate standard errors using the standard deviation of the marginal effect estimates from a bootstrap procedure with 500 iterations. The identification assumption behind the marginal effect estimates 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). We plot the results in Fig. 8.

Fig. 8 shows a large heterogeneity in impacts, with the evidence pointing to an increasing function based on an applicant's growth potential. The shape of this function helps us understand the IV results from Section 3.3 and more generally sheds light on the types of applicants who benefit most from acceleration. The figure plots the marginal effects at different values of acceleration propensity in the common support, as well as the bootstrapped confidence intervals. For values above/below those thresholds of the acceleration propensity score, we cannot estimate marginal effects, as there are no selection mistakes to use in the estimation (i.e., no applicants with an acceleration propensity below (above) 0.35 (0.75) were mistakenly selected (rejected) by the program).

The heterogeneity patterns in Fig. 8 show that the positive LATE results in Section 3.3 come from type one selection mistakes rather than from type two selection mistakes. The figure shows a remarkable growth of high-potential participants relative to similar quality applicants who were mistakenly rejected (i.e., type one selection mistakes) because of the strictness of the judges. Instead, no apparent performance improvements are visible in the figure for the low-potential participants who were mistakenly accelerated (i.e., type two selection mistakes) because they were assigned to more generous judges. This pattern of impact heterogeneity demonstrates that accelerators can add value by boosting top growers but not necessarily by transforming low-potential entrepreneurs into high growers. In Section 4, we discuss in detail the interpretation of these results in the wider context of constraints to high-growth entrepreneurship in developing countries and of firm capabilities' interventions.

Our approach is made in a similar spirit as the estimation of marginal treatment effects (MTEs) in the microeconomics literature (e.g., Heckman and Vytaclil, 2005) and particularly in the judge leniency literature (e.g., Doyle, 2007). MTE estimates in our context would illustrate how the outcomes of applicants on the margin of acceleration change as we move from more strict to more generous judges. Given the increasing function of marginal effects in the acceleration propensity score of Fig. 8, we expect the MTE function for revenue to be decreasing in scoring generosity. This decrease occurs because the margin for relatively generous judges should entail relatively fewer deserving applicants, as measured by our acceleration propensity score. Consistent with this intuition, we show in unreported analysis that the MTE function for revenue conditional on adjusted scores has a decreasing slope in the probability of acceleration, as predicted by scoring generosity.

One important departure between our analysis of treatment heterogeneity and MTEs' estimation in the judge leniency literature is that we exploit the multiple judge assignment per applicant to control for applicant heterogeneity using the adjusted scores from the fixed effects models estimated in Section 3.1. These controls are not possible in settings with single judge assignment, such as in most prior applications of judge leniency techniques (e.g., children's welfare, Doyle, 2007; bankruptcy courts, Dobbie and Song, 2015), but they are crucial in our setting, as an unconditional revenue MTE function does not show the decreasing slope in the probability of acceleration. The importance of these controls is as expected given the skewness in young firm growth, as well as the ability of ValleE judges to distinguish high growers, which is captured in the adjusted scores (see Section 3.1).

3.5. Challenges in measuring program impacts on young firm growth

We end the presentation of our results by illustrating how the unobservable heterogeneity in young firms' potential affects the interpretation of estimates from different identification strategies typically used in the analysis of entrepreneurship interventions, including business accelerators. In this section, we discuss the advantages and limitations of the different approaches and compare their estimates with our preferred specifications.

Table 11 summarizes results from different estimations of Eq. (3). To conserve space, we do not present further details of these additional exercises. Columns 1 and 2 replicate the OLS and IV estimates from Table 9 for ease of comparison. Column 3 in Table 11 reports the estimate from the most popular methodology used to quantify the effect of business accelerators: propensity score matching (PSM; see Bone, Allen, and Haley, 2019). To produce this estimate, we match participants with similar rejected applicants as measured by characteristics at time of application only. The number of observations decreases by design relative to those in Column 2; dependent variables include 2015–2017 revenues only, as the match parameters include revenue at application. The popularity of this method relies mostly on the fact that many of these programs are not designed to be evaluated, and therefore evaluations designed ex post must rely on constructing control groups using matching procedures based on observable predictors of growth potential (e.g., serial founder). The main drawback of these methods is their inability to control for heterogeneity in unobservable growth potential.

Column 4 in the table reports results from a second estimate based on PSM, in which we match applicants based on characteristics at application and adjusted scores. The number of observations decreases relative to Column 3, as matches are additionally required to be in the common support of adjusted scores. This strategy uses judges' scores as a proxy of projects' growth potential that is unobservable to the econometrician but observable to the judges. Table 5 validates this proxy by showing that adjusted scores predict growth (even after controlling for participation), which demonstrates the predicting ability of ValleE judges. In settings in which no such ability is demonstrated by judges, adjusted scores will be poor proxies of unobservable growth potential (see, e.g., McKenzie and Sansone, 2017).

Finally, Column 5 presents estimates from a regression discontinuity (RD) approach, in which we exploit the program's ex-ante determined capacity threshold, which implied that only the top 35 projects based on average scores were chosen. We subtract revenues at application from the dependent variable and estimate the model over the 2015–2017 period to allow for the comparison of coefficients with the rest of the estimates in the table. The advantage of this strategy is the exogeneity of the cutoff. One of the main drawbacks is its reliance on the continuity at the threshold, which in a setting like ours does not hold due to the variation in unobservable quality near the threshold. Another drawback is the inability to directly extrapolate results beyond observations close to the threshold.

There are four main insights from the results in Table 11. First, the similarity between the estimates in Columns 1 and 3 highlights the concern that approaches using matching/selection on observables will be severely biased. The simple PSM does not correct the OLS bias, as participants and matched rejected applicants differ on dimensions that are unobservable to the econometrician (i.e., projects' quality). Second, the similarity between Columns 2 and 4 highlights the intuition behind our IV approach, which is to infer impacts from outcome differences between applicants with similar potential (i.e., adjusted score and covariates) but exogenously different acceleration status due to differences in judge assignment luck. Third, the positive difference between Columns 5 and 2 highlights the challenges in using RD to assess impacts. In our setting, the RD inflates the impact because several high-potential companies rank close to the qualifying threshold. These companies, as we showed in Section 3.4, are the ones that benefit the most from acceleration. Using the IV estimate instead averages out the large impacts on high-potential projects and the small effects on low-potential entrepreneurs. The main drawback, however, is external validity, as the estimates are only representative for applicants at the margin of selection (i.e., type one and type two selection mistakes), which can differ from the potential impact on other applicants in the population.

Finally, the comparative analysis between the IV and RD estimates (Columns 2 and 5, respectively) highlights the utility of exploring treatment heterogeneity in Section 3.4. The results in Fig. 8 help explain the difference between the IV and RD estimates as stemming at least partially from heterogeneity in impact for different levels of acceleration potential.

3.6. *Channels of impact*

Why are there such large benefits caused by acceleration? In this section, we explore this question in several ways. First, we look at which accelerator services have the largest apparent effects according to surveys of participants based on different data cuts. We then consider other channels in addition to the accelerator services through which the program could also affect firm capabilities, such as certification effects that improve market recognition. While we cannot provide rigorous evidence of these additional mechanisms with our current data, we discuss related suggestive evidence.

Online Appendices 10 and 11 provide compelling evidence that customized advice and visibility, rather than the standardized business training, are the most impactful aspects of the intervention. Online Appendix 10 shows that, consistent with entrepreneurs' perceptions in business accelerator programs worldwide (see Roberts et al., 2016), 74% of ValleE participants selected advice and visibility as the program's most valuable aspects in follow-up surveys. In contrast, only 8% of surveyed ValleE participants described the grouped business training as a key impact driver. Regarding the business practices that were most impacted by the program, 54% of entrepreneurs reported having found a business contact thanks to the program, and 34% reported using the program to show their products/services to other businessmen that shared their interests. Online Appendix 11 shows evidence that impact effects are larger for entrepreneurs who would presumably benefit the least from grouped business training, such as the more educated entrepreneurs (those with a college degree or graduate studies). Impacts are also larger for projects that indicated in the baseline that strategic advice (as given by mentors) was their main constraint on growth. Impacts are instead not visible for applicants who did not indicate that they needed strategic advice at application.

This additional evidence confirms prior findings on the outperformance of customized interventions over standardized business training programs (Karlan and Valdivia, 2011; Campos et al., 2018; Ubfal et al., 2019; Bruhn, Karlan, and Schoar, 2018; Lafortune, Riutor, and Tessada, 2018). It also suggests that impact mechanisms other than the accelerator services per se could also be at play, as customized advice and visibility can trigger network, certification, and validation effects that can affect firms' capabilities and have been shown to have large impacts in other contexts, such as among students at prestigious business schools (e.g., Lerner and Malmendier, 2013).[29] As is common in the literature on returns from education, we cannot rigorously distinguish these additional mechanisms with our current data, and thus we only discuss suggestive evidence (e.g., Chevalier et al., 2004). For example, the heterogeneity of impacts discussed in Section 3.4 goes against the certification or validation effects, as by definition, these effects should be higher for the low-potential applicants (or

[29] For evidence on the value of network and certification effects, see Megginson and Weiss (1991); Hsu (2004); Fafchamps and Quinn (2016); Brooks et al. (2018); Cai and Szeidl (2018).

homogenous across applicants of different quality). Evidence from follow-up surveys points to potential network effects, as 53% of participants reported an improved ability in finding business contacts as a consequence of participating in the program (see Online Appendix 10).

4. Interpretation of results and external validity

Overall, the results in Section 3 provide compelling evidence that alleviating constraints to firm capabilities has a first-order effect on young firm growth, specifically by unlocking innovative entrepreneurs' potential, and not necessarily by transforming subpar businesses. The implications are twofold. On the one hand, the results imply that high-potential entrepreneurs face barriers to growth besides financial constraints, which can be mitigated by business accelerators. These results echo the investment thesis in venture capital that is based on the provision of nonfinancial services rather than on the provision of seed capital on its own.

On the other hand, the results also highlight the limitations of policies using business accelerators to promote firm growth. The impact of such policies would seem to rely on the ability of accelerator programs to identify and attract high-potential businesses given that no discernible impacts are visible for low-potential participants. However, we note that the standard "joint test" caveat in firm capabilities' interventions applies here as well. That no impacts are visible in low-quality projects does not necessarily mean that business accelerators cannot add value to projects in the left tail of the growth potential distribution or that firm capabilities cannot be taught to these types of entrepreneurs. An alternative explanation for why impacts are not visible in low-quality applicants is that the firm capabilities typically provided by business accelerators are not enough for these types of projects and more intensive interventions are needed (see Bruhn, Karlan, and Schoar, 2018).

In terms of magnitude, our impact estimates are similar to those found in evaluations of business accelerators in developing countries (e.g., Goñi and Reyes, 2019). However, they are generally larger than those in 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 1 and 1.5 years), whereas we estimate an increase of 166% in annual revenues over the first three post-application years for rejected applicants. Buttressed by the heterogeneous impact patterns and the survey evidence, the differences in magnitude relative to interventions on microenterprises are likely explained by two distinct factors: the high-growth focus of our sample relative to subsistence enterprises and the additional provision of more intensive customized advice and visibility, which is not common in business training interventions. This last factor can also explain why prior work on business plan competitions finds that training has little effect for high-potential firms (Fafchamps and Woodruff, 2016; McKenzie, 2017), as the short-term nature of business plan competitions also prevents the inclusion of more intensive advice and visibility in these programs.

In terms of external validity, several aspects of our setting suggest that these results are not only confined to ValleE's experience. For starters, ValleE is very similar to the average ecosystem accelerator on many dimensions. For example, the location of the program outside the capital city of Colombia is a common trait among ecosystem accelerators. Roughly 38% of these programs are located in underdeveloped regions. Forty percent are in the United States but are outside Silicon Valley, Massachusetts, New York, or Washington DC; the rest are in Europe but are 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.

Going back to the questions we posed in the introduction, what do we learn from our findings about high-growth young firms, and why are they missing in developing countries? Our findings provide evidence that firm capability constraints obstruct the growth of some high-potential entrepreneurs in developing countries and business accelerators can help identify these constrained high-potential entrepreneurs and boost their growth. Using a back-of-the-envelope calculation based on our findings, we show that high-potential participants, on average, reach about three times their initial revenue by the time their companies are four years of age. These participants' four-year life cycle growth rate is roughly double that of Colombian top growers, who instead approximately increased their initial revenue 1.5 times (see Fig. 3 in Eslava and Haltiwanger, 2018). The back-of-the-envelope calculation is as follows: by 2017, rejected applicants increased their initial revenue 1.8 times, from \$22M COP to \$40M COP. Our IV estimates imply that marginal applicants grew 1.66 times more than rejected applicants, so roughly 3 (1.66×1.8) times their revenue at baseline.

5. Robustness checks

The main concern with the results in Section 3 is the potential biases from the small cross-section; namely, 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 robustness of the results to using different measures of scoring generosity, sets of controls, and methodologies, as shown in Section 3, as well as to different outcomes variables (see Online Appendix 7).

Nevertheless, to further address this concern, we use a randomization inference (RI) approach that conducts exact finite sample inference and remains valid even when the number of observations is small (cf. Rosenbaum, 2002). This approach is somewhat similar to the bootstrap approach but is 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 RI instead helps the econometrician address 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 in which 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 being 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 RI 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 Eq. (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. The results are summarized in Table 12 and are illustrated in Fig. 9.

The results in Table 12 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 RI test are that the distribution of the score is the same for all observations in the subsample and initial outcome variables are statistically similar between groups. We present supporting evidence in Panel A of Table 12, in which 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.

6. Conclusion

We show compelling evidence that alleviating constraints to firm capabilities has a first-order effect on high growth, particularly by unlocking innovative entrepreneurs' potential rather than by transforming low-quality projects. Our research laboratory is an accelerator in Colombia that provides participants with grouped training and customized advice and visibility but no cash. We measure firm growth using administrative data. Our empirical strategy exploits the selection mistakes made by the program because it did not control for the heterogeneity in scoring generosity across the judges who were randomly allocated to evaluate the applicants. Our approach based on selection mistakes can work even when organizations are trying to choose the best firms, precisely by showing the amount of randomness that comes from human decision-making in judging future success. We estimate that over the first three post-application years, participating in the accelerator increases annual revenue by 166% relative to rejected applicants. These results are representative for the applicants whose selection decision is altered by the judge assignment and would (not) have been selected but for the strictness (generosity) of their randomly assigned judges.

We provide the first exploration of accelerator treatment effect heterogeneity by exploiting the continuity of our instrument and the wide span of selection mistakes along the distribution of applicant growth potential. We demonstrate that the IV results come from type one selection mistakes rather than type two. There is remarkable growth of high-potential participants relative to unlucky high-potential applicants who were mistakenly rejected by strict judges. By contrast, we find no apparent performance improvements of low-potential participants who were mistakenly accepted by generous judges. Using a back-of-the-envelope calculation, we quantify revenue improvements of 31%–40% for the accelerator had it accounted for judge heterogeneity in scoring generosity and thus selected higher-potential firms (see Online Appendix 12). The results demonstrate that accelerators add value in developing countries by identifying and boosting top growers. It follows that these programs could contribute to solving the development problem by boosting high-potential entrepreneurs, thus reducing the shortage of high-growth entrepreneurship between developed and emerging economies.

References

Ahmad, N., Petersen, D.R., 2007. High-growth enterprises and gazelles–preliminary and summary sensitivity analysis. Unpublished working paper. OECD-FORA.

Autio, E., Arenius, P., Wallenius, H., 2000. Economic impact of gazelle firms in Finland. Unpublished working paper. Helsinki University of Technology, Institute of Strategy and International Business.

Ayyagari, M., Demirguc-Kunt, A., Maksimovic, V., 2014. Who creates jobs in developing countries?. Small Business Economics 43, 75–99. 10.1007/s11187-014-9549-5.

Banerjee, A., 2013. Microcredit under the microscope: what have we learned in the past two decades, and what do we need to know? Annual Review of Economics 5, 487–519.

Bertrand, M., 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? Cognetics Inc., Boston.

Birch, D.L., Medoff, J., 1994. Gazelles. In: Solmon, L.C., Levenson, A.R. (Eds.), Labor Markets, Employment Policy and Job Creation. Westview Press, Boulder and London, pp. 159–167.

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

Bone, J., Allen, O., Haley, C., 2017. Business incubator and accelerators: the national picture. Research paper. Nesta.

Brooks, W., Donovan, K., Johnson, T. R., 2018. Mentors or teachers? Microenterprise training in Kenya. American Economic Journal: Applied Economics 10, 196–221.

Brooks, C., Oikonomou, I., 2018. The effects of environmental, social and governance disclosures and performance on firm value: a review of the literature in accounting and finance. The British Accounting Review 50, 1–15.

Bruhn, M., Karlan, D., Schoar, A., 2010. What capital is missing in developing countries? The American Economic Review 100, 629–633.

Cai, J., Szeidl, A., 2018. Interfirm relationships and business performance. The Quarterly Journal of Economics 133, 1229–1282.

Calderon, G., Cunha, J.M., De Giorgi, G., 2013. Business literacy and development: evidence from a randomized controlled trial in rural Mexico. Unpublished working paper. NBER.

Campos, F., Frese, M., Goldstein, M., Iacovone, L., Johnson, H.C., McKenzie, D., Mensmann, M., 2018. Is personal initiative training a substitute or complement to the existing human capital of women? Results from a randomized trial in Togo. AEA Papers and Proceedings 108, 256–61.

Card, D., 2001. Estimating the return to schooling: progress on some persistent econometric problems. Econometrica 69, 1127–1160.

Chevalier, A., Harmon, C., Walker, I., Zhu, Y., 2004. Does education raise productivity, or just reflect it? The Economic Journal 114, F499–F517.

Clarysse, B., Wright, M., Van Hove, J., 2015. A look inside accelerators: building businesses. Research paper. Nesta.

Cho, T., 2019, Turning Alphas into betas: arbitrage and endogenous risk. Journal of Financial Economics, forthcoming.

Cohen, S.G., Hochberg, Y.V., 2014. Accelerating startups: the seed accelerator phenomenon. Unpublished working paper. University of Georgia and Rice University.

Colombo, M. G., Delmastro, M., 2002. How effective are technology incubators?: Evidence from Italy. Research Policy 31, 1103–1122.

Deschryvere, M., 2008. High-growth firms and job creation in Finland. Unpublished working paper. Research Institute of the Finnish Economy.

Dobbie, W., Song, J., 2015. Debt relief and debtor outcomes: measuring the effects of consumer bankruptcy protection. American Economic Review 105, 1272–1311.

Doyle, J., 2007. Child protection and child outcomes: measuring the effects of foster care. American Economic Review 97, 1583–1610.

Doyle, J., 2008. Child protection and adult crime: using investigator assignment to estimate causal effects of foster care. Journal of Political Economy 116, 746–770.

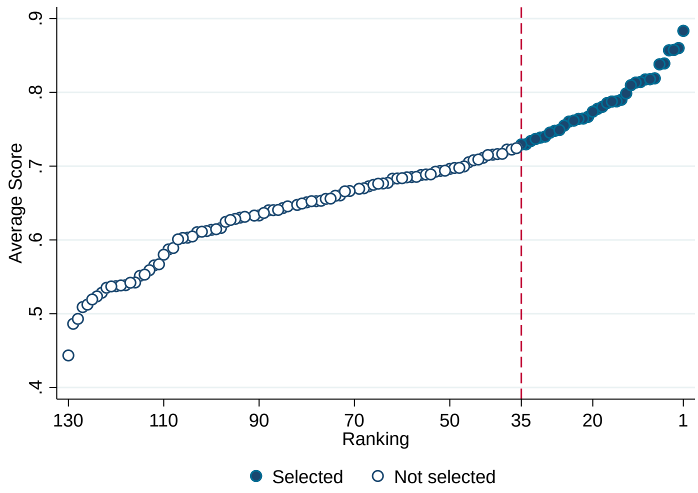
de Mel, S., McKenzie, D., Woodruff, C., 2008. Returns to capital in microenterprises: evidence from a field experiment. The Quarterly Journal of Economics 123, 1329–1372.
———, 2014. Business training and female enterprise start-up, growth, and dynamics: experimental evidence from Sri Lanka. Journal of Development Economics 106, 199–210.
Eslava, M., Haltiwanger, J., 2018. The life-cycle growth of plants: the role of productivity, demand and distortions. Research paper. SSRN.
Eslava, M., Haltiwanger, J.C., Pinzon, A., 2018. Job creation in Colombia vs the U.S.: 'Up or out dynamics' meets 'the life cycle of 'plants.' Research paper. SSRN.
Ewens, M., Rhodes- Kropf, M., 2015. Is a VC partnership greater than the sum of its partners? The Journal of Finance 70, 1081–1113.
Fafchamps, M., Quinn, S., 2016. Networks and manufacturing firms in Africa: results from a randomized field experiment. The World Bank Economic Review 32, 656–675.
Fafchamps, M., Woodruff, C.M., 2016. Identifying gazelles: expert panels vs. surveys as a means to identify firms with rapid growth potential. Unpublished working paper. World Bank.
Fee, C. E., Hadlock, C.J., Pierce, J.R., 2013. Managers with and without style: evidence using exogenous variation. The Review of Financial Studies 26, 567–601.
Fritsch, M., Weyh, A., 2006. How large are the direct employment effects of new businesses? An empirical investigation for West Germany. Small Business Economics 27, 245–260.
González-Uribe, J., 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. Leatherbee, M., 2018b. Selection issues. In: Wright, M. (Ed.), Accelerators. Imperial College Business School and Israel Drori, VU, Amsterdam., 81-100.
Goñi, E.A.G., Reyes, S., 2019. On the role of resource reallocation and growth acceleration of productive public programs. Unpublished working paper. Inter-American Development Bank.
Greene, W.H., 2001. Estimating econometric models with fixed effects. Working paper. New York University.
Greene, W.H., 2001. Fixed and random effects in nonlinear models. Working paper. New York University.
Grover, A.G., Medvedev, D., Olafsen, E., 2018. High-growth firms: facts, fiction, and policy options for emerging economies. Unpublished working paper. The World Bank.
Hallen, B. L., Bingham, C. B., Cohen, S. L., 2014. Do accelerators accelerate? A study of venture accelerators as a path to success. Academy of Management Proceedings.
Haltiwanger, J.C., Jarmin, R.S., Miranda, J., 2013. Who creates jobs? Small versus large versus young. The Review of Economics and Statistics 95, 347–361.
Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. Review of Economic Studies 64, 605–654.
Heckman, J J., Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73, 669–738.
Henrekson, M., Johansson, D., 2008. Gazelles as job creators—a survey and interpretation of the evidence. Unpublished working paper. Research Institute of Industrial Economics.
Hsieh, C.-T., Klenow, P.J., 2014. The life cycle of plants in India and Mexico. The Quarterly Journal of Economics 129, 1035–1084.
Hsu, D.H., 2004. What do entrepreneurs pay for venture capital affiliation? The Journal of Finance 59, 1805–1844.
Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467–75.
Imbens, G.W., Rosenbaum, P., 2005. Randomization inference with an instrumental variable. Journal of the Royal Statistical Society Series A 168, 109–126.
Karlan, D., Valdivia, M., 2011. Teaching entrepreneurship: impact of business training on microfinance clients and institutions. The Review of Economics and Statistics 93, 510–526.
Karlan, D., Knight, R., Udry, C., 2012. Hoping to win, expected to lose: theory and lessons on micro enterprise development. Unpublished working paper. NBER.

King, E.M., Behrman, J.R., 2009. Timing and duration of exposure in evaluations of social programs. The World Bank Research Observer 24, 55–82.
Kirchhoff, B.A., 1994. Entrepreneurship and Dynamic Capitalism. Praeger, Westport.
Kling, J.R., 2006. Incarceration length, employment, and earnings. American Economic Review 96, 863–876.
Klinger, B., Schundeln, M., 2011. Can entrepreneurial activity be taught? Quasi-experimental evidence from Central America. World Development 39, 1592–1610.
Lafortune, J., Riutort, J., Tessada, J., 2018. Role models or individual consulting: the impact of personalizing micro-entrepreneurship training. American Economic Journal: Applied Economics 10, 222–245.
Lasrado, V., Sivo, S., Ford, C., O'Neal, T., Garibay, I., 2016. Do graduated university incubator firms benefit from their relationship with university incubators? The Journal of Technology Transfer 41, 205–219.
Lerner, J., Malmendier, U., 2013. With a little help from my (random) friends: success and failure in post-business school entrepreneurship. Review of Financial Studies 26, 2411–2452.
Lerner, J., Schoar, A., 2010. Introduction. In: Lerner, J. and Schoar, A. (Eds.), International Differences in Entrepreneurship. University of Chicago Press, pp. 1–13.
Mano, Y., Akoten, J., Yoshino, Y., Sonobe, T., 2014. Teaching KAIZEN to small business owners: an experiment in a metalworking cluster in Nairobi. Journal of the Japanese and International Economies 33, 25–42.
McKenzie, D., 2017. Identifying and spurring high-growth entrepreneurship: experimental evidence from a business plan competition. American Economic Review 107, 2278–2307.
McKenzie, D., Puerto, S., 2017. Business training for female microenterprise owners in Kenya grew their firms without harming their competitors. Unpublished working paper. The World Bank.
McKenzie, D.J., Sansone, D., 2017. Man vs. machine in predicting successful entrepreneurs: evidence from a business plan competition in Nigeria. Unpublished working paper. The World Bank.
McKenzie, D., Woodruff, C., 2008. Experimental evidence on returns to capital and access to finance in Mexico. The World Bank Economic Review 22, 457–482.
———, 2014. What are we learning from business training and entrepreneurship evaluations around the developing world? World Bank Research Observer 29, 48–82.
Megginson, W.L., Weiss, K.A., 1991. Venture capitalist certification in initial public offerings. The Journal of Finance 46, 879–903.
Picot, G., Dupuy, R., 1998. Job creation by company size class: The magnitude, concentration and persistence of job gains and losses in Canada. Small Business Economics 10, 117–139.
Puri, M., Zarutskie, R., 2012. On the life cycle dynamics of venture capital and non venture-capital financed firms. The Journal of Finance 67, 2247–2293.
Quinn, S., Woodruff, C., 2019. Experiments and entrepreneurship in developing countries. Annual Review of Economics 11, 225–248.
Roberts P., Lall S., Baird, R.B., Eastman, E., Davidson, A., Jacobson, A., 2016. What's working in startup acceleration: insights from fifteen village capital programs. Unpublished working paper. Emory University.
Rosenbaum, P.R., 2002. Observational Studies, 2nd Edition. Springer, New York.
Rosenbaum, P.R., 2010. Design of Observational Studies. Springer-Verlag, New York.
Schoar, A., 2010. The divide between subsistence and transformational entrepreneurship. In: Lerner, J., Stern, S. (Eds.), Innovation Policy and the Economy. MIT Press, Cambridge, pp. 57–81.
Schreyer, P., 2000. High-growth firms and employment. Unpublished working paper. OECD Science.
Schwartz, M., Göthner, M., 2009. A novel approach to incubator evaluations: the PROMETHEE outranking procedures. Unpublished working paper. Halle Institute for Economic Research.
Smith, S. W., Hannigan, T. J., 2015. Swinging for the fences: how do top accelerators impact the trajectories of new ventures? Paper presented at the Druid 2015 Conference on the Relevance of Innovation, Rome.

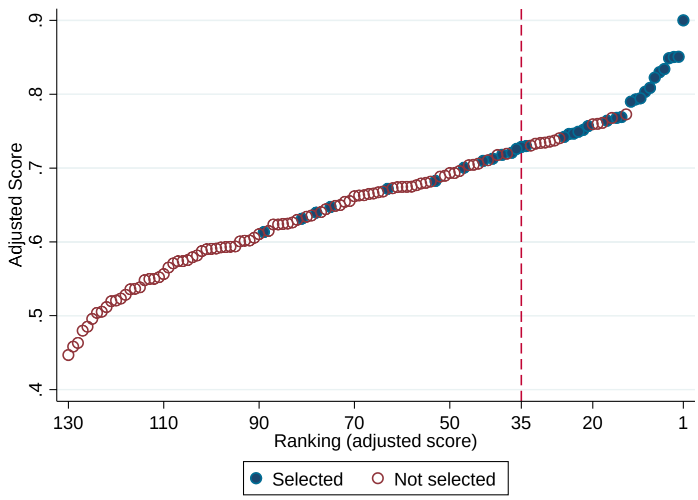
Smith, J., Todd, P., 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics 125, 305–353.
Stock, J., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Andrews, D.W.K. (Ed.), Identification and Inference for Econometric Models. Cambridge University Press, New York, pp. 80–108.
Sutton, J., 2012. Competing in Capabilities: The Globalization Process. Oxford University Press.
Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge.
Ubfal, D., Arraiz, I., Beuermann, D. W., Frese, M., Maffioli, A., Verch, D., 2019. The impact of soft-skills training for entrepreneurs in Jamaica. Unpublished working paper. IZA Institute of Labor Economics.
Valdivia, M., 2015. Business training plus for female entrepreneurship? Short and medium-term experimental evidence from Peru. Journal of Development Economics 113, 33–51.
Xie, Y., Brand, J., Jann, B., 2012. Estimating heterogeneous treatment effects with observational data. Sociological Methodology 42, 314–347.
Young, A., 2018. Consistency without inference: Instrumental variables in practical application. Unpublished working paper. London School of Economics.
Yu, S., 2019. How do accelerators impact the performance of high-technology ventures?

Figure 1. Distribution of Applicant Scores and Selection into the Accelerator

Panel A—Average Scores

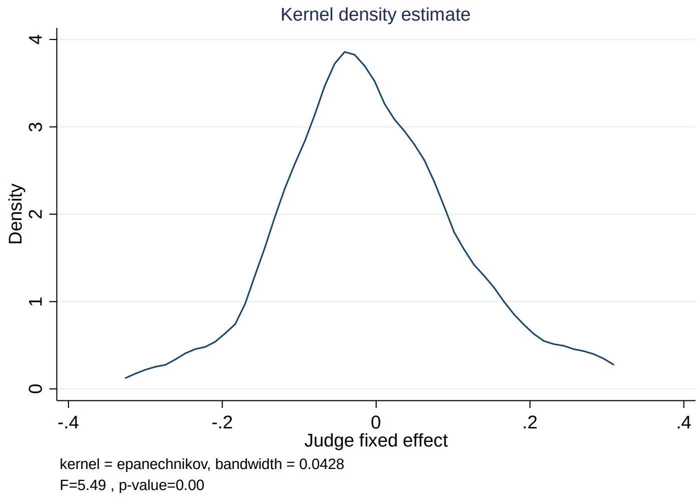


Panel B—Adjusted Scores



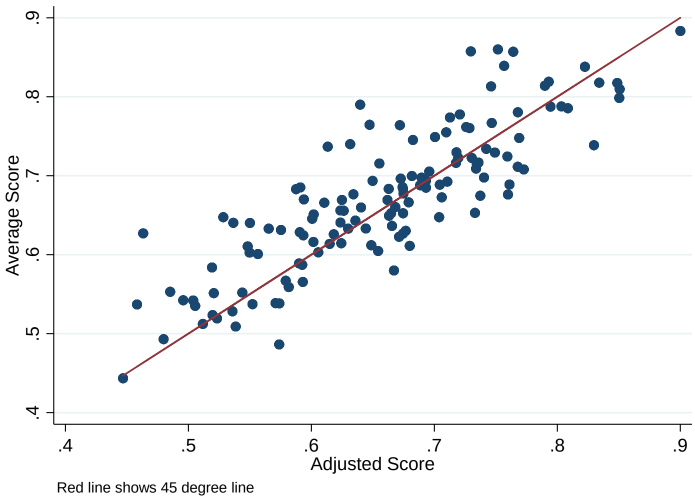
Panel A plots average scores against rankings based on the average score. Panel B plots adjusted scores, estimated as projects' fixed effects from Eq. (1), against rankings based on the adjusted score. In each panel, each dot represents an applicant; the solid (open) dots indicate the applicants that were (were not) selected into the accelerator.

Figure 2. Distribution of Judges' Fixed Effects



This figure plots the distribution of the estimated judge fixed effects from Eq. (1), which regresses project scores (by individual judges) against applicant fixed effects and judge fixed effects. Each project was evaluated by three randomly selected judges. Judges evaluated an average of eight projects. The table reports the statistics of an *F*-test showing that the judge fixed effects are jointly significant (*p*-value of 0.00).

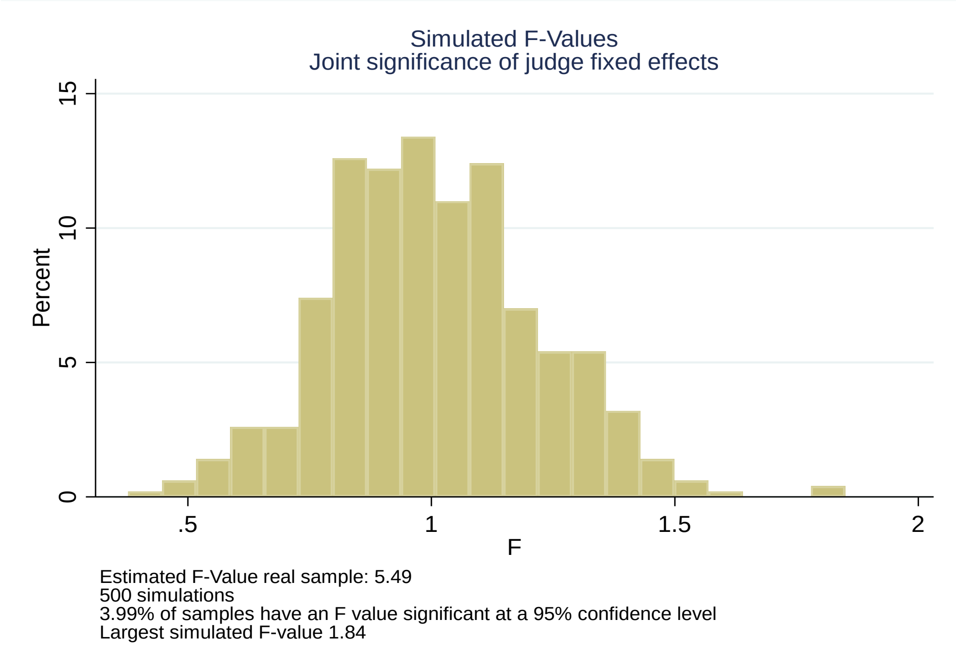
Figure 3. Average Scores and Adjusted Scores



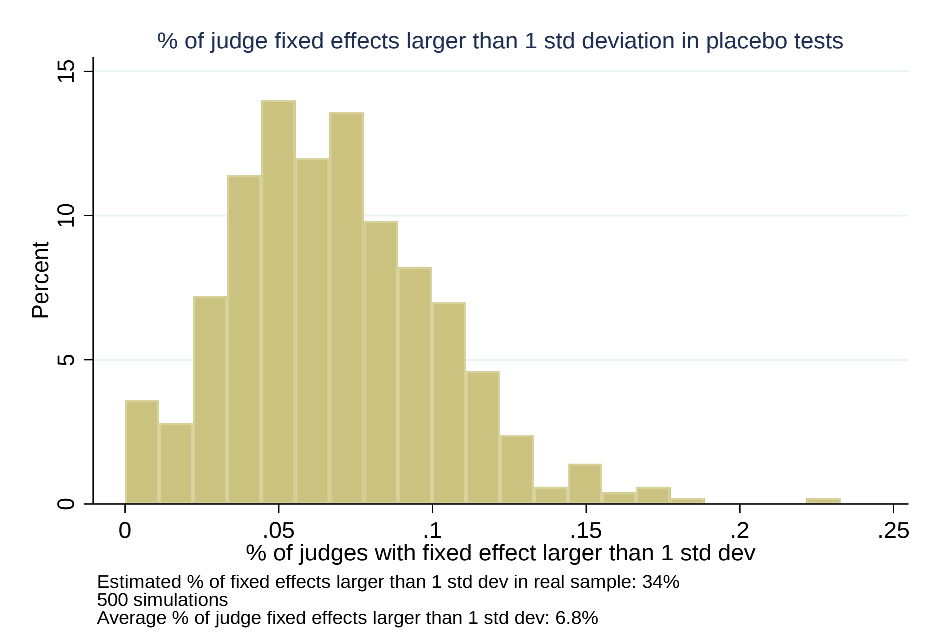
This figure plots average scores against adjusted scores. Each dot represents an applicant. The red line shows the 45-degree line. Applicants with adjusted scores above the 45-degree line were “lucky” in that they drew a generous judge panel, while applicants with average scores below the 45-degree line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

Figure 4. Placebo Assignment of Judges' Scores

Panel A—Distribution of *F*-values

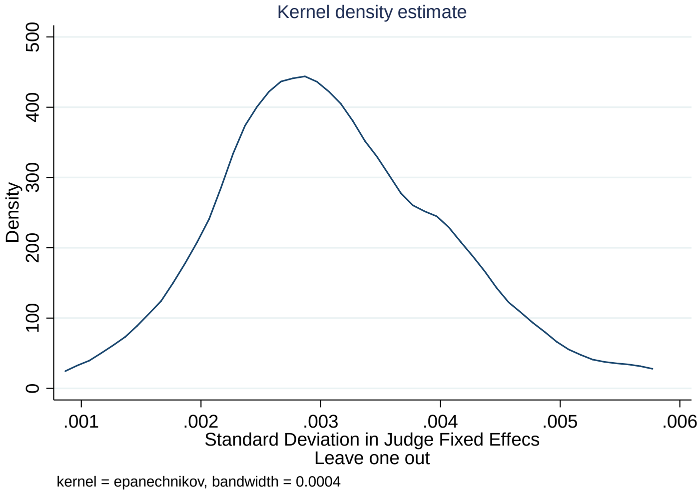


Panel B—Fixed Effects One Standard Deviation Above/Below Project Effect



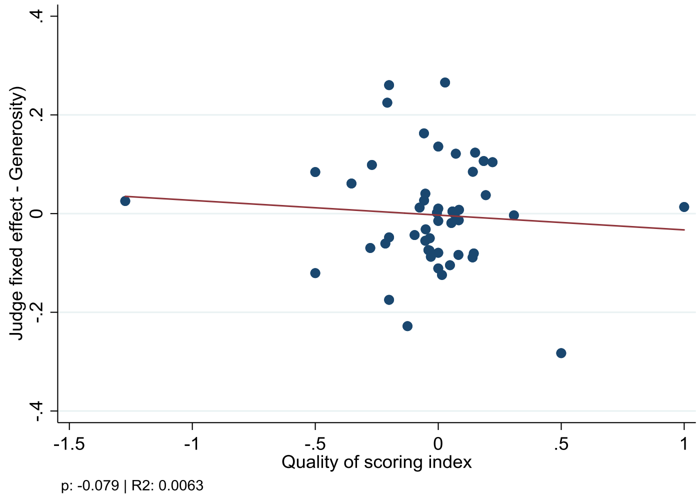
This figure plots the distribution of *F*-tests on the joint significance of the judge fixed effects in 500 placebo assignments.

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 135 judge fixed effects. We produce each estimate by sequentially leaving out of the sample one of the projects. Each judge has more leave-one-out fixed effect estimates than scored projects. This is because the estimated fixed effect of a given judge A from Eq. (1) varies as we leave out the projects she evaluated, but also as we leave out the projects of other co-judges that judge A did not also evaluate. By co-judges, we mean judges with whom judge A independently co-evaluated at least one project. The average standard deviation per judge of the leave-one-out fixed effects is 0.003 and the maximum is 0.006.

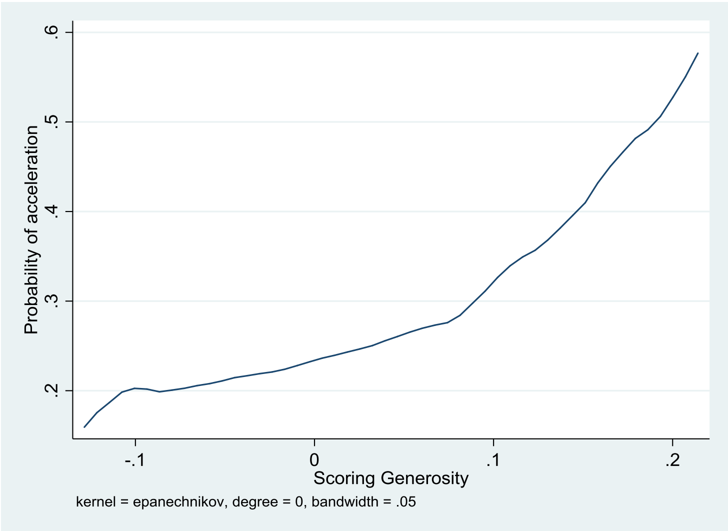
Figure 6. Scoring Generosity and Ranking Ability of Judges



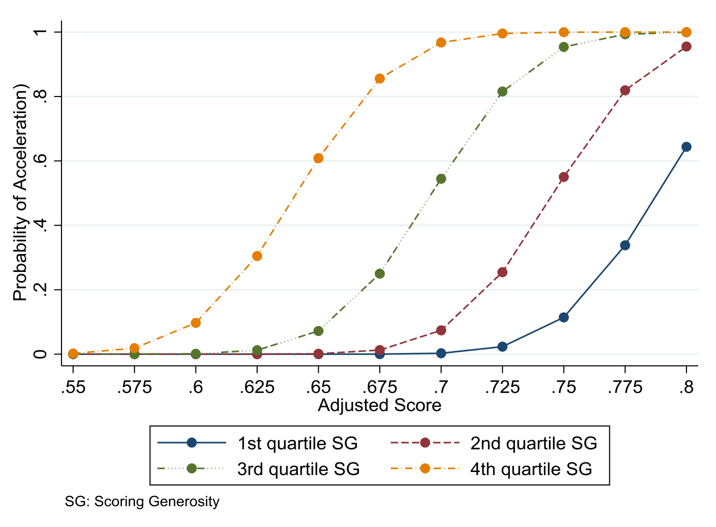
This plot is a scatter plot of judges' generosity and ranking ability. We measure judges' ranking ability using the correlation between a "judge's rank" and "actual rank." To produce this correlation, for every judge we rank the companies she evaluated based on (i) 2017 revenue ("actual ranking") and (ii) the judge's score ("judge's rank")

Figure 7. Acceleration Probability and Scoring Generosity

Panel A—Acceleration Probability and Generosity

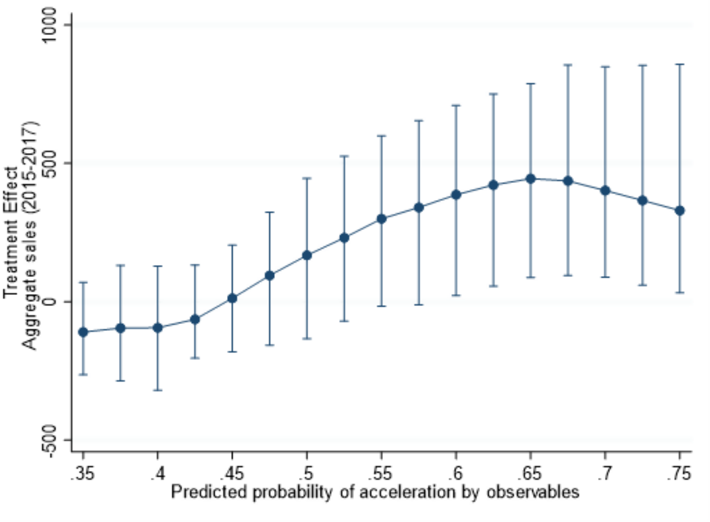


Panel B— Acceleration Probability and Generosity, by Quartiles of Adjusted Score



Panel A plots the probability of acceleration against adjusted score. Panel B 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.

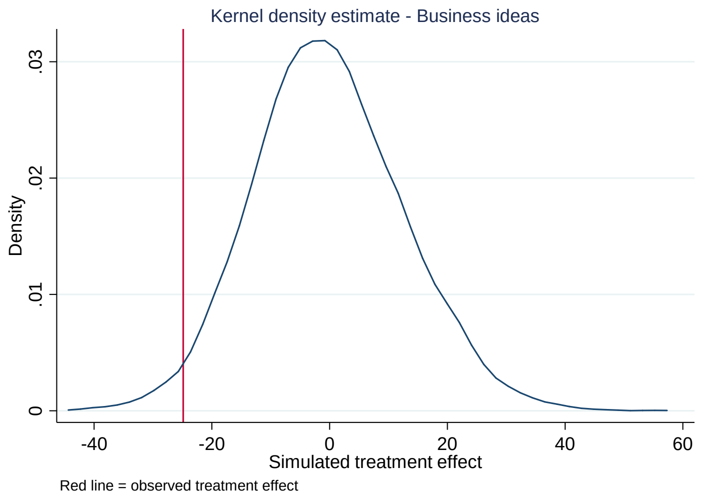
Figure 8. Heterogeneous Acceleration Effects



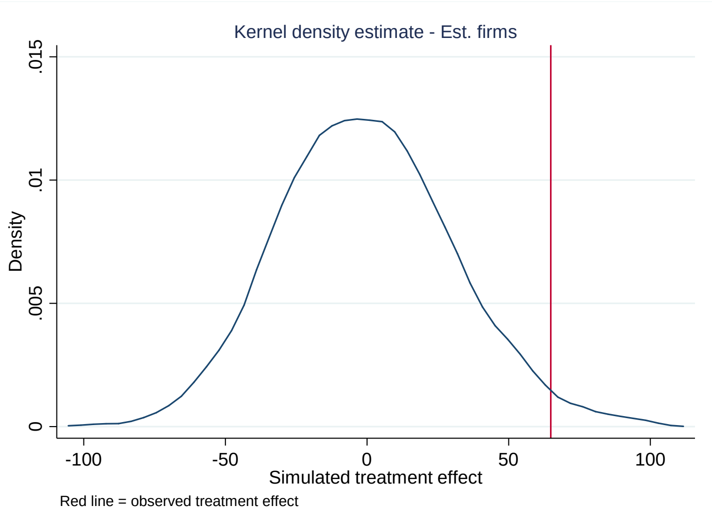
This figure plots the marginal effects of acceleration at different percentiles of applicants' acceleration propensity in the common support. For all other percentiles of the acceleration propensity score we cannot estimate marginal effects, as there are no selection mistakes to use in the estimation—i.e., no applicants with an acceleration propensity below (above) 0.35 (0.75) were mistakenly selected (rejected) by the program. We calculate standard errors using the standard deviation of the marginal effect estimates from a bootstrap procedure with 500 iterations.

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%
# of 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
Revenue 2013 (million pesos)	10.62	0	290	1.27	21.48
Revenue 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 Nonaccelerated 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
# of 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
Revenue 2013 (million pesos)	1.25	1.34	0.96	13.37	39.84	0.07
Revenue 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 nonaccelerated 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
Revenue 2013	135	10.62	37.22
Revenue 2014	135	25.80	56.14
Profits 2014	135	8.20	15.90
Total employees 2014	135	4.03	3.93
Revenue 2015	135	51.66	124.25
Profits 2015	104	14.04	44.42
Total employees 2015	104	5.35	6.90
Fundraising 2015	104	0.086	0.283
Revenue 2016	135	58.32	128.77
Profits 2016	92	16.65	39.86
Total employees 2016	92	4.68	4.65
Fundraising 2016	92	0.077	0.268
Revenue 2017	135	50.64	118.61
Profits 2017	86	8.73	21.62
Total employees 2017	86	4.09	4.88
Fundraising 2017	86	0.135	0.345

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, profits, and fundraising 2015–2017), and the Colombian business registry (revenues 2015–2017).

Table 4 Judges' Ranks and Project Growth Cuts

	(1) Actual Rank	(2) Actual Rank
Judge's Rank	0.295***	0.061***
	(0.0239)	(0.0231)
Judge Fixed Effect	No	Yes
Observations	405	405
R-squared	0.266	0.575

This table shows results from simple regressions of judges' ranks against projects' actual ranks. In column 2 we include judge fixed effects. To implement this regression, for every judge we rank the companies she evaluated based on (i) 2017 revenue ("actual rank") and (ii) the judge's score ("judge's rank"). The total number of observations is 405, as three different judges evaluated each of the 135 applicants. Standard errors are clustered at the applicant level and bootstrapped. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5. Applicant Scores and Project Growth

Panel A-Regression Results

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

Panel B—Shapley Owen Participation

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

The table presents results from estimating equation (2). 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 6. 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 acceleration				Yes
Pseudo R^2	0.0873	0.0881	0.184	
Observations	135	135	135	135

Panel B—Revenue Growth Rates

	Initial revenue (2014)	Final revenue (2017)	Growth from baseline	Implied annual growth
Gazelles	63.5	302.3	376%	68.24%
Nongazelles	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% of applicants according to 2017 revenue (and splitting the sample into business ideas and established firms). Panel B summarizes revenue growth rates across gazelles and nongazelles.

Table 7. Unconditional Probability of Acceleration and Scoring Generosity

	(1)	(2)	(3)	(4)	(5)
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 acceleration, 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) All	(2) Business Ideas	(3) 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 Eq. (4). The outcome variable is Acceleration × after. The variables acceleration 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

	All			Business ideas			Established firms		
	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) OLS	(6) IV	(7) OLS	(8) OLS	(9) IV
Acceleration × 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 Eq. (3). The outcome variable is revenue. The variables acceleration 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)
Acceleration × 2015	-0.259*
	(0.133)
Acceleration × 2016	0.297*
	(0.160)
Acceleration × 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 Acceleration 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—Comparison Impact Estimates Based on Different Methodologies

	(1)	(2)	(3)	(4)	(5)
	OLS	IV	PSM without Adjusted Score	PSM with Adjusted Score	RD
Treatment * After	42.91**	66.31**	40.73**	59.10**	81.38**
	(20.71)	(30.62)	(20.31)	(28.12)	(28.02)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	675	675	399	354	405
Number of ids	135	135	133	118	135

This table presents the results from different methodologies to estimate the accelerator impacts. For ease of exposition, Columns 1 and 2 replicate the OLS and IV estimates from Table 9. Column 3 estimates the effects using propensity score matching based on observables including initial revenue, firm's age, entrepreneur's age entrepreneur's sex, education, location, and sectorial and entrepreneurial experience. Column 4 estimates the effects using propensity score matching based on the same observables as Column 3, but also including the adjusted score in the matching procedure. Column 5 estimates the effects based on a discontinuity regression approach that exploits the 35th rank threshold on average score. The dependent variable is the change of revenue (each after year vs. 2014) variable. Standard errors are bootstrapped and clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 12. 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*
Revenue 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 the 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.