# Policy Papers

## Do Entrepreneurship Policies Work? Evidence from 460 Start-Up Program Competitions Across the Globe

Geoffrey Barrows

PP 2018-02

Suggested citation:

G. Barrows *(2018).* Do Entrepreneurship Policies Work? Evidence from 460 Start-Up Program Competitions Across the Globe. *FAERE Policy Paper, 2018-02.*

www.faere.fr

# Do Entrepreneurship Policies Work? Evidence From 460 Start-Up Program Competitions Across the Globe

Geoffrey Barrows*

December 4, 2017

**Abstract**

Many organizations around the world implement programs designed to encourage entrepreneurship, including grant prize awards, accelerator programs, incubators, etc. The goal of these programs is to supply entrepreneurs with early-stage support and visibility to help develop ideas and attract capital; but, if capital markets are efficient, good business ideas should find funding anyways. In this paper, I present evidence from the first global-scale, quasi-experimental study of whether entrepreneurship programs improve outcomes for start-up firms. I employ a regression discontinuity design to test whether winners of start-up program competitions perform better ex-post than losers, where the threshold rank for winning the competition provides exogenous variation in program participation. With 460 competitions across 113 countries and over 20,000 competing firms, I find that winning a competitions increases the probability of firm survival by 64%, the total amount of follow-on financing by $260,000 USD, and total employment by 47%, as well as other web-based metrics of firm success. Impacts are driven by medium-size prize competitions, and are precisely estimated both in countries where the costs of starting a business are low and where these costs are high. These results suggest that capital market frictions indeed prohibit start-up growth in many parts of the world.

*Keywords:* Start-ups, Entrepreneurship, Credit Constraints, Prizes, Accelerators
*JEL codes:* G24, L26, M13, O16

*CREST, CNRS, École Polytechnique, Université Paris-Saclay. Address: Ecole polytechnique, Departement d'Economie, 91128 Palaiseau Cedex, France. Email: geoffrey-masters.barrows@polytechnique.edu.
Special thanks to Torsten Kolind, Adele Burnes and the entire staff at YouNoodle. Emerson Ang, Tina Balacano, Kerim Bugdayci, Ruth Carter, Antonio Gonzalez, and Fredrik Madso provided excellent research assistance. All RAs were subject to legally binding confidentiality agreements, and all subsequent analysis was conducted on anonymized data to protect the privacy of individual start-ups. The paper benefited immensely from comments by Antoine Dechezleprêtre, Teevrat Garg, Mattieu Glachant, Hélène Ollivier, Gregor Singer, and seminar participants at EAERE Athens, AERE Pittsburgh, The World Bank, and the 6th FAERE Thematic Workshop on Innovation in Paris. I gratefully acknowledge support from the Chair for Energy and Prosperity at École Polytechnique.

# 1 Introduction

Entrepreneurship is widely considered instrumental both for lifting individuals out of poverty and for stimulating aggregate growth. As a result, many governments, NGOs, and private institutions around the world implement programs designed to encourage entrepreneurship, including grant prize awards, accelerator programs, and incubators.[1] The rationale for such programs is straightforward. First, they supply entrepreneurs with the necessary early-stage funding and support to develop their ideas when traditional forms of capital are difficult to secure. Second, the signal of having won a prize or participated in a prestigious incubator program helps start-ups secure capital from traditional sources down the line, which in turn helps the firms grow. While these arguments imply that program participation should correlate with firm success, it is not clear that entrepreneurship programs *causally* impact firm outcomes. If capital markets work efficiently, then good business ideas should find funding anyways.[2]

In this paper, I estimate the causal impact of entrepreneurship programs on start-up firm success from a global-scale quasi-natural experiment across multiple industries, countries, and program types. To isolate the causal impact, I exploit a discontinuity in the probability of program entry based on the selection process. Grant prizes and accelerator and incubator participation are often awarded as the result of competitions among potential recipients. Donor organizations/program administrators[3] either evaluate competition entrants themselves or hire industry experts to do so, generating an observable ranked order of entrant "quality" for each competition. Availability of funding dictates the number of prizes, which are usually only awarded to the highest ranked firms in the competition. This structure generates a discontinuity in program participation around a

[1] The practice of offering cash prizes for innovation has a long history, going back at least to the Spanish Longitudinal Prize of 1567, in which the king of Spain offered 6000 gold ducats + 2000 ducats/year for life for the first person to discover a method finding longitude at sea (Masters & Delbecq, 2008). A recent McKinsey report estimated the annual aggregate value of prize awards today at $1-2 billion USD globally (Bays et al., 2009).

[2] There is a large literature connecting entrepreneurship with capital or liquidity constraints. See Kerr & Nanda (2015) for a recent survey. Relevant mechanisms that could generate frictions in the private capital market, including gender bias (Fay & Williams, 1993), in-group bias (Beck et al., 2011; Fisman et al., 2017), racial discrimination (Hanson et al., 2016; Blanchflower et al., 2003), and moral hazard, among others.

[3] Winning a competition may either result in a pure grant prize, training, or some other kind of non-monetary benefit. I will usually refer to all three possibilities simply as a "prize." In the case that a grant is awarded, then the competition organizer can best be thought of as a "donor organization", while when the prize involves training, the organization can more accurately be considered a program administrator. I will usually refer to the organization simply as a "donor."

pre-determined threshold, such that firms that just barely beat the threshold are more likely to receive a prize, while firms that barely miss the threshold likely will not receive the prize.[4]

The dataset for the study is based on confidential administrative records from the Internet platform website YouNoodle Inc. (YN), which organizes competitions on behalf of donor organizations. For each competition, YN collects submissions, which are then routed to one or multiple judges to evaluate along pre-determined metrics. YN then aggregates the judge scores, computes rankings, and submits the rankings back to the donor organization, who then makes the final selection. From these judge data, I recreate the rank order of firm placements from 460 competitions organized by YN between 2010 and 2015 in many different countries around the world, with firms operating in a variety of industries. Competition winners are identified along with the prize values from donor organization publications. Based on these data, I infer the threshold used for selecting the winners, and then construct the running variable as ordinal distance between each firm's rank and the threshold.[5]

To estimate returns to program participation, I employ two sets of outcome metrics. The first set of metrics are based on the subjective evaluations of firms' websites and general web presence, collected manually by a team of research assistants in 2015 and 2016. These metrics are based on the idea that internet activity for modern start-ups is critical to success, and hence should constitute an informative (yet noisy) signal of firm success. The second set of outcomes metrics come from the aggregator site Mattermark, which reports information such as number of employees, funding, unique web visitors, facebook likes for many start-ups (though not traditional performance metrics like revenue, profits, or patents, for reasons that I'll discuss below). The Mattermark data rely on more standard economic outcomes (employees, funding), but the coverage is smaller than the first set of metrics. In fact, I can only match about 35% of firms in the YN dataset to Mattermark. Hence, I will use the Mattermark data to check the quality of the signal in the first set of metrics (which I will call "subjective metrics"), and to benchmark magnitudes, but the analysis mostly relies on the former metrics.

The analysis proceeds in two steps. First, I perform a standard Regression Discontinuity

[4] The selection decision ultimately rests with the donor organization, which means that donors can disregard the quality rankings and select any firm they choose. Also, competition entrants can refuse a prize. Thus, beating the critical threshold does not guarantee program participation, but it does generate a discrete jump in the probability of participation.

[5] All RAs were subject to YN's legally binding confidentiality agreements, and all subsequent analysis was conducted on anonymized data to protect the privacy of individual start-ups.

(RD) evaluation of the impact of entrepreneurship programs on the subjective metrics. Here, with a sample of 7,883 firms from 460 competitions, I find that winning a competition increases the chance that a firm is still in operation 2-5 years later by 64%, with similar increases in both the web score and the general web presence scores. These results are statistically significant, and robust to varying the bandwidth around the threshold and the set of controls. Additionally, placebo tests in which the threshold is counterfactually located at different points in the rank order yield small and statistically insignificant results.

A secondary result is that when donor organizations deviate from the YN rankings, they systematically choose lower quality firms. I can infer this by comparing the OLS correlations between wining a competition with future success, and the IV results based on the RD. Since donor organizations do not choose all and only firms that beat the YN threshold, the OLS estimate of the impact of winning includes the selection bias stemming from the donors' decision to deviate from the rankings. By contrast, since placement on either side of the threshold is random within an appropriate bandwidth, the IV estimates represent just the causal impact of winning the competition. If the OLS estimate is larger (more positive) than the IV, then one could conclude that selection bias is positive. However, since I find that the IV estimate is significantly higher than the OLS estimate, I conclude that the selection effect is in fact negative. The donors would have had greater positive impact if they had stuck to the YN rankings.

Next, I turn to the Mattermark data – the "objective metrics" – to assess economic significance. I first show that the subjective metrics are highly correlated with the objective metrics, which I take as evidence that the subjective metrics are a reasonable proxy for firm success. Based on these correlations, I project out of sample what the Mattermark scores would be for YN firms missing from that database. With these projected Mattermark measures, I again perform the RD analysis. Here, I find that winning a competition generates 55% more follow-on funding for the firm, along with similar increases in employees, LinkedIn connections, Facebook likes, and Twitter followers. In absolute terms, the average funding received by a firm to the right of the threshold is $470,000 USD. A treatment impact of 55% implies an increase of $260,000 from winning a competition, which can be compared to the average winning grant value of about $26,000.

Finally, I test for heterogeneous impacts by firms' country of origin, competition type, donor organization type, gender of entrepreneur, and dollar value of competition prize. I find that positive impacts are driven by mid-size prize competitions and competitions organized by NGOs. I find impacts precisely estimated both in countries where the costs of starting a business are low and where these costs are high. I also estimate statistically significant impacts both for pure grant prizes as well as accelerator and incubator programs. These results suggest that capital market frictions indeed prohibit start-up growth in many parts of the world, and that small-scale program intervention can encourage entrepreneurship.

The paper contributes to a burgeoning literature on entrepreneurship that also exploits discontinuities in program participation based on competition structures. McKenzie (2017) finds that grant prizes of $50,000 for entrepreneurs in Nigeria lead to substantial increases in firm survival and employment. Fafchamps & Quinn (2016) finds substantial impacts from small grants ($1,000 USD in their case) focusing on African countries exclusively. Klinger & Schündeln (2011) find that training increases firm-level outcomes in a sample of Central American entrepreneurs who already employ on average 10 workers at the time of application, while Gonzalez-Uribe & Leatherbee (2017) finds no impact from basic accelerator services in Chile. Smith & Viceisza (2017) find impacts on survival rates for entrepreneurs competing in the ABC televised competitions "Shark Tank." These entrepreneurs represent mostly American firms, and as in Klinger & Schündeln (2011), they are observed at a later stage in their growth trajectories than the firms competing in YN competitions. Finally, Howell (2017b) also finds evidence of program effects from competitions in the US. The key difference between my study and previous work is that I use a dataset with global coverage (or at least, a dataset that represents a great majority of countries in the world). This wide coverage enables me to control for country and industry specific effects and to estimate heterogeneous effects based on development and capital market efficiency. Additionally, the start-ups in my dataset tend to be younger even then those in other start-up research (e.g., not even incorporated yet), which are more likely to be constrained in their access to credit.

The paper also contributes to a broader literature on alternative financing. Early work by Kortum & Lerner (2000) and Lelarge et al. (2010) estimate the impact of venture capital and government grants, respectively, on firm success. Yu (2016) studies the impact of prominent accelerator programs in the US and finds that accelerators resolve uncertainty about profitability of investment ideas. In a closely related paper, Lerner et al. (2017) estimate by regression discontinuity the returns to angel investing across different markets, finding that, in general, angel investments increase the probability of firm survival and the quantity of future funding received by the firms. Finally, Howell (2017a) and Dechezleprêtre et al. (2016) study US government grants for clean technology innovation and UK subsidies for R&D innovation in general, and find that government policies tend to increase patent activity. The present study complements Lerner et al. (2017) and Howell (2017a) by demonstrating impacts from smaller programs and for younger firms.[6]

# 2 Data

In this section, I present the dataset, including the different metrics that measure start-up success. I also present evidence in support of a strong positive correlation between the subjective and objective metrics.

## 2.1 Competitions

The dataset for this paper is based on confidential administrative records from the Internet platform YouNoodle Inc. (YN), which organizes competitions for donor organizations. Donor organizations define the parameters of the competition – number and value of prizes, submission requirements, judge criteria, etc – while YN posts the call, collects submissions, aggregates judge scores, computes rankings, and submits the rankings back to the donor organization. Donor organizations then award prizes or program acceptance based on the YN rankings, though the donor is free to select different firms than the ones recommended by the YN rankings (which happens quite often).[7] Given that the number of prizes is dictated by available funds, I can take the threshold rank for winning a prize if the donors *did* follow the YN recommendations as exogenous. Additionally, since YN ranks are determined by impartial judges, manipulation around the threshold is unlikely.[8] Thus, the structure of the competitions delivers a fuzzy RD design, with placement on either side of the threshold cutoff (within an appropriate bandwidth) as good as randomized.

Two key features of the dataset bear mention. First, the number of prizes varies substantially by competition. Second, some competitions feature multiple rounds, wherein YN does not always judge the final round of the competition. Together, these points imply

[6] For example, in Lerner et al. (2017) and Howell (2017a) the average age of firms are 8 and 9 years respectively, where as firms in our dataset are often not even a year old.

[7] Out of the 460 competitions, 59% offer just a grant prize award, 12% offer accelerator or incubator support, and the rest offers some other type of awards, such as travel allowances for a conference or pure recognition.

[8] And even if the judges had some motivation to manipulate a given firm's evaluation, competitions are usually evaluated by many judges, so it would be difficult for any one judge to know exactly what score would yield a winning rank.

that the threshold will vary by competition (see Appendix Table A.3), and that the candidates whose ranks are higher than the threshold may not end up winning any prize. This heterogeneity could pose a challenge for the estimation: a firm that won the only prize in a 1-prize competition might be of significantly higher quality than a firm that won the 10th prize in a 10-prize competition. I develop a novel re-scaling strategy to address this empirical issue in Section 3. Appendix A describes the YN dataset in detail, along with the process for computing competition thresholds and normalized ranks.

The distribution of firms by rank is illustrated in Figure 1. Here, I have normalized the threshold to 0; i.e., the last firm to beat the threshold is assigned rank = -1, the first firm that just barely missed the cut-off is assigned rank = 1, the firm that beat the threshold by 2 ranks has normalized rank of -2, etc. In Panel A, I plot the distribution of all 20,594 entrants across all 460 competitions in the dataset (after cleaning), and find that the distribution is skewed to the right of the threshold – there are a lot more firms that missed the cutoff than ones that beat the cutoff.[9] Since the RD analysis will only focus on firms around the cutoff, I only collected outcome data for a subset of the total population (7,649 firms).[10] Focusing on just this sample in Panel B, I find that the distribution of firms is roughly symmetric around the cut-off.

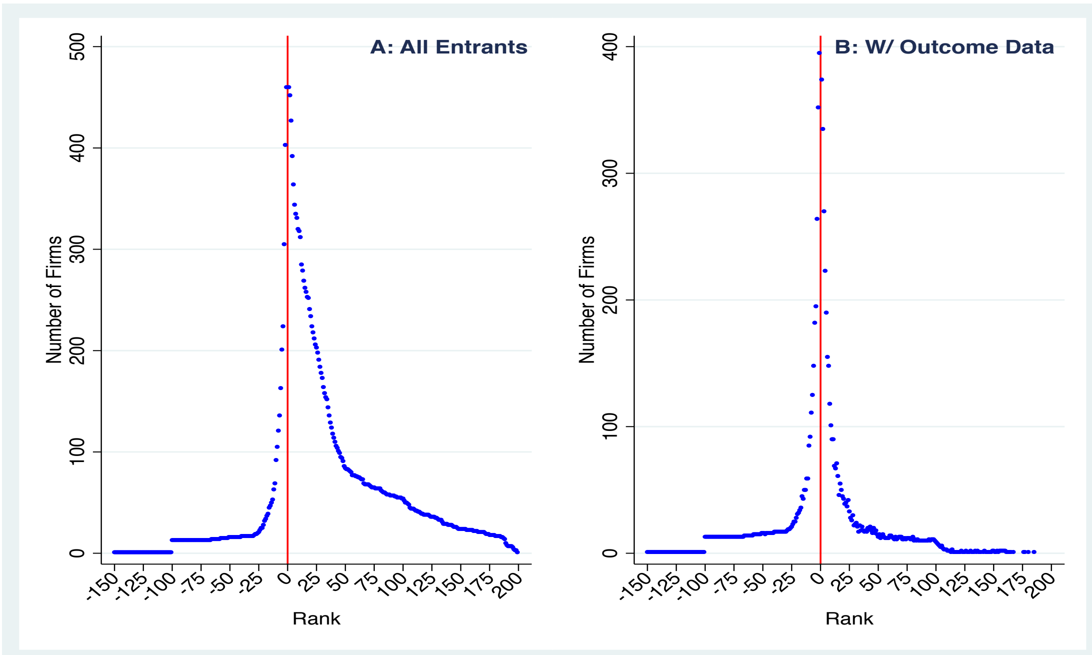
## 2.2 Outcome Metrics

A significant challenge to conducting empirical research on start-up firms is the dearth of available "success" metrics. What does it mean to be a successful start-up? Start-ups are usually too young or too small to show up in traditional firm surveys, like the census of manufacturers. And, even if they were to appear in these surveys, it usually takes many years for start-ups to post profits, or even generate revenue, so traditional balance sheet data would likely be uninformative. Additionally, these surveys would likely suffer from selection bias. For instance, patent data is a popular way to measure success in Research and Development, but start-ups are usually too early in their development to apply for patents. And patenting is notoriously sparse, which means one would need to observe a very large number of competition entrants to have enough statistical power to estimate

[9] The total number of entrants in Figure 1 is lower than the number quoted in Table A.3 because it was not always possible to match some firms identified as "winners" from competition publications to a specific firm name in the YN dataset (234 firms). These firms are included in Table A.3 as having participated in the competition, but cannot be assigned a YN rank, since they are not in the YN database.

[10] As in the previous footnote, this number is lower than the corresponding figure quoted in Table A.3 because of winners that have no official YN rank (234 firms).

Figure 1: Distribution of Firms by Rank




Notes: Panel A presents data from all competition entrants (20,594 entrants across 460 competitions), while Panel B restricts to observations with outcome data (7,649 entrants across 460 competitions).

effects.[11]

For this paper, I consulted industry experts at YN to design a novel set of start-up "success" metrics that can be easily evaluated for every firm in the database. These metrics are based on the simple idea that internet activity for modern start-ups is critical to success. In today's economy, almost all firms have a home website where they describe their products, list contact information, report recent activity, etc. Quality of web design and ease of use are critical for attracting customers and financing. Additionally, since many products today are web-based, the web page is in many cases the product itself of the start-up. Hence, the quality of the web page should yield an informative, yet noisy, signal of firm success. This is a fundamental assumption of the paper; however, with auxiliary "objective" metrics from Mattermark, I will present evidence that the subjective metrics track traditional measures of firm performance quite well.

[11] If one observed firms many years after they won an entrepreneurship prize, it is possible that patenting activity would be a relevant outcome metric. With data coming only from the first half of the 2010s, it is too early to look for patenting effects in this population of firms. Though it might be a promising avenue for future research.

Data collection worked as follows. In 2015 and 2016, several research assistants (RAs) were hired to search for competition entrants on the Internet. Each RA was given a subset of the firms in the YN database to search for. RAs knew everything about the entrants that YN knew from the competition application – name of firm, name of entrepreneur, maybe the country of origin, perhaps some information about the product or industry. RAs spent a few minutes searching for each firm. On average, an RA was able to find a live web page for a firm within 5 minutes that could plausibly be identified as the same firm that entered the competition (understanding that name or product might have changed since the time of the competition). If the RA could not find a live homepage for the firm within about 20 minutes, then the firm was deemed "not alive."[12] This binary variable taking the value 1 for alive and 0 otherwise serves as the first subjective metric.[13]

Thanks to the training offered by YN experts, the RAs were also able to create a score for the quality of the web page, as well as for the general web presence of each firm. Conditional on finding a live homepage for the firm, the RA rendered a subjective rating of the web page from 0 to 5 (where 0 corresponds to no live homepage). This evaluation was based on a variety of dimensions, including ease of navigation, attractiveness of design, evidence of recent posting, and usefulness of information. Obviously, subjective ratings might vary between RAs, but this variation can be absorbed by fixed effects in the regression analysis. Lastly, RAs searched for "general" web presence, beyond the home page, and again assigned a rating of 0-5 ( where 0 corresponds to no general web presence). This general score is based on links to other web sites, news clippings, twitter feeds, etc. The general score captures more the general activity or "buzz" around a start-up. This measure should correlate with the homepage ("web") score, though not necessarily perfectly.

Descriptive statistics are reported for competition winners vs. losers in Table 1. The researchers were able to find live website for 62% of winning firms (column 1), compared to just 47% of losing firms (column 5). Column 9 performs a t-tests for equality of means and rejects the null of no difference for the "alive" variable at the 1% level. Winning firms had

[12]A possible explanation for nonexistent homepages could be that the firm were successfully bought-out, which is something I cannot identify in the data. If winning a competition has positive causal impacts on firm success, then there should be more of these "buy-out" cases above the threshold than below the threshold. I.e., the alive metric should be systematically biased downwards for winning firms, which means that if I find positive and statistically significant effects on the alive metric, it is despite the possibility of buy-outs.

[13]One might argue that the "alive" metric is closer to an objective measure – the firm is either alive or it is not. However, it can rarely be determined with 100% confidence that a firm is actually not still in operation. Rather, a firm is judged "not alive" when the researcher determines that he or she has searched sufficiently long enough that if the firm were alive, he or she would have found evidence of its existence. Thus, the determination that a firm "is not alive" depends on a subjective call from the researcher.

an average web score of 3.74, compared to 3.65 for losing websites, but these are conditional on finding a live web site. In the regressions below, missing web scores will be coded as 0. Finally, winning firms together have higher average general scores (2.38), compared to losing firms (1.81), difference in mean significant at 1% level.

In Table 1, I also report descriptive statistics for the objective metrics. First, researchers were able to find winning firms in the Mattermark database 44% of the time, compared to 28% of losing firms. From Mattermark, we learn that firms in the sample had on average 14 employees, though the distribution is highly skewed. Only 45% of firms had more than 6 employees, and 35% had 3 employees or less. The average firm had 37,000 unique web visits, 11,000 facebook likes, 186 LinkedIn connections, 2,900 twitter followers, and generated 3.5 million dollars US in total funding. The distribution of most of the variables are also highly skewed. For example, median funding was only $125,000. Additionally, while the winning average is surprisingly lower than the losing average for all these objective values, the difference is rarely statistically significant, and never significant at even the 5% level.

In the last four panels of Table 1, I report firm-level covariates by winning vs losing firms. For each firm, I observe the home country of the entrepreneur, the industry of the firm, the product type, and two variables regarding gender - a binary indicator for whether the lead entrepreneur is female, and another indicator for whether there is a female on the team at all. These data will be helpful for identifying heterogeneous treatment effects and for checking balance across the threshold. The sample is heavily skewed towards North America, South America, and Europe, but there are some firms from Asia and Africa as well. Firms operate in a variety of industries – though the distribution is probably skewed towards high tech areas relative to the underlying industrial composition of most developed countries. And finally, while male-headed firms and all-male teams dominate the dataset, we see a nontrivial mass of female-led and female-on-team firms (21% and 29%, respectively).

## 2.3 Comparing the subjective and objective metrics

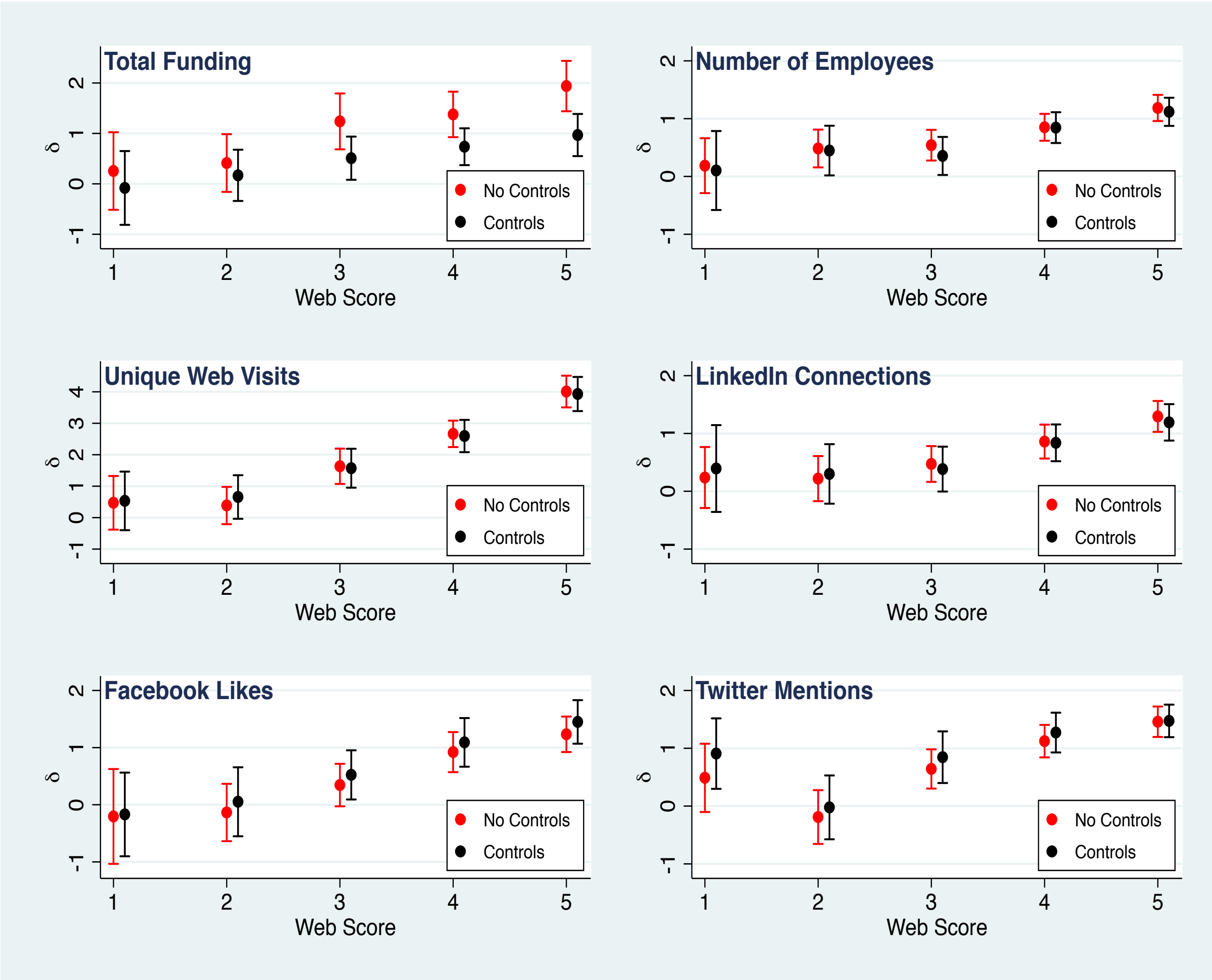
The analysis of the paper depends on the assumption that subjective evaluations of home web pages by RAs proxy for firm quality. To test for that, I compare the subjective metric scores to the objective metrics from Mattermark. As I mentioned above, I only find 35% of YN firms in the Mattermark dataset, which offers a large enough sample to assess the correlations.

Table 1: Firm-level Descriptive Statistics

| | Winners | | | | Losers | | | | Difference |
|---|---|---|---|---|---|---|---|---|---|
| | Mean (1) | Min (2) | Max (3) | Obs (4) | Mean (5) | Min (6) | Max (7) | Obs (8) | (9) |
| *Subjective Metrics* | | | | | | | | | |
| Alive (0,1) | 0.62 | 0 | 1 | 2870 | 0.47 | 0 | 1 | 5013 | *** |
| Web | 3.74 | 1 | 5 | 1771 | 3.65 | 1 | 5 | 2335 | ** |
| General | 2.38 | 0 | 5 | 2870 | 1.81 | 0 | 5 | 5013 | *** |
| *Objective Metrics* | | | | | | | | | |
| Mmark (0,1) | 0.44 | 0 | 1 | 2870 | 0.28 | 0 | 1 | 5013 | *** |
| Employees | 12.65 | 1 | 345 | 532 | 16.23 | 1 | 611 | 709 | * |
| Uniques (Ths) | 32.52 | 0.001 | 3241 | 870 | 38.59 | 0.001 | 2235 | 1076 | |
| FB Likes (Ths) | 9.19 | 0.001 | 590 | 699 | 13.09 | 0.002 | 1271 | 940 | |
| Funding (mil) | 2.59 | 0.011 | 204 | 370 | 4.34 | 0.005 | 149 | 322 | * |
| Twitter (Ths) | 2.64 | 0.001 | 250 | 782 | 3.11 | 0.001 | 411 | 1054 | |
| Linked-in (Ths) | 0.15 | 0.001 | 10 | 573 | 0.21 | 0.001 | 13 | 775 | * |
| *Firm Location* | | | | | | | | | |
| NAmerica | 0.35 | 0 | 1 | 2870 | 0.36 | 0 | 1 | 5013 | |
| SAmerica | 0.29 | 0 | 1 | 2870 | 0.24 | 0 | 1 | 5013 | *** |
| Europe | 0.24 | 0 | 1 | 2870 | 0.24 | 0 | 1 | 5013 | |
| Asia | 0.10 | 0 | 1 | 2870 | 0.11 | 0 | 1 | 5013 | |
| Africa | 0.02 | 0 | 1 | 2870 | 0.03 | 0 | 1 | 5013 | |
| Unknown | 0.01 | 0 | 1 | 2870 | 0.02 | 0 | 1 | 5013 | *** |
| *Industry* | | | | | | | | | |
| Clean Tech | 0.11 | 0 | 1 | 2870 | 0.11 | 0 | 1 | 5013 | |
| Services | 0.31 | 0 | 1 | 2870 | 0.30 | 0 | 1 | 5013 | |
| Products | 0.19 | 0 | 1 | 2870 | 0.20 | 0 | 1 | 5013 | * |
| Web | 0.19 | 0 | 1 | 2870 | 0.18 | 0 | 1 | 5013 | ** |
| Life Sciences | 0.12 | 0 | 1 | 2870 | 0.12 | 0 | 1 | 5013 | |
| *Product Type* | | | | | | | | | |
| Services | 0.25 | 0 | 1 | 2870 | 0.25 | 0 | 1 | 5013 | |
| Products | 0.24 | 0 | 1 | 2870 | 0.23 | 0 | 1 | 5013 | |
| Software/Apps | 0.41 | 0 | 1 | 2870 | 0.35 | 0 | 1 | 5013 | *** |
| *Gender Variables* | | | | | | | | | |
| Female Lead | 0.20 | 0 | 1 | 1901 | 0.22 | 0 | 1 | 3600 | |
| Female on Team | 0.28 | 0 | 1 | 1947 | 0.29 | 0 | 1 | 3680 | |

Notes: Column 9 tests for equality of mean value for winners vs losers. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Figure 2: Subjective Metrics vs Objective Metrics




Notes: Log dependent variable for each subfigure is listed in top left corner. Specifications without controls are reported in red, while estimates in black control for country of origin, industry, product-type, and competition. 95% confidence intervals depicted with vertical bars. Web score = 0 serves as omitted category. All standard errors are clustered to allow for arbitrary correlation within application group.

In Figure 2, I present nonparametric point estimates from regressions of the form

$$y_{ic} = \alpha + \sum_{j\in[0,5]} \delta_j * [\mathbb{1}|web_{ic} = j] + X_{ic}\Gamma + \epsilon_{ic} \qquad (1)$$

where $y_{ic}$ is the log of one of 6 Mattermark variables listed in Table 1, $X_{ic}$ is a set of firm controls including dummy variables for country of origin, industry, product-type, and competition, and $\epsilon_{ic}$ is an error term. The coefficients of interest are the $\delta_j$s, the coefficients on the indicator variables for firm $i$ receiving a web score of $j \in [0, 5]$. If the subjective metrics capture some signal about firm quality, one would expect these $\delta_j$s to be increasing in $j$.

Each subfigure in Figure 2 presents estimates from two separate regressions, each taking the indicated Mattermark variable as the dependent variable. Raw correlations (with no control) are depicted in red, and the estimated $\delta_j$s controlling for the full suite of fixed effects are reported in black. In each case, 95% confidence intervals are plotted with vertical bars. Web score = 0 is the omitted category so all estimates are relative to firms for which that the RAs could not find a live web page. Standard errors are clustered on the application group.[14]

In Figure 2, I find that the subjective metrics lineup extremely well with the objective metrics from Mattermark.[15] In almost every case, point estimates are positive, increasing in $j$, and mostly statistically significant. Hence, when the RAs score the home webpage highly, Mattermark reports that the firm has more funding, employees, unique web visits, etc. In terms of magnitudes, the point estimates in Figure 2 are in log points, which indicates economically significant differences. For example, a firm that scored 5 in terms of website earned $e^{0.96} = 2.61$ times more funding than a firm with web score==0, hired $e^{1.11} = 3.02$ times more employees, and generated $e^{3.93} = 50$ times more unique web visits. While it would be useful to have data on revenue, profits or patents, the fact that the subjective scores from the RAs correlate positively with all available metrics provides a reassuring check that subjective measures proxy start-up success.

# 3 Empirical Strategy

With the administrative records from YN and the subjective and objective outcomes, I test for causal impacts of grant prize funding by estimating the standard RD equation

$$y_{ic} = \alpha + \tau * [\mathbb{1}|Rank_{ic} < 0] + f(Rank_{ic}) + X_{ic}\Gamma + \delta_c + \epsilon_{ic} \quad , \qquad \text{with} \quad -r < Rank_{ic} < r \tag{2}$$

[14]See Appendix A for definition and discussion of "application groups." Essentially, these are super competitions comprising many individual competitions judges by YN at the same time.

[15]I only present results from the web score. Below, when I project out of sample, I will use the general score as well. Correlations look quite similar.

where $y_{ic}$ is outcome for firm $i$ after entering competition $c$, $Rank_{ic}$ is the normalized rank assigned by YN to firm $i$ entering competition $c$, $[\mathbb{1}|Rank_{ic} < 0]$ is an indicator for firm $i$ beating the threshold (i.e., $Rank_{ic} < 0$), $X_{ic}$ is a vector of firm-level controls that might include country fixed effects, industry fixed effects, gender fixed effects, and product-type fixed effects, $f(Rank_{ic})$ is a polynomial control for normalized rank, $\delta_c$ is a competition fixed effect, $r$ is the bandwidth, and $\epsilon_{ic}$ is the idiosyncratic error term.

Conditional on the assumption that placement on either side of the threshold is exogenous, $\tau$ can be interpreted as a causal impact. However, as mentioned above, beating the threshold does not guarantee selection for an award. Donor organizations often disregard the YN rankings and award prizes to firms that failed to beat the threshold. The decision to deviate from the YN rankings is based on unobservable information that is potentially endogenous to firm quality. Hence, I employ the fuzzy RD design and instrument winning the competition (receiving any award) with beating the YN threshold.

As mentioned above, a complication arises from the heterogeneity in the (un-normalized) rank of the threshold cut-off. When a small number of prizes are awarded (hence, the threshold cut-off is small), winning firms beat comparatively more firms than when a large number of prizes are awarded. This could lead to heterogeneous effects based on the (un-normalized) rank of the threshold. This problem was also noted by Howell (2017a), whose solution was to control for quintile of rank of firm within the competition. This is a reasonable approach when the heterogeneity in number of prizes is fairly small (as in the case of Howell (2017a)); however, since I find significant heterogeneity in the number of prizes awarded, I propose a novel re-scaling of the data to account for difference in competition structure.

To address this issue, I re-scale ranks so that the difference between firms on either side of the threshold are comparable across competitions. Formally, I define

$$\begin{aligned} Rank^s_{ic} &= \frac{Rank_{ic} - T_c + 1}{T_c} \quad \text{for} \quad Rank_{ic} < 0 \\ &= Rank_{ic} \qquad\qquad\quad \text{for} \quad Rank_{ic} > 0 \end{aligned} \tag{3}$$

where $Rank_{ic}$ is the ordinal distance from the threshold cut-off, and $T_c$ is the number of awarded prizes in competition $c$. Equation (4) effectively spreads out competition winners along a continuum normalized by the total number of winners without changing any ordering. I then bin $Rank^s_{ic}$ into 20 discrete bins left of the cutoff ($Rank_{ic} < 0$). I refer to the resulting rank as $Rank^{sb}_{ic}$ - the binned rescaled normalized rank. I rely on $Rank^{sb}_{ic}$ for the

Table 2: Covariate Balance

| Bandwidth: | 1 | 10 | 1 | 10 |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Female Lead (0/1) | -0.125 | -0.000 | | |
| | (0.135) | (0.017) | | |
| Female on Team (0/1) | | | -0.048 | -0.002 |
| | | | (0.131) | (0.016) |
| # Observations | 414 | 2323 | 428 | 2380 |
| # Competitions | 207 | 343 | 214 | 349 |
| R squared | 0.217 | 0.736 | 0.216 | 0.735 |
| Mean Dep. Var | 0.500 | 0.460 | 0.500 | 0.461 |

Notes: Dependent variable is indicator for beating the threshold. All regressions include fixed effects for country of firm, product type, industry, and competition, while columns 2 and 4 add linear controls for the running variable (rank) separately on each side of the threshold. Standard errors are clustered to allow for arbitrary correlation within application group.

main specification, but show that results hold for the quintile method of Howell (2017a) as well (Appendix B).

To implement the empirical strategy described in Equation (2), I must assume that unobservable determinants of quality are equally balanced across the threshold. While this assumption is inherently untestable, I can check that at least observable co-variates balance across the threshold. In this case, the list of observable co-variates is quite short. The co-variates available at the firm-level are country of firm, product type, industry, gender of the team lead, and a binary variable for whether there is a female on the team. Unfortunately, only the gender variables yield informative tests. All the other variables should balance for the most part by construction. Indeed, competitions are usually organized around a particular theme, so country of origin, industry, and product-type are likely to balance already. However, no competition was specifically targeted at women entrepreneurs. Hence, the likelihood of having a female head or a female on the team should not balance across the threshold by construction.

In Table 2, I present estimates of Equation (2) where the outcome variable is an indicator for beating the threshold. The first two columns tests for balance in the propensity to have a female lead, and the second two columns tests for balance in having a female on the team at all. In columns 1 and 3, I take just a single firm on either side of the threshold (bandwidth = 1), and in columns 2 and 4, I allow up to 10 firms on either side of the threshold. When the bandwidth is greater than 1, I control linearly for rank, allowing for separate slopes on either side of the threshold. All regressions control for firms' country of origin, product type, industry, and competition, and standard errors are clustered on the application group (see Appendix A). I cannot reject the null hypothesis of no difference in either gender variable across the threshold. While the point estimates are a bit noisy with the bandwidth of 1, at a bandwidth of 10, I can reject at the 5% level all values outside the interval $[-.0346, 0.0341]$ and $[-.0341, 0.0300]$ for female lead and female team, respectively, which I take as re-assuring evidence against differential sorting across the threshold.

# 4 Results

In this section, I present the results of the RD design, using both the subjective metrics and the objective metrics with an out-of-sample prediction of objective scores.

## 4.1 First Stage

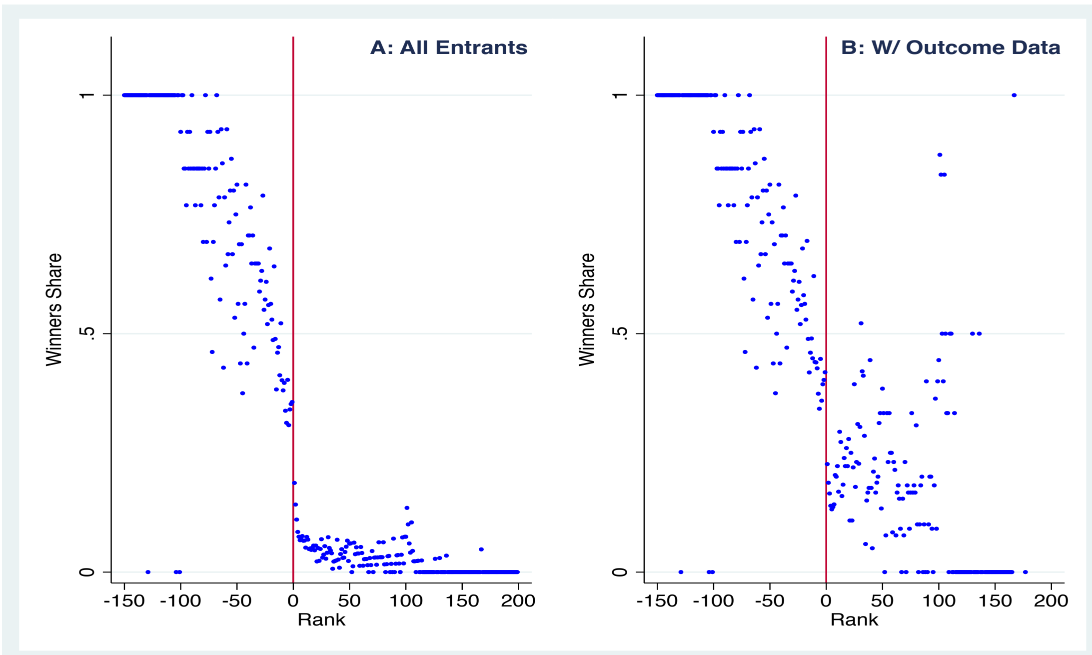
Figure 3 first plots the share of firms that win the competition by normalized rank for all entrants (panel A) and only for firms with outcome data (panel B). In panel A, it can be seen that some competitions award a large amount of prizes (more than 100).[16] The probability of winning the competition is high far from the threshold, and then falls as normalized rank approaches 0. At the threshold – indicated by a red vertical line– the probability of winning the competition drops from about 40% down to 20%, and then descends toward zero as firms miss the threshold by further and further.

Panel A reveals two key points. First, there is a clear discrete drop in the probability of winning the competition right at the threshold. This validates the RD design. Second, receiving a rank to the left of the cutoff ($Rank_{ic} < 0$) does not ensure victory in the competition. This implies that to estimate the impact of winning a competition, I must divide this impact by the first-stage impact of beating the threshold. Clearly, the YN rankings are informative with respect to winning the competition (since the probability of winning is correlated with the rank), but donors sometimes elect not to award firms that beat the threshold a prize, and vice versa. If the donors exploit private information (unobservable to the judges) to select higher quality firms, then one would expect an OLS estimate of success on winning to be biased upwards. However, donors might select firms

[16] These large competitions are mostly from the same donor that runs different generations of the same competitions each year. Large numbers of prizes are actually rare (Figure 1).

based on political reasons or some other criteria unrelated to success, which would bias OLS estimate downward.

Figure 3: First stage impact on winning the competition




Notes: Panel A presents data from all competition entrants (460 competitions, 20,594 firms), while panel B restricts to observations with outcome data (445 competitions, 7,579 firms). In panel B, only competitions with observations both to the right and left of the threshold are included. Y-axis plots the share of firms that eventually won the competition by normalized YN rank.

In panel B, I find same downward sloping relationship left of the threshold as in panel A, but there is a surprising upward trend in the probability of winning as one moves further away from the threshold in the positive direction. This is an artifact of the data collection process. Prior to researching a competition, the research assistants did not know exactly what the cut-off threshold was for the competition. Lacking this information, researchers were instructed to collect outcome data for all competition winners, and then a handful of highly ranked losers as well. Thus, winners have been oversampled relative to the total population. Firms with ranks far to the right of the threshold will only be included in the outcome dataset if, in fact, they won the competition. Thus, one finds a high share of winning far right of the threshold. This is not a problem for the estimation, since I will focus on a narrow bandwidth around the threshold. What is crucial for the empirical strategy is that there is a discrete jump in the probability of winning at the threshold.

## 4.2 RD Results for Subjective Metrics

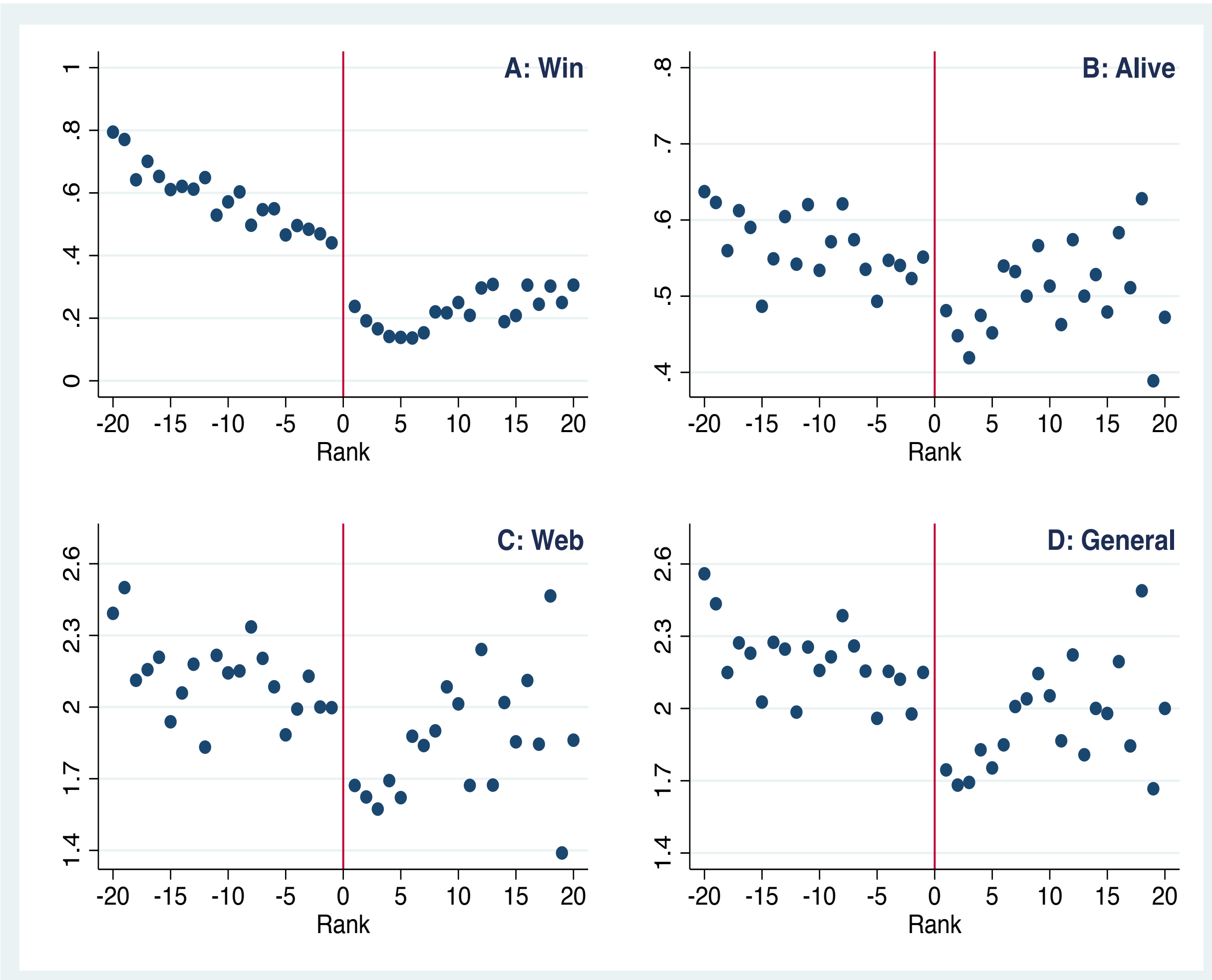
How do the subjective metrics evolve with respect to distance from the threshold? In Figure 4, I zoom in on the threshold (bandwidth of 20) and plot average values of the first stage binary win value (Panel A), along with the 3 subjective metrics (Panels B-D) by normalized rank. For all four panels, metrics are falling as the rank approaches the threshold from the left, as one would expect if judges ranks are informative at all. Furthermore, there is a discrete drop right at the threshold. Finally, right of the threshold, values tend to rise, which makes sense in light of the sampling phenomenon mentioned above. Figure 4 presents visual evidence that before controlling for any co-variates, it appears that winning a competition causally improves firm success.

In Table 3, I present regression estimates of equation (2) that confirm the visual results from Figure 4. Column 1 presents the first stage impact of beating the threshold on winning the competition. I adopt a bandwidth of 1 in Panel A and 10 in Panel B. As in Table 2, all regressions control for country of firm, product type, industry, and competition fixed effects, and estimates in Panel B control linearly for rank on either side of the threshold. Also, standard errors are clustered on the application group. In column 1, I find that, as in Figure 4, there is roughly a 20 percentage point jump in the probability that a firm wins a grant prize (or participates in an accelerator or incubator) when the firm beats the threshold. In both panels, the point estimate is statistically significant at the 1% level. The F-statistics for joint significance of all explanatory variables are 28 and 20, respectively.

In columns 2-7, I present OLS and IV estimates on the three subjective metrics – alive, web, and general scores. Even-numbered columns present the OLS estimates of winning a competition, while odd-numbered columns instrument the binary "win" variable with the first stage from column 1. First, in the even columns, I find that winning a competition is associated with better future firm outcomes. Regardless of the bandwidth, point estimates are statistically significant at the 1% level (except for Panel A column 2) and economically meaningful. In column 2, I find that winners are 13 - 16 percentage points more likely to be alive than losers, on a base rate of 51% – or about 30% more likely to have a live website. For web and general scores, winning firms have between 0.62 and 0.78 higher scores, depending on the specification. Against base rates in the neighborhood of 1.8-1.9, these estimates translate into 40-50% higher web and general scores.

If winning the competition were randomly assigned, the estimates in columns 2, 4, and 6 would constitute causal impacts of program participation. However, since donor organizations sometimes choose to award prizes to firms that missed the cut-off instead of

Figure 4: Competition Outcome and Subjective Metrics by Rank




Notes: Firm-specific metrics averaged by normalized YN rank. Figure includes only 1-round and 2-round competitions with observations both to the right and left of the threshold within the bandwidth 20 (5,182 observations across 407 competitions).

following the YN rankings, it is possible that winning the competition still correlates with some unobservable determinant of firm success. If the donor organizations have private information that leads them to choose systematically higher-quality firms, then the OLS estimates will include a positive selection effect. On the other hand, donor organizations might systematically choose firms of lower quality when they go off the YN rankings if they are driven by political reasons or personal preferences. In this case, the OLS would understate the true causal impact.

To isolate the causal impact from the selection effect, in columns 3, 5, and 7, I instrument participation in the program with the exogenous indicator for beating the threshold. Since placement on either side of the threshold should be as good as randomly assigned, the IV estimates identify just the causal impact. In columns 3, 5, and 7, I find positive and statistically significant coefficients which are significantly higher than the OLS. The impact on finding a live website jumps by a factor of 2 or 3 compared to the OLS, with similar increases in the coefficients on web and general scores. These estimates imply that the causal impact of program participation are quite large, and that if anything, the selection effect is *negative*. When the donor organizations disregard the rankings, they systematically choose lower-quality firms.

How robust are the results in Table 3? In Appendix B, I perform several standard checks. In Table B.1, I adopt the alternative strategy of Howell (2017a) for addressing differences in competition sizes, using the un-scaled underlying ranks $Rank_{ic}$ and controlling for quintile of firm placement. In Table B.1 I find coefficients of similar magnitudes. Standard errors are a bit larger, though most estimates are still statistically significant at the 1% or 5% level. In Figure B.1, I vary the bandwidth between 1 and 25, and find fairly stable results. Finally, in Figure B.2, I re-estimate the IV placing the threshold cut-off counterfactually at ranks between -10 and 10. As one would expect, the IV estimates peak at the true normalized threshold of 0, with placebo tests mostly statistically indistinguishable from 0.

## 4.3 Potential Mechanisms

The results in Table 3 indicate that winning a competition causally improves start-up outcomes. But what drives the result? Before turning to the objective metrics, I investigate two potential explanations.

First, it is possible that these results are driven entirely by the extensive margin – perhaps grants and incubators merely help start-ups survive. Columns 2-3 provide direct

Table 3: RD Results for Subjective Metrics

| | 1st Stage | 2nd Stage | | | | | |
|---|---|---|---|---|---|---|---|
| | | Alive | | Web | | General | |
| | | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A: Bwidth=1* | | | | | | | |
| $1\|Rank < 0$ | 0.224*** | | | | | | |
| | (0.042) | | | | | | |
| Win | | 0.134** | 0.443** | 0.745*** | 2.271*** | 0.780*** | 2.861*** |
| | | (0.056) | (0.174) | (0.230) | (0.729) | (0.230) | (0.777) |
| # Observations | 606 | 606 | 606 | 606 | 606 | 606 | 606 |
| # Competitions | 303 | 303 | 303 | 303 | 303 | 303 | 303 |
| R squared | 0.632 | 0.687 | 0.652 | 0.688 | 0.633 | 0.666 | 0.556 |
| Mean Dep. Var | 0.338 | 0.518 | 0.518 | 1.838 | 1.838 | 1.955 | 1.955 |
| *Panel B: Bwidth=10* | | | | | | | |
| $1\|Rank < 0$ | 0.179*** | | | | | | |
| | (0.034) | | | | | | |
| Win | | 0.162*** | 0.300*** | 0.616*** | 1.246*** | 0.620*** | 1.417*** |
| | | (0.019) | (0.069) | (0.082) | (0.271) | (0.076) | (0.308) |
| # Observations | 3239 | 3239 | 3239 | 3239 | 3239 | 3239 | 3239 |
| # Competitions | 384 | 384 | 384 | 384 | 384 | 384 | 384 |
| R squared | 0.321 | 0.382 | 0.370 | 0.372 | 0.357 | 0.372 | 0.345 |
| Mean Dep. Var | 0.333 | 0.508 | 0.508 | 1.872 | 1.872 | 1.954 | 1.954 |

Notes: All regressions include fixed effects for country of firm, product type, industry, and competition, while panel B adds linear controls for the running variable (rank) separately on each side of the threshold. Standard errors are clustered to allow for arbitrary correlation within super competition. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

evidence of an extensive-margin effect, but columns 4-7 pool both live and dead firms together, so the results represent a mix of the intensive and extensive margin. In Table B.2, I condition on live firms and re-estimate the OLS and IV just for web and general scores. Here, I find some evidence that intensive margin plays a role. Point estimates are all positive and on the same order as in Table 3, though the standard errors are larger. The clearest evidence comes from the general score. In Panel B, I find that results are statistically significant at the 5% level. Conditional on having a live website, winning a competition yields at least a better overall web presence.

Second, even if credit markets were perfect, one might expect entrepreneurship programs to at least keep start-ups alive for a while because they offer resources for free. To test this hypothesis, I exploit variation in timing of the competition to see if benefits are merely short-lived. In Tables B.3 and B.4, I split the sample into early competitions (run between 2010-2013) and later competitions (run between 2014-2015). Since all outcomes metrics were assessed in 2015-2016, firms in the early sample had 2-5 years for benefits to evolve as a result of program participation, while firms in the late sample had only a year or two. If programs merely offer a temporary benefit, effects should only materialize for the latter sample. In Table B.3, I find that impacts are large and statistically significant even in the early sample. The point estimates are a bit higher for the later sample, but there are still large effects for firms 2-5 years after a competition, indicating that resources from entrepreneurship programs do not merely fall straight to the bottom line.

## 4.4 RD Results for Objective Metrics

To assess the economic significance of previous results, I re-estimate program impacts for the Mattermark variables. These variables track closer to traditional performance measures; however, since the overlap between Mattermark and YN firms is only about 35%, statistical power will be low. To increase precision, I rely on the correlations in Figure 2 and predict out of sample the Mattermark scores for the 65% of firms with missing data. I then perform the RD analysis on both the raw Mattermark data and the projected values.

To predict out of sample, I estimate equation (1) for each of the 6 Mattermark outcomes, including dummy variables for both web score values and general score values on the right hand side. I then compute fitted values based on the estimated coefficients, exponentiate them to put in levels, winsorize the top 5 percent, and set all values to 0 for firms that the RAs could not find a live website for (alive=0). In-sample comparisons are reported (in log values) in Figure B.3. Regressing predicted values on true values without a constant, I find slope parameters that range between 0.82 and 0.99, indicating a reasonable in-sample fit.

Table 4: RD Results - Objective Metrics

| | Non Projected | | Projected | | | |
|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV |
| Bandwidth: | 10 | 10 | 1 | 1 | 10 | 10 |
| *Panel A : Funding* | | | | | | |
| Win | -0.647 | 1.170 | 0.165*** | 0.440** | 0.078*** | 0.260*** |
| | (0.534) | (1.027) | (0.056) | (0.208) | (0.018) | (0.071) |
| # Observations | 135 | 135 | 606 | 606 | 3239 | 3239 |
| # Competitions | 30 | 30 | 303 | 303 | 384 | 384 |
| R squared | 0.814 | 0.769 | 0.799 | 0.783 | 0.730 | 0.716 |
| Mean Dep. Var | -1.31 | -1.31 | 0.34 | 0.34 | 0.26 | 0.26 |
| *Panel B : Employees* | | | | | | |
| Win | 0.042 | 0.685 | 1.439** | 4.769** | 1.182*** | 2.840*** |
| | (0.210) | (0.632) | (0.699) | (2.022) | (0.198) | (0.763) |
| # Observations | 272 | 272 | 606 | 606 | 3239 | 3239 |
| # Competitions | 51 | 51 | 303 | 303 | 384 | 384 |
| R squared | 0.478 | 0.452 | 0.739 | 0.707 | 0.530 | 0.515 |
| Mean Dep. Var | 1.88 | 1.88 | 4.15 | 4.15 | 3.74 | 3.74 |
| *Panel C : Web Uniques* | | | | | | |
| Win | 0.177 | 4.597** | 0.178*** | 0.401* | 0.120*** | 0.263*** |
| | (0.438) | (2.032) | (0.067) | (0.220) | (0.029) | (0.094) |
| # Observations | 590 | 590 | 606 | 606 | 3239 | 3239 |
| # Competitions | 87 | 87 | 303 | 303 | 384 | 384 |
| R squared | 0.336 | 0.166 | 0.823 | 0.815 | 0.528 | 0.522 |
| Mean Dep. Var | -1.81 | -1.81 | 0.25 | 0.25 | 0.29 | 0.29 |

Notes: All regressions include fixed effects for country of firm, product type, industry, and competition. Columns marked bandwidth =10 add linear controls for the running variable (rank) separately on each side of the threshold. Non-projected outcome metrics are in logs, and projected outcomes are in levels. Standard errors are clustered to allow for arbitrary correlation within application group. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 4 presents the OLS and IV impacts for bandwidths of 1 and 10 for the projected Mattermark values and just the bandwidth of 10 for the non-projected (i.e. reported) values for all 6 outcomes. I find that, using the reported Mattermark data, point estimates are positive in the IV, though only statistically significant for unique web visits. By contrast,

Table 4 Continued: RD Results - Objective Metrics

| | Non Projected | | Projected | | | |
|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV |
| Bandwidth: | 10 | 10 | 1 | 1 | 10 | 10 |
| *Panel D : LinkedIn* | | | | | | |
| Win | 0.611*** | 0.972 | 0.051 | 0.252* | 0.069*** | 0.175*** |
| | (0.221) | (1.109) | (0.041) | (0.141) | (0.012) | (0.053) |
| # Observations | 333 | 333 | 606 | 606 | 3239 | 3239 |
| # Competitions | 62 | 62 | 303 | 303 | 384 | 384 |
| R squared | 0.488 | 0.476 | 0.820 | 0.800 | 0.567 | 0.557 |
| Mean Dep. Var | -0.81 | -0.81 | 0.26 | 0.26 | 0.26 | 0.26 |
| *Panel E : Facebook* | | | | | | |
| Win | 0.496* | 0.835 | 0.595*** | 0.877 | 0.379*** | 0.400* |
| | (0.251) | (1.307) | (0.195) | (0.562) | (0.061) | (0.206) |
| # Observations | 437 | 437 | 606 | 606 | 3239 | 3239 |
| # Competitions | 78 | 78 | 303 | 303 | 384 | 384 |
| R squared | 0.485 | 0.480 | 0.802 | 0.799 | 0.535 | 0.535 |
| Mean Dep. Var | 0.16 | 0.16 | 0.86 | 0.86 | 0.91 | 0.91 |
| *Panel F : Twitter* | | | | | | |
| Win | -0.138 | 1.145 | 0.072* | 0.150 | 0.076*** | 0.155*** |
| | (0.204) | (1.264) | (0.039) | (0.108) | (0.013) | (0.048) |
| # Observations | 537 | 537 | 606 | 606 | 3239 | 3239 |
| # Competitions | 85 | 85 | 303 | 303 | 384 | 384 |
| R squared | 0.329 | 0.281 | 0.848 | 0.843 | 0.543 | 0.536 |
| Mean Dep. Var | -1.00 | -1.00 | 0.18 | 0.18 | 0.22 | 0.22 |

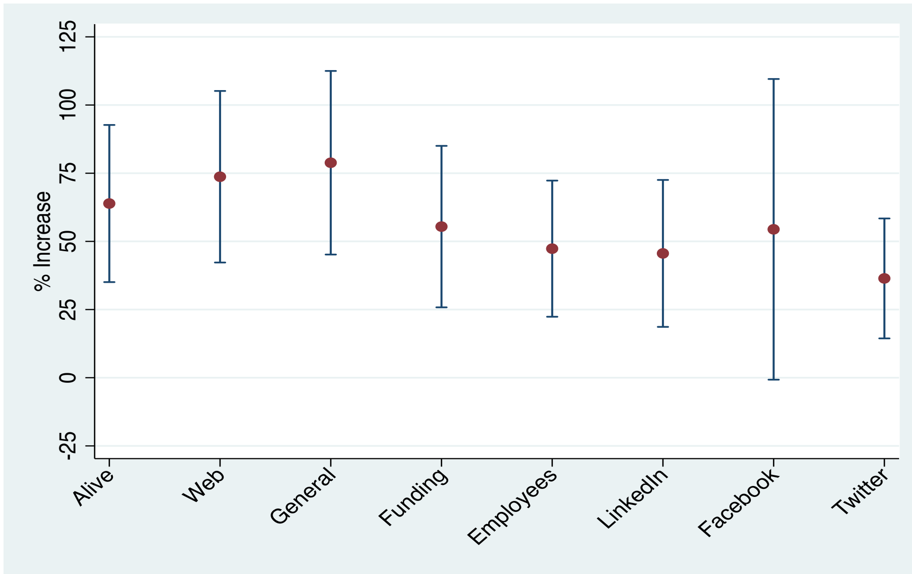
Notes: All regressions include fixed effects for country of firm, product type, industry, and competition. Columns marked bandwidth =10 add linear controls for the running variable (rank) separately on each side of the threshold. Non-projected outcome metrics are in logs, and projected outcomes are in levels. Standard errors are clustered to allow for arbitrary correlation within application group. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

using the projected values, I find positive and statistically significant coefficients in the IV for all 6 outcome variables. Additionally, IV estimates are roughly 2-3 times larger than the OLS estimates, as in Table 3.

Figure 5 summarizes the IV impacts for all 3 subjective metrics and 5 objective metrics.[17] Since the RD estimates a local average treatment effect, the relevant population

[17]Unique web visits are omitted for ease of viewing. The impacts on unique web visits is 321% with a

Figure 5: Summary of Results




Notes: Point estimates and 95% confidence intervals based on IV results from Tables 3 and 4.

consists of firms within the preferred bandwidth (10) that missed the cutoff. In Figure 5, I scale the IV estimates from Tables 3 and 4 by the mean values in this population (median for objective metrics), and report point estimates with red dots, and 95% confidence intervals with blue bars. Point estimates imply that winning a competition increases subjective metrics 64-79%. It would be difficult to benchmark the web and general impacts against the literature (since these metrics are novel, to my knowledge), but I can compare the increased survival probability to other studies. Lerner et al. (2017) and Howell (2017a) both estimate that angel investment and US grants increase firm survival on the order of 20%.[18] A potential explanation for the difference is that the firms studied in Lerner et al. (2017) are likely much larger and more established – the firm has to be fairly well along if it is applying for millions of dollars from an angel investor. By contrast, the firms in this study are just beginning operations, and thus might be in greater need of financial assistance and support.

In Figure 5, I also find that winning a competition increases objective metrics by 36-55%. Looking at the first objective metric, I find that winning a competition on average

95% confidence interval of [96% ,545%]

[18] Lerner et al. (2017), Table 7. Howell (2017a) page 1150. Both Lerner et al. (2017) and Howell (2017a) distinguish between a failed business and a successful exit, which I do not. Thus, if anything, I am understating the impact on firm survival relative to Lerner et al. (2017) and Howell (2017a) .

generates \$260,000 USD in aggregate follow-on funding, which is exactly 10 times greater than the average size of a prize in the dataset (\$26,000 USD). While funding is different than profits, this is still an enormous return on a small investment. By comparison, Lerner et al. (2017) finds that angel investment increases follow-on financing around \$3 million USD, though the point estimates are not precise.[19] On an average deal size of \$1.2 million dollars, this represents only a 3x return, which is still quite substantial. On the other side of the spectrum, Howell (2017a) estimates that a grant of \$150,000 USD from the US DOE generates more than \$2.7 million USD in follow-on financing, or 18x return.[20]

Finally, I can compare the impact on employment to other estimates. In Figure 5, I find that winning a competition leads to hiring just under 3 extra employees. Dividing by the average value of a prize, that equals roughly \$9,000 in up-front cost for every full-time employee hired by a start-up.[21] If these results hold in general equilibrium, (i.e., if these are 3 *extra* jobs), then they suggest that entrepreneurship policies might represent a cost effective way to create jobs. This 48% increase is right in line with the 40% figure quoted by Lerner et al. (2017) with respect to angel investments.[22]

### 4.5 Heterogeneous Impacts

Lastly, I investigate heterogeneous treatment effects by both competition and firm-specific co-variates, which speak to potential mechanisms.

In terms of competition variables, I classify competitions along three dimensions. First, I groups competitions by the size of the largest prize offered in the competition.[23] Small-prize competitions offer a prize of less than \$1,000 USD, medium-prize competitions offer a prize of between \$1,000 USD and \$10,000 USD, and large-prize competitions offer a prize over \$10,000 USD. Next, I characterize the donor organization as either Firm, Government Agency, NGO, or University. Finally, I characterize the competition types as either Grant, Accelerator/Incubator, Idea, or Pitch/Other. The distribution of competitions along these three dimensions is reported in Table 5.

In Figures 6, 8, and 7, I separately estimate IV impacts on the each of the subjective metrics and report point estimates and 95% confidence intervals by group and outcome

[19] Lerner et al. (2017), Table 4.
[20] Howell (2017a), Table 3.
[21] The full calculation is 26,315.33/2.84 = \$9,265.
[22] Lerner et al. (2017) Table 7.
[23] In the case that prizes are non-monetary (e.g., office space for 3 months in Silicon Valley, or free travel to investor meetings, etc) I convert to USD valuation as best I can. Also, incubators and accelerators often offer cash prizes in addition to administrative support and/or office space.

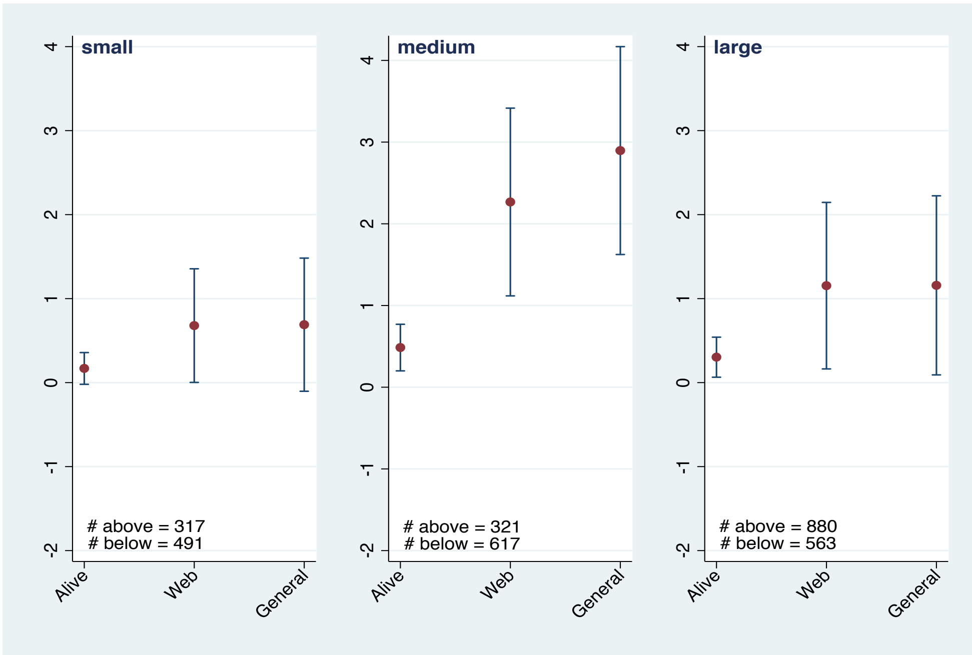
Table 5: Competition-level Variables

| | # Competitions (1) | Percentage (2) |
|---|---|---|
| *Competition Prize Size* | | |
| Small ($x \leq$ 1,000 USD) | 120 | 0.261 |
| Medium (1,000 USD $< x \leq$ 10,000 USD) | 212 | 0.461 |
| Large (10,000 USD $< x$ ) | 128 | 0.278 |
| Total | 460 | 1.000 |
| *Donor Organization Type* | | |
| Firm | 64 | 0.139 |
| Government | 51 | 0.111 |
| NGO | 239 | 0.520 |
| University | 106 | 0.230 |
| Total | 460 | 1.000 |
| *Competition Prize Type* | | |
| Idea | 84 | 0.183 |
| Accelerator | 25 | 0.054 |
| Grant | 273 | 0.593 |
| Incubator | 32 | 0.070 |
| Pitch/Other | 46 | 0.100 |
| Total | 460 | 1.000 |

variables. For all regressions, the bandwidth is 10, and all the same controls are included as in Table 3. Standard errors are clustered on the application group.

In Figure 6, I find that medium-size prizes generate the largest impacts, with smaller impacts coming from small and large-prize competitions, though still statistically significant. The fact that program impacts are increasing in program size towards the bottom of the distribution is perhaps not surprising. If entrepreneurship programs alleviate financial constraints, then one would expect the dollar value of the prize to matter. However, it could also be that larger prizes simply carry more weight when applying for future financing, so it is difficult to separate the financial channel from the pure signaling channel. The fact that large-prize competitions have smaller impacts than medium-size competitions is surprising. It could be that firms sort into competitions based on prize size, and only the highest quality firms go after the largest prizes. If financial constraints play less of a role for the high quality firms, this could explain the non-monotonicity. Note that Howell (2017a) also find declining impacts towards the top of the prize-size distribution (no impact from grants above $1 million USD), though even the "small" grants from Howell (2017a) would

Figure 6: Prize Size




Notes: Bandwidth is set to 10 for all regressions, and all the same controls and restrictions apply as in Table 3. Standard errors are clustered on the application group. Text indicates number of observations included above and below the threshold.

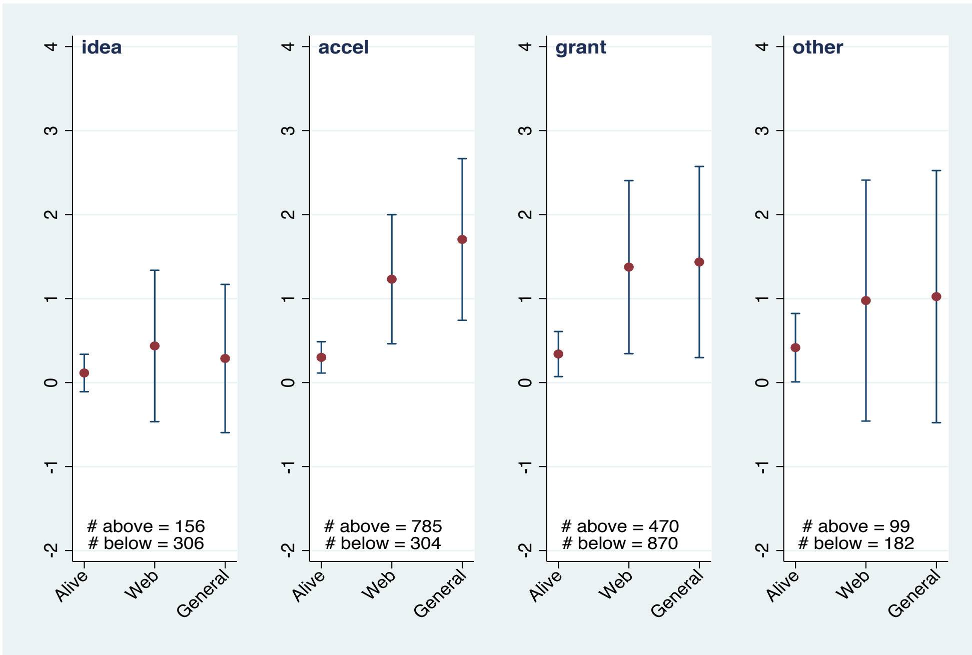
be considered large in the context of YN competitions.

Looking at donor organization and prize type, I find in Figure 8 that programs run by NGOs have especially large impacts, programs run by governments and firms have modest impacts, and programs run by universities yield small and statistically insignificant results. In Figure 7, I find that grants and accelerators or incubators programs yield comparable results. This suggests that both support and financing are important for start-ups.

Next, I investigate heterogeneous effects by firm-specific co-variates. In particular, I estimate separate effect on the subjective metrics by gender composition of the team, origin of founder, and industry. Along any given dimension, I group firms into one of the mutually exclusive bins associated to the dimension (e.g., female lead vs non-female lead), and estimate the IV just on the sub-sample included in the bin. This approach is equivalent to interacting a dummy variable for each bin with all co-variates and fixed effects and estimating interaction effects, as in Howell (2017a).[24]

[24]The procedure drops any competition that does not have at least one firm in the relevant bin on either side of the threshold.

Figure 7: Competition Type



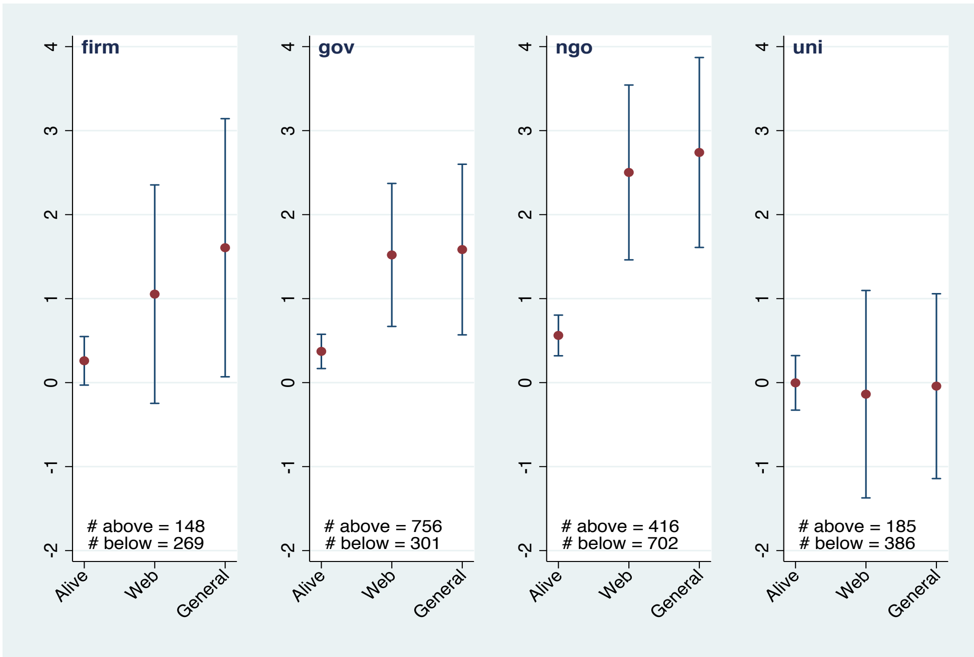

Notes: Bandwidth is set to 10 for all regressions, and all the same controls and restrictions apply as in Table 3. Standard errors are clustered on the application group. Text indicates number of observations included above and below the threshold.

Figures 9 presents results by gender composition. The two panels on the left splits the sample between female-lead firms (left) and male-lead firms (right), and the two panels of the right splits the sample between teams that have at least 1 female on the team (left) and all-male teams (right). In both cases, I find large and statistically significant impacts for the male samples, but not for the female samples. For firms with at least 1 female on the team, point estimates are close to the all-male sample, but standard errors are much larger. The discrepancy could simply be an artifact of the small number of firms with female lead and female workers in the sample. Alternatively, it might be that female entrepreneurs require support and mentoring or other auxiliary services beyond grants in order to be successful. Without more data, I cannot distinguish between these explanations.

Next, I break the sample by origin of the team leader. There are many ways to group countries. In Figure 10, I break the sample by continent (top panel) or I group countries by quartile of per-capita income (below panel).[25] Nothing much stands out across continents,

[25]Per-capita GDP come from the World Bank Development Indicators. Quartiles represent the universe of countries, not just the countries included in the YN dataset.

Figure 8: Organization Type

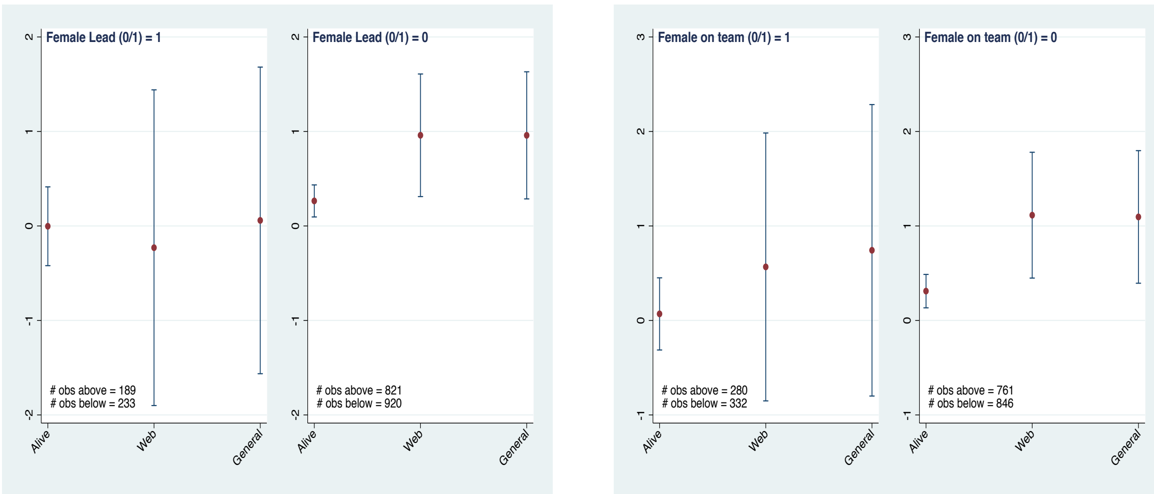



Notes: Bandwidth is set to 10 for all regressions, and all the same controls and restrictions apply as in Table 3. Standard errors are clustered on the application group. Text indicates number of observations included above and below the threshold.

except that the Africa sample generates large, but very imprecise point estimates. For the quartile decomposition, the first quartile (Q1) includes all countries with per-capita income less than $1,100 USD, second quartile (Q2) includes all countries with per-capita income between $1,100 USD and $4,000, third quartile (Q3) includes all countries with per-capita income between $4,000 USD and $14,000, and fourth quartile (Q4) includes all countries with per-capita income above $14,000. I find precise estimates only for the Q3 and Q4 samples, i.e. the richer countries. Estimates for the Q1 and Q2 samples are large, but statistically indistinguishable from 0. Again, this could be an artifact of small samples.

In Figure 11, I break the sample by impediments to starting a business. The top panel groups countries by time required to start a business, while the below panel breaks the sample by number of procedures. Both metrics come from the WDI. While costs of starting a business likely correlate with GDP per-capita, the sample are not exactly the same. I find significant observations counts in the Q1 sample, where estimates in Figure 10 were underpowered. Here, with sufficient coverage at the low end of the distribution, I find statistically significant impacts. Point estimates are no higher than in the Q4 samples,

Figure 9: Gender Heterogeneity for Lead Female (Left) and Female on Team (Right)

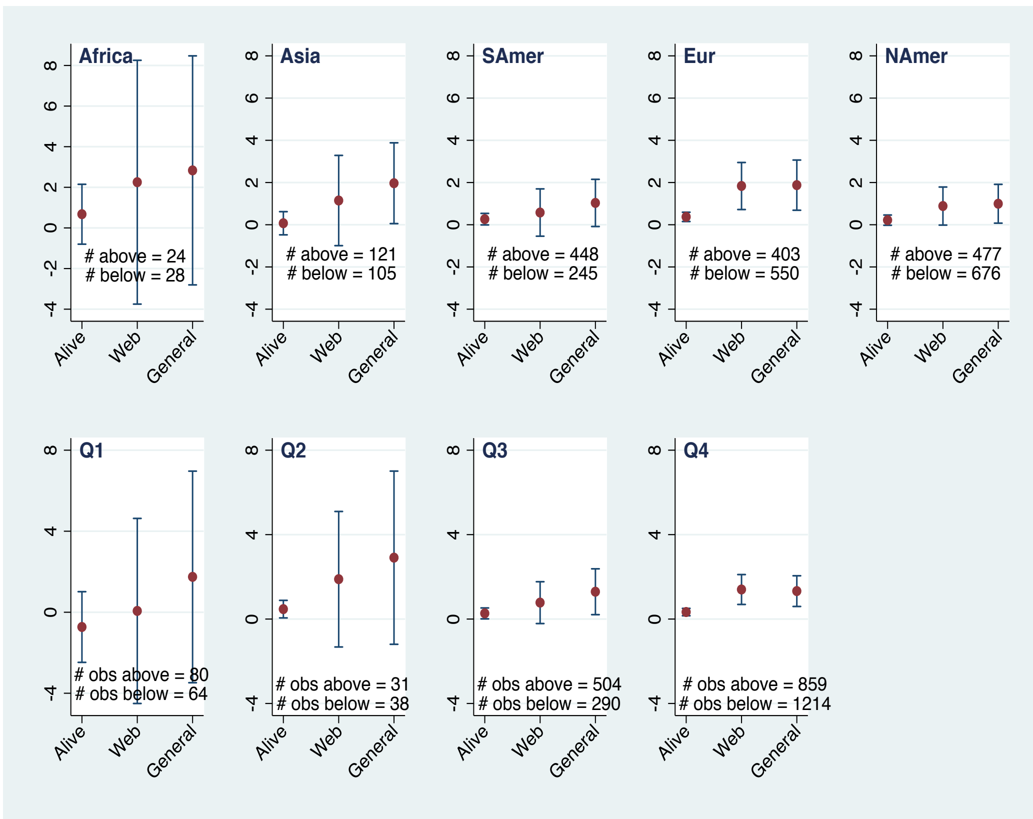



Notes: Bandwidth is set to 10 for all regressions, and all the same controls and restrictions apply as in Table 3. Standard errors are clustered on the application group. Text indicates number of observations included above and below the threshold.

but still I find economically meaningful impacts in countries where it is difficult to start a business. One might expect impacts to be larger in the Q1 samples because these are the markets where we expect credit to be allocated least efficiently. However, if ancillary services such as infrastructure, internet speed, etc are also lower quality in these markets, then perhaps it is harder for programs to lift entrepreneurs up.

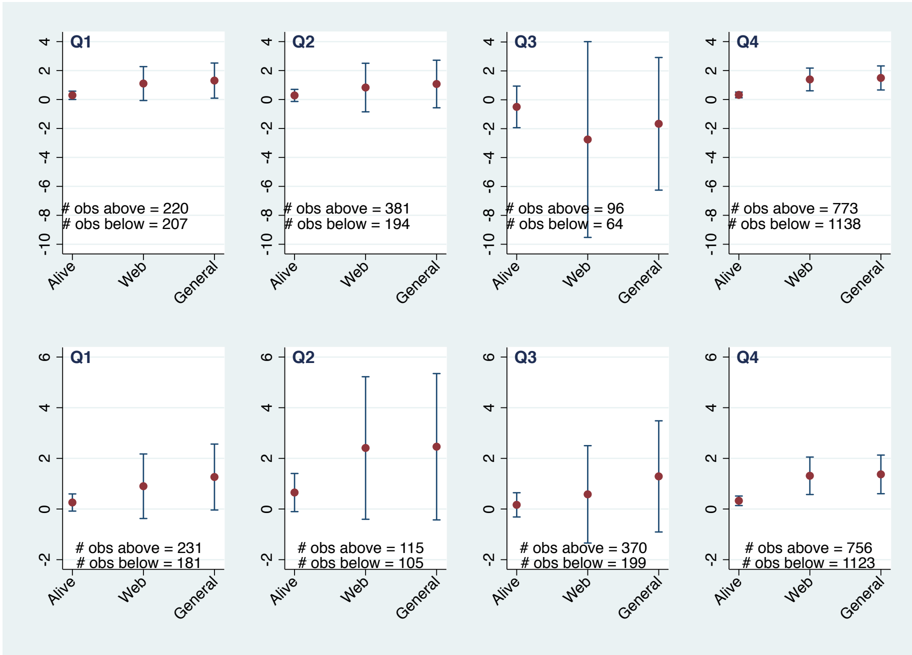
Finally, in Figure 12, I separately estimate impacts by industry. Impacts are largely evenly distributed across industries. A potential outlier is the clean technology sector ("ergy"). Here, I find a large impact on the general score, though the point estimates are imprecise. Howell (2017a) also finds large impacts of grants on firms in the clean tech sector. Here, I find that these impacts generalize to other sectors as well.

Figure 10: GDP1



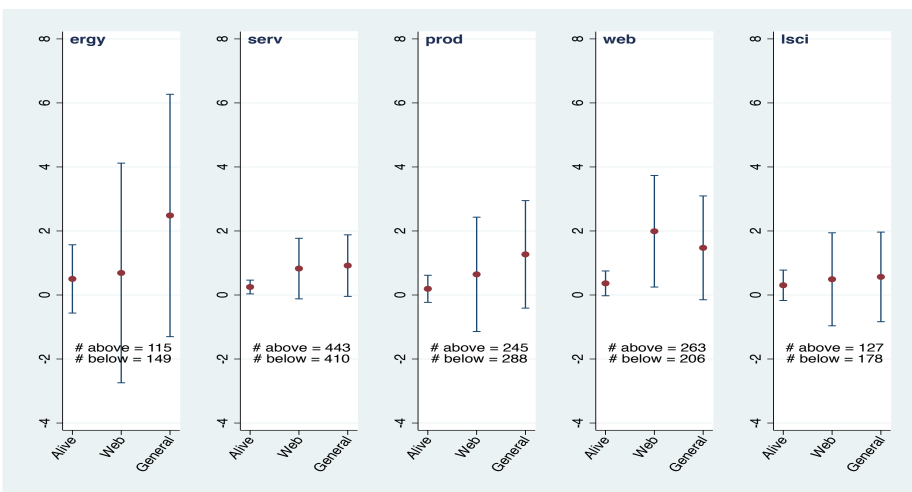

Notes: Bandwidth is set to 10 for all regressions, and all the same controls and restrictions apply as in Table 3. Standard errors are clustered on the application group. Text indicates number of observations included above and below the threshold.

Figure 11: Heterogeneity in Entering Time (Top) and in Entering Procedures (Below)




Notes: Bandwidth is set to 10 for all regressions, and all the same controls and restrictions apply as in Table 3. Standard errors are clustered on the application group. Text indicates number of observations included above and below the threshold.

Figure 12: Industry




Notes: Bandwidth is set to 10 for all regressions, and all the same controls and restrictions apply as in Table 3. Standard errors are clustered on the application group. Text indicates number of observations included above and below the threshold.

# 5 Conclusion

This paper estimates the impacts of entrepreneurship programs (grants and accelerators or incubators) on early stage start-ups in different markets and industries across the globe. I exploit a quasi-natural experimental setting owing to the competition structure of program selection and make use of confidential proprietary data on thousands of firms participating in several hundred competitions. I develop novel metrics of firm success based on the subjective evaluations of research assistants and demonstrate that these metrics correlate very well with auxiliary data from a business aggregator web site.

Exploiting the discontinuous jump in probability around the threshold of selection, I find that entrepreneurship programs increase the survival probability of a firm, as well as web-based measures of performance. In addition, I find that program participation increases follow-on funding significantly, as well as employment and measures of internet activity. In short, programs have, on average, a causal impact on firm success.

The largest impacts are found for medium-size prize competitions, indicating a potential non-monotonicity in returns to program size. This non-monotonicity is echoed by Howell (2017a), and might speak to differential sorting based on prize size. Additionally, while impacts are precisely estimated for male-lead firms and all-male teams, impacts on female-lead firms and firms with at least 1 female on the team are indistinguishable from zero. This is likely due to smaller sample sizes, but might reflect the need to couple entrepreneurship programs with other ancillary services, such as mentoring, for female entrepreneurs. Finally, statistically significant impacts are found for firms coming from countries both where it is easy to start a business and where it is difficult. The fact that estimates are significant even in the latter case implies that even where support services for entrepreneurs may be weak, entrepreneurship programs can succeed.

# References


Bays, J., Chakravorti, B., Goland, T., Harris, B., Jansen, P., McGaw, D., Newsum, J., Simon, A., & Tallento, L. (2009). And the winner is…": Capturing the promise of philantropic prizes. *McKinsey & Company*.

Beck, T., Behr, P., & Madestam, A. (2011). Sex and credit: Is there a gender bias in lending?

Blanchflower, D. G., Levine, P. B., & Zimmerman, D. J. (2003). Discrimination in the small-business credit market. *Review of Economics and Statistics*, 85(4), 930–943.

Dechezleprêtre, A., Einiö, E., Martin, R., Nguyen, K.-T., & Van Reenen, J. (2016). *Do tax incentives for research increase firm innovation? An RD design for R&D*. Technical report, National Bureau of Economic Research.

Fafchamps, M. & Quinn, S. (2016). Aspire. *The Journal of Development Studies*, (pp. 1–19).

Fay, M. & Williams, L. (1993). Gender bias and the availability of business loans. *Journal of Business Venturing*, 8(4), 363 – 376.

Fisman, R., Paravisini, D., & Vig, V. (2017). Cultural proximity and loan outcomes. *American Economic Review*, 107(2), 457–92.

Gonzalez-Uribe, J. & Leatherbee, M. (2017). The effects of business accelerators on venture performance: evidence from start-up chile. *The Review of Financial Studies*.

Hanson, A., Hawley, Z., Martin, H., & Liu, B. (2016). Discrimination in mortgage lending: Evidence from a correspondence experiment. *Journal of Urban Economics*, 92, 48 – 65.

Howell, S. T. (2017a). Financing innovation: evidence from r&d grants. *The American Economic Review*, 107(4), 1136–1164.

Howell, S. T. (2017b). *Learning from Feedback: Evidence from New Ventures*. Technical report, National Bureau of Economic Research.

Kerr, W. R. & Nanda, R. (2015). Financing innovation. *Annual Review of Financial Economics*, 7, 445–462.

Klinger, B. & Schündeln, M. (2011). Can entrepreneurial activity be taught? quasi-experimental evidence from central america. *World Development*, 39(9), 1592–1610.

Kortum, S. & Lerner, J. (2000). Assessing the contribution of venture capital to innovation. *RAND journal of Economics*, (pp. 674–692).

Lelarge, C., Sraer, D., & Thesmar, D. (2010). Entrepreneurship and credit constraints: Evidence from a french loan guarantee program. In *International differences in entrepreneurship* (pp. 243–273). University of Chicago Press.

Lerner, J., Schoar, A., Sokolinski, S., & Wilson, K. (2017). The globalization of angel investments: Evidence across countries. *Journal of Financial Economics*.

Masters, W. A. & Delbecq, B. (2008). Accelerating innovation with prize rewards. *History and Typology of Technology Prizes and a New Contest Desgin for Innovation in African Agriculture*.

McKenzie, D. (2017). Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition. *American Economic Review*.

Smith, B. & Viceisza, A. (2017). Bite me! abc's shark tank as a path to entrepreneurship. *Small Business Economics*, (pp. 1–17).

Yu, S. (2016). How do accelerators impact the performance of high-technology ventures? *Browser Download This Paper*.

# Appendix

## A Data Appendix

The dataset for this paper is based on the administrative records of YN. Founded in 2010, YN has adjudicated hundreds of competitions for various donor organizations over the past 7 years. For each competition, YN saves the name and unique entry id (assigned by YN) of each competition entrant. YN also saves the identification of each judge, the criteria by which the judge evaluated a given firm, and the raw judge scores. While these records are confidential, I was given access to the raw data as a private consultant for YN on site at the YN headquarters, as were all RAs. Though RAs had to know the actual names of firms when searching online, firm names and competition names were removed from the final dataset once the outcome metrics were collected, thus anonymizing the data for analysis.

In the raw data, YN collects all firms together that applied in the same "application group." An "application group" is a set of firms that apply under the same submission call and are judged by the same criteria. The number of application groups and entrants by year are reported in Table A.1. In total, YN received 26,856 applications between 2010-2015, spread across 387 application groups. There was substantial heterogeneity in the size of each group, with the mean number of applicants over all equaling 69. The smallest application groups received only a handful of applications (sometimes only 1), and the largest application group included 558 entrants. For the analysis, I must observe firms both above and below the critical threshold in the same competition, so clearly application groups with only 1 or maybe a handful of entrants will be dropped.

An "application group" is commissioned by a donor organization and organized by YN, but in fact, a single "application group" might comprise multiple mutually exclusive simultaneous "competitions." The distinction is that firms in the same competition are directly competing for the same award, while firms in the same "application group" might not be. For example, a donor might run two distinct competitions for "clean technology" firms and "social good" firms at the same time. Firms in the "clean technology" track compete only against other firms in the "clean technology" track, and the same holds for firms in the "social good" track, though all firms in the application group are judged at the same time along the same metrics. There are really two sets of prizes (one for each track) and two distinct thresholds, but only 1 ordered ranking of firms. In order to have an apples-to-apples comparison, I broke up these "application groups" by hand into individuated competitions and computed a unique threshold for each competition.

Table A.1: Applications Rounds and Entrants by Year

| | # Applications Rounds | # Entrants | Entrants/Applications Round | | |
|---|---|---|---|---|---|
| | | | Mean | Min | Max |
| Year | (1) | (2) | (3) | (4) | (5) |
| 2010 | 16 | 1206 | 75 | 1 | 199 |
| 2011 | 53 | 3569 | 67 | 2 | 259 |
| 2012 | 69 | 5237 | 76 | 2 | 558 |
| 2013 | 105 | 6445 | 61 | 3 | 209 |
| 2014 | 135 | 9572 | 71 | 3 | 260 |
| 2015 | 9 | 827 | 92 | 1 | 212 |
| All Years | 387 | 26856 | 69 | 1 | 558 |

The procedure for individuating competitions, identifying winners, and assigning thresholds went as follows. First, RAs searched for the competition website by name on the Internet. Donor organizations almost always publicize their competitions (publicity for the winners is largely the point of the competitions). Once a website for the competition can be identified, RAs searched for a list of winners. These public lists were the only means available to identify the winners. If the RAs could not find a list of winners, the competition was not used in the analysis.

Next, surveying the list of winners, the RAs matched the names of winning firms with the names of firm entrants in the YN database. Comparison did not always yield identical matches across the two information sets. RAs attempted to resolve naming discrepancies when they arose.

Next, RAs evaluated whether the application group featured multiple tracks – i.e. multiple competitions. This evaluation was based on public information about the competition, as well as the description of the winning firms. For example, winning firms were often identified as winner of a specific track. For example "Winner – Clean Technology Track" vs, "Winner – Social Good Track." In this case of multiple simultaneous tracks, the RAs separated winners by track. These tracks then serve as the individuated competitions.

After breaking apart application groups and dropping competitions for which the RAs could not find winners, there are 460 individuated competitions with a total 20,828 entrants. The distribution of competitions by tracks is reported in Table A.2 (column 4). Of the 460 usable competitions, 231 of these competitions were not tracked with other competitions.

I.e., For 231 application groups, the application group was a single competition. The other 229 competitions were housed under a common application group with at least one other competition. At maximum, 16 individuated competitions were organized within a single application group (this happened twice).

Table A.2: Competitions by Rounds and Tracks

| | # Rounds | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| # Tracks | 1 Round | 2 Rounds | 3 Rounds | Total |
| 1 | 147 | 68 | 16 | 231 |
| 2 | 4 | 18 | 2 | 24 |
| 3 | 9 | 12 | | 21 |
| 4 | 12 | 52 | 20 | 84 |
| 5 | | 40 | | 40 |
| 6 | 6 | | | 6 |
| 7 | 14 | | | 14 |
| 8 | 8 | | | 8 |
| 16 | 16 | 16 | | 32 |
| Total | 216 | 206 | 38 | 460 |

Notes:

After identifying winners and individuating competition, we then collected the outcome data by firm. The RAs were first instructed to collect outcome data for all winning firms (those firms listed on the competition website).

Next, we needed outcomes for some loser firms in order to compute the RD estimates. When there was only one competition per application group and the number of entrants was not too large, then the RAs just collected outcome data for all the loosing firms. However, when there were multiple tracks or the number of entrants was large, RAs had to use their own judgment with respect to two issues. First, if the application group featured multiple tracks, the RAs did their best to assign loosing firms to specific tracks (competitions). To make these assignments, RAs consulted the application material to YN and tried to match the company to the likely track. There is more uncertainty with respect to the track of the losers compared to the winners because the winners were usually associated to a track right on the website, but the RAs were able in many cases to make sensible assignments. If the RA felt that he or she did not have enough information to make an assignment, I assigned the firm randomly to one of the possible tracks. Second, RAs had to choose how many loosing firms to evaluate. I instructed the RAs loosely to collect about twice as many loosing firms as winning firms and to focus attention on firms around the threshold. In total, the RAs collected outcome data for 7,883 firms.[26]

One final complication requires mention with respect to the computation of thresholds, which is that competitions often feature multiple rounds. For example, donor organizations often contract with YN to judge only the first round of a competition, leaving final award assignment for a later round, in which selection criteria is unobserved. In this case, YN ranks are influential (though perhaps not 100% deterministic) for advancement in the competition. In this case, only the threshold used to determine advancement right after YN judging is exogenous, and hence useful for the analysis.

Table A.3: Competitions and Entrants by Threshold

| | Competitions | | # Entrants (All) | | # Entrants (w/ Outcome Data) | |
|---|---|---|---|---|---|---|
| Threshold<br>(1) | Number<br>(2) | Cumulative<br>(3) | Count<br>(4) | Mean<br>(5) | Count<br>(6) | Mean<br>(7) |
| 1 | 57 | 0.12 | 1226 | 22 | 212 | 4 |
| 2 | 98 | 0.34 | 1652 | 17 | 509 | 5 |
| 3 | 81 | 0.51 | 2415 | 30 | 628 | 8 |
| 4 | 23 | 0.56 | 1199 | 52 | 201 | 9 |
| 5 | 38 | 0.65 | 1466 | 39 | 412 | 11 |
| 6 | 27 | 0.70 | 1426 | 53 | 325 | 12 |
| 7 | 15 | 0.74 | 829 | 55 | 187 | 12 |
| 8 | 16 | 0.77 | 575 | 36 | 240 | 15 |
| 9 | 13 | 0.80 | 646 | 50 | 233 | 18 |
| 10+ | 92 | 1.00 | 9394 | 102 | 4936 | 54 |
| Total | 460 | 1.00 | 20828 | 45 | 7883 | 17 |

Notes: The threshold reported in column 1 corresponds to the number of firms that advance in the competition as a result of YN judging. For a 1-round competition, the threshold corresponds to the number of firms that win the competition. For a 2-round competition, the threshold corresponds to the number of firms that advance to the "finals" round. Columns 4-5 report descriptive statistics for all entrants, while columns 6-7 restricts to observations with non-missing outcome data. Ranking data is missing for some firms with outcome data (234).

To fix ideas, consider an example in which a donor organization contracts with YN to judge the first round of a competition. Upon receiving the YN rankings, the donor

[26] We were not able to match some winning firms to entrants in the YN database, but collected outcome data for them anyways. Hence, the number of firms with both YN rank data and outcome data is less than this figure (7,649 firms with both Yn rank data and outcome data).

organization selects the top 10 firms ranked by YN to advance to a "finals" round. In this round, all 10 finalists are invited to present their plan live at a "pitch day" event to a new panel of judges (not related to YN). The new panel of judges then selects some subset of the finalists to award prizes or incubator space to. In this example, the threshold that is relevant for the RD is 10 - the rank necessary for advancement to the finals round. This threshold is exogenous, and beating the threshold increases the probability that a a firm wins the competition, though it does not ensure it. Thus, the threshold serves as a valid instrument for winning the competition.

Table A.2 breaks out number of competitions by number of rounds (and tracks). In 216 of the competitions, there was only 1 round, which was obviously judged by YN. I.e., the YN ranks should have been deterministic for winning a prize (or award, or placement in an accelerator) - though of course, the donor organization may have chosen to disregard the YN rankings. In this case, the threshold cut-off is the same as the number of prizes awarded.

For the other 244 competitions, some quantity of judging took place subsequent to the YN judging (for which I have data). In 206 competitions, there was a single round of judging subsequent to the YN judging – i.e. two rounds in total (column 2). In this case, the threshold cut-off is computed as the YN rank required for advancement to the final round (i.e., the number of finalists). For example, in the fictitious example above, I would set the threshold cut-off equal to 10.

In the final 38 competitions, there were 3 rounds total. I.e., YN judged for entry into a "semi-final" round, after which two more rounds of judging took place before winners were selected. In this case, the threshold cut-off is computed as the rank necessary for advancement to the semi-finals round (i.e., the number of semifinalists). In the baseline specification, I drop all 3-round competitions because the signal is likely very weak in these competitions.

The distribution of competitions and entrants by threshold cut-off is reported in Table A.3 both for the entire dataset (column 4-5) and just those firms with outcome data (columns 6-7). Thresholds range from 1 to 150, but the bulk of the competitions have threshold ranks in the single digits. In fact, 80% of competitions have thresholds under 10 (column 3). Across all 460 competitions, the mean number of entrants was 45 overall, and 17 among firms with outcome data.

# B Auxiliary Results

In this appendix, I present auxiliary results referred to in the main text.

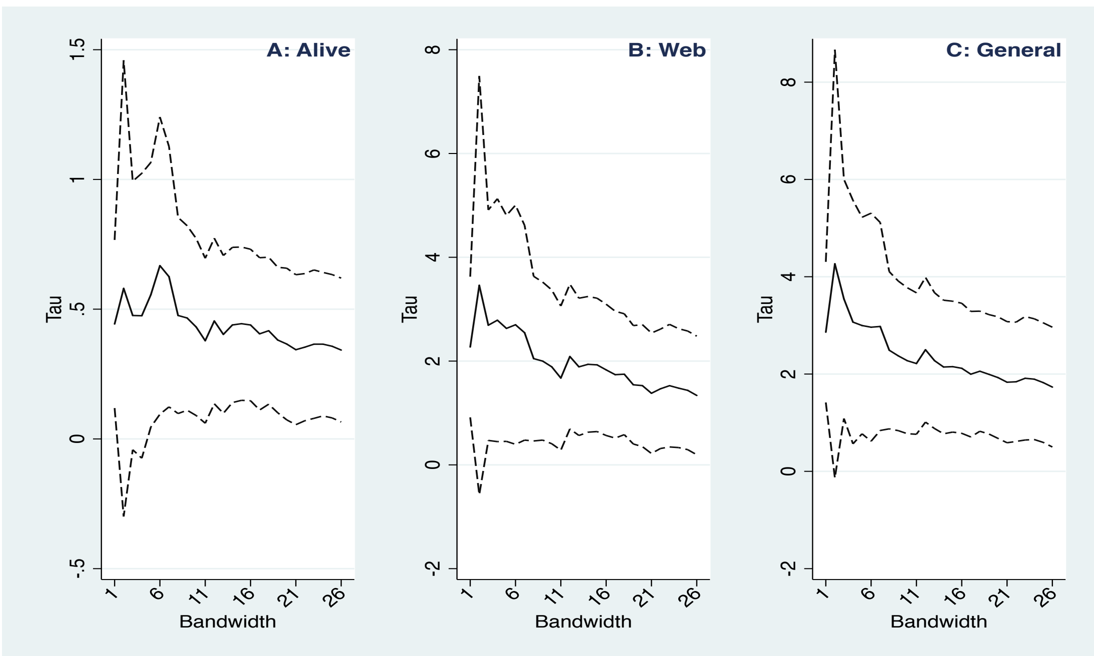
In Table B.1, I present RD results using unscaled rank $Rank_{ic}$ instead of $Rank^{sb}_{ic}$ as the running variable, and controlling for quintile of firm outcome within competition as in Howell (2017a). Qualitatively, the same pattern holds as in the preferred specification in Table 3 in the main text.

Table B.1: RD Results for Subjective Metrics - Alternative Scaling

| | First Stage | Second Stage | | | | | |
|---|---|---|---|---|---|---|---|
| | | Alive | | Web | | General | |
| | | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A : Bwidth=1* | | | | | | | |
| $1\|Rank < 0$ | 0.149*** | | | | | | |
| | (0.057) | | | | | | |
| Win | | 0.108* | 0.370 | 0.573** | 1.682 | 0.614** | 3.148* |
| | | (0.056) | (0.317) | (0.229) | (1.340) | (0.249) | (1.731) |
| # Observations | 606 | 606 | 606 | 606 | 606 | 606 | 606 |
| # Competitions | 303 | 303 | 303 | 303 | 303 | 303 | 303 |
| R squared | 0.649 | 0.699 | 0.677 | 0.703 | 0.677 | 0.679 | 0.536 |
| Mean Dep. Var | 0.338 | 0.518 | 0.518 | 1.838 | 1.838 | 1.955 | 1.955 |
| *Panel B : Bwidth=10* | | | | | | | |
| $1\|Rank < 0$ | 0.157*** | | | | | | |
| | (0.032) | | | | | | |
| Win | | 0.147*** | 0.345*** | 0.524*** | 1.316** | 0.533*** | 1.840*** |
| | | (0.021) | (0.129) | (0.087) | (0.507) | (0.077) | (0.541) |
| # Observations | 3420 | 3420 | 3420 | 3420 | 3420 | 3420 | 3420 |
| # Competitions | 404 | 404 | 404 | 404 | 404 | 404 | 404 |
| R squared | 0.283 | 0.420 | 0.395 | 0.417 | 0.392 | 0.428 | 0.352 |
| Mean Dep. Var | 0.310 | 0.505 | 0.505 | 1.829 | 1.829 | 1.931 | 1.931 |

Notes: All regressions include fixed effects for country of firm, product type, industry, competition, and quintile of rank, while panel B adds linear controls for the running variable (rank) separately on each side of the threshold. Rank corresponds to $Rank_{ic}$, not the rescaled $Rank^{sb}_{ic}$. Standard errors are clustered to allow for arbitrary correlation within super competition. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Figure B.1: Point Estimates by Bandwidth




Notes: Each subfigure plots the point estimate and 95% confidence interval for the threshold dummy $(1|Rank < 0)$ at different bandwidths. All regressions include linear controls for the running variable (rank) on each side of the threshold along with fixed effects for country of firm, product type, industry, and competition. Standard errors are clustered to allow for arbitrary correlation within super competition. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Figure B.2: Placebo Thresholds

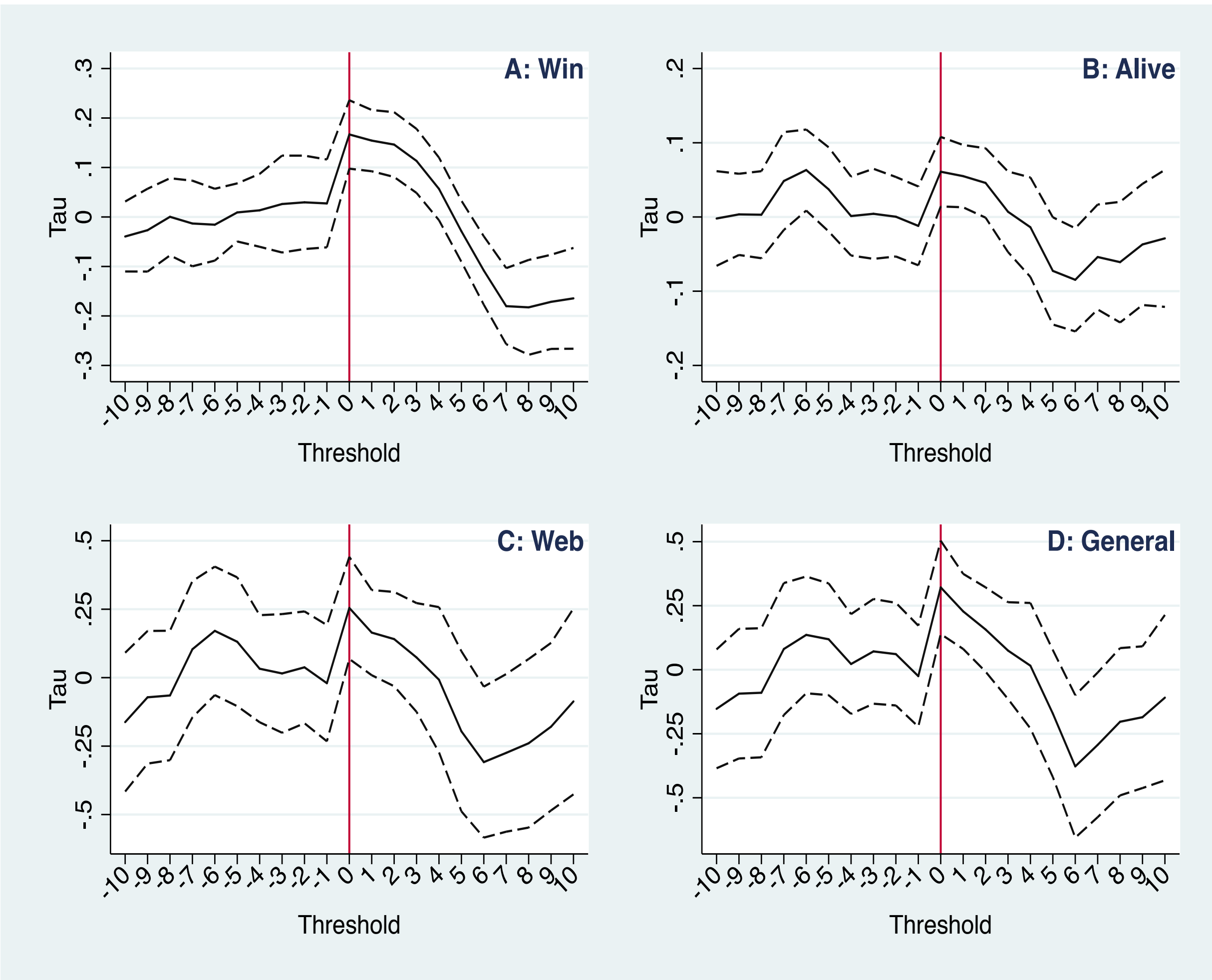


Table B.2: RD Results for Subjective Metrics Intensive Margin

| | First Stage | Second Stage | | | |
|---|---|---|---|---|---|
| | | Web | | General | |
| | | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) |
| *Panel A : Bandwidth=1* | | | | | |
| $1\|Rank < 0$ | 0.161* | | | | |
| | (0.090) | | | | |
| Win | | 0.213 | 1.429 | 0.483* | 2.401 |
| | | (0.246) | (1.234) | (0.283) | (1.651) |
| # Observations | 186 | 186 | 186 | 186 | 186 |
| # Competitions | 93 | 93 | 93 | 93 | 93 |
| R squared | 0.737 | 0.834 | 0.764 | 0.820 | 0.673 |
| Mean Dep. Var | 0.414 | 3.532 | 3.532 | 3.500 | 3.500 |
| *Panel B : Bandwidth=10* | | | | | |
| $1\|Rank < 0$ | 0.152*** | | | | |
| | (0.050) | | | | |
| Win | | 0.071 | 0.366 | 0.170** | 0.740** |
| | | (0.076) | (0.271) | (0.080) | (0.330) |
| # Observations | 1457 | 1457 | 1457 | 1457 | 1457 |
| # Competitions | 211 | 211 | 211 | 211 | 211 |
| R squared | 0.374 | 0.323 | 0.312 | 0.355 | 0.324 |
| Mean Dep. Var | 0.412 | 3.732 | 3.732 | 3.579 | 3.579 |

Notes: All regressions include fixed effects for country of firm, product type, industry, and competition, while panel B adds linear controls for the running variable (rank) separately on each side of the threshold. Standard errors are clustered to allow for arbitrary correlation within super competition. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.3: RD Results for Subjective Metrics - Early Sample

| | First Stage | Second Stage | | | | | |
|---|---|---|---|---|---|---|---|
| | | Alive | | Web | | General | |
| | | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A : Bwidth=1* | | | | | | | |
| $1\|Rank < 0$ | 0.198*** | | | | | | |
| | (0.048) | | | | | | |
| Win | | 0.120 | 0.339 | 0.750** | 2.131** | 0.652** | 2.472** |
| | | (0.080) | (0.226) | (0.321) | (0.977) | (0.311) | (0.999) |
| # Observations | 356 | 356 | 356 | 356 | 356 | 356 | 356 |
| # Competitions | 178 | 178 | 178 | 178 | 178 | 178 | 178 |
| R squared | 0.695 | 0.743 | 0.728 | 0.739 | 0.699 | 0.726 | 0.648 |
| Mean Dep. Var | 0.346 | 0.472 | 0.472 | 1.615 | 1.615 | 1.756 | 1.756 |
| *Panel B : Bwidth=10* | | | | | | | |
| $1\|Rank < 0$ | 0.165*** | | | | | | |
| | (0.046) | | | | | | |
| Win | | 0.160*** | 0.284*** | 0.580*** | 1.162*** | 0.552*** | 1.243*** |
| | | (0.024) | (0.076) | (0.102) | (0.282) | (0.093) | (0.318) |
| # Observations | 2103 | 2103 | 2103 | 2103 | 2103 | 2103 | 2103 |
| # Competitions | 236 | 236 | 236 | 236 | 236 | 236 | 236 |
| R squared | 0.330 | 0.346 | 0.337 | 0.335 | 0.322 | 0.331 | 0.310 |
| Mean Dep. Var | 0.324 | 0.463 | 0.463 | 1.698 | 1.698 | 1.775 | 1.775 |

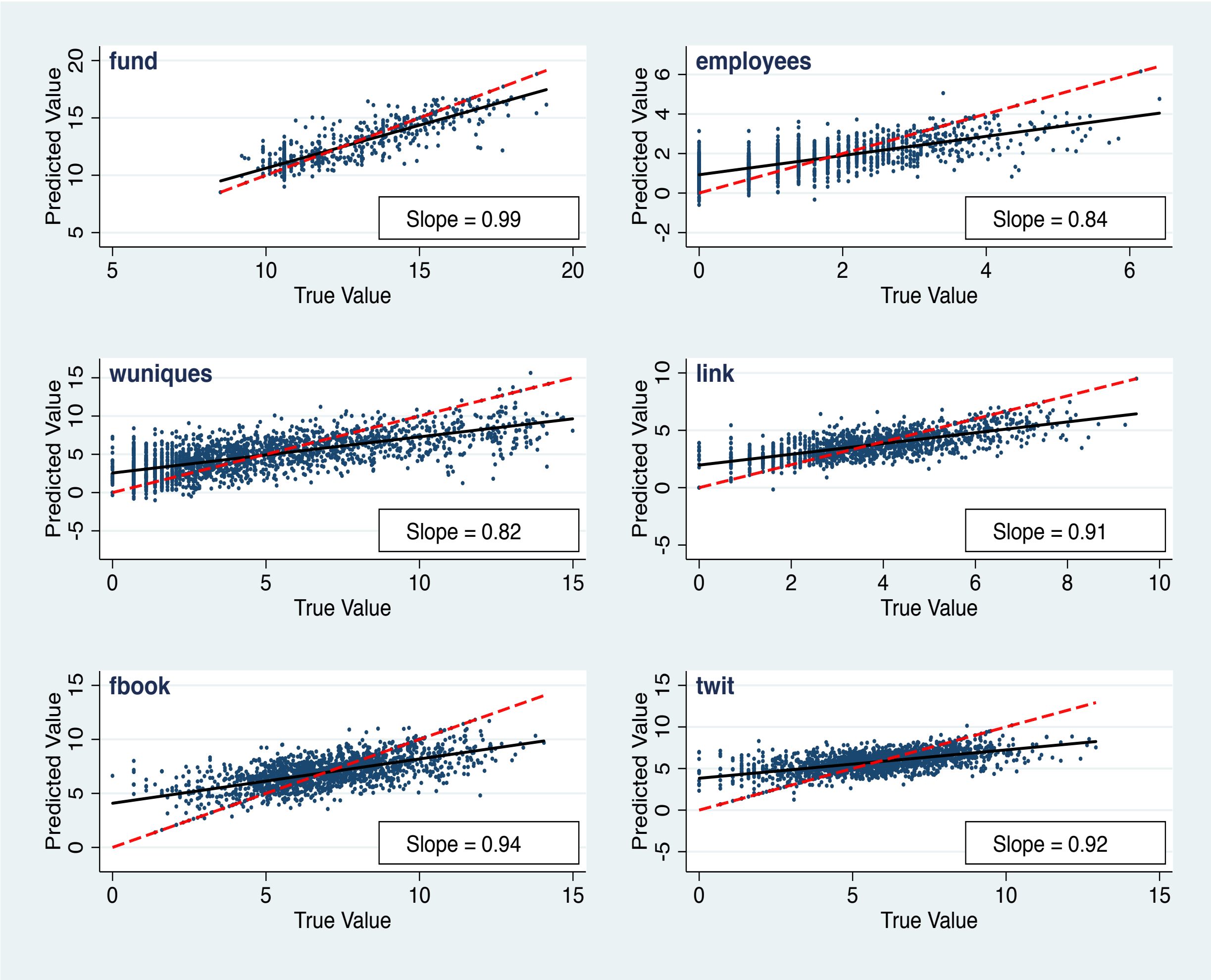
Notes: All regressions include fixed effects for country of firm, product type, industry, and competition, while panel B adds linear controls for the running variable (rank) separately on each side of the threshold. Standard errors are clustered to allow for arbitrary correlation within super competition. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.4: RD Results for Subjective Metrics - Late Sample

| | First Stage | Second Stage | | | | | |
|---|---|---|---|---|---|---|---|
| | | Alive | | Web | | General | |
| | | OLS | IV | OLS | IV | OLS | IV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A : Bwidth=1* | | | | | | | |
| $1\|Rank < 0$ | 0.327*** | | | | | | |
| | (0.073) | | | | | | |
| Win | | 0.191* | 0.560** | 0.830* | 2.209** | 0.813* | 3.119*** |
| | | (0.113) | (0.249) | (0.466) | (0.935) | (0.458) | (1.156) |
| # Observations | 228 | 228 | 228 | 228 | 228 | 228 | 228 |
| # Competitions | 114 | 114 | 114 | 114 | 114 | 114 | 114 |
| R squared | 0.666 | 0.681 | 0.631 | 0.693 | 0.650 | 0.675 | 0.554 |
| Mean Dep. Var | 0.329 | 0.561 | 0.561 | 2.035 | 2.035 | 2.136 | 2.136 |
| *Panel B : Bwidth=10* | | | | | | | |
| $1\|Rank < 0$ | 0.190*** | | | | | | |
| | (0.055) | | | | | | |
| Win | | 0.171*** | 0.377*** | 0.733*** | 1.601*** | 0.787*** | 1.995*** |
| | | (0.031) | (0.140) | (0.143) | (0.588) | (0.136) | (0.681) |
| # Observations | 1110 | 1110 | 1110 | 1110 | 1110 | 1110 | 1110 |
| # Competitions | 147 | 147 | 147 | 147 | 147 | 147 | 147 |
| R squared | 0.363 | 0.465 | 0.439 | 0.460 | 0.432 | 0.461 | 0.404 |
| Mean Dep. Var | 0.348 | 0.590 | 0.590 | 2.183 | 2.183 | 2.282 | 2.282 |

Notes: All regressions include fixed effects for country of firm, product type, industry, and competition, while panel B adds linear controls for the running variable (rank) separately on each side of the threshold. Standard errors are clustered to allow for arbitrary correlation within super competition. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Figure B.3: Subjective Metrics vs Objective Metrics




Notes: Figure plots log true Mattermark value against log projected value, where projected values are computed from equation (1). Regressing projected on true values without a constant yields solid black line with slope reported in lower right corner. Red dashed line represents 45 degree line.