THREE ESSAYS ON ENTREPRENEURIAL QUALITY

By

Jorge Guzmán

B.S. Computer Engineering
Instituto Tecnológico y de Estudios Superiores de Monterrey, 2006

M.B.A.
Massachusetts Institute of Technology, 2011

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2017

Signature of Author...... Signature redacted ....
Department of Management
January 12, 2017

Certified By................. Signature redacted ....
Scott Stern
David Sarnoff Professor of Management of Technology
Thesis Supervisor

Accepted By......... Signature redacted ....
Catherine Tucker
Sloan Distinguished Professor of Management
Professor of Marketing
Chair, MIT Sloan PhD Program

This page intentionally left blank

THREE ESSAYS ON ENTREPRENEURIAL QUALITY

by
Jorge Guzman

Submitted to the MIT Sloan School of Management
Department of Technological Innovation, Entrepreneurship and Strategic Management (TIES)

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy (Ph.D.) in Management
January, 2017

ABSTRACT

This dissertation introduces a new method to measure entrepreneurial quality and applies it to three questions of theoretical interest in entrepreneurship. The dissertation is presented in five chapters. Chapter 1 is an introductory chapter. It overviews the entrepreneurial quality approach, indicates the key choices made in this dissertation, and explains how researchers can build on this approach for their own applications. Chapters 2 through 4 are specific research applications. Chapter 2 uses this approach to measure the "state of American entrepreneurship" once quality is included. Chapter 3 studies differences in the quality of entrepreneurship by gender and the gender gap in VC financing conditional on quality. Chapter 4 looks at entrepreneurial migration and the quantity and quality of migrants in high growth entrepreneurship across regions. Chapter 5 provides a short conclusion. Extended abstracts are provided at the beginning of each chapter.

Thesis Supervisor: Scott Stern
Title: David Sarnoff Professor of Management

DEDICATION

To my daughter Juliana,
the bravest girl in the world.

To my wife Aileen,
I just love you.

To my daughter Cata and my son Manolo,
Hi, guys! *wave*
This was so much fun!

ACKNOWLEDGEMENTS

My original plan for this section was for it to be a long, heartfelt, thoughtful account, that adequately captured the rich effect this PhD had on my personal development as a scholar and individual, as well as that of my family. Alas, MIT dissertation requirements are such that acknowledgements should be limited to a single page, perhaps two, forcing me to keep this short and focused—a good trait for academics, but very uncharacteristic of a Latin-American 'thank you' section.

So, excuse me for wrestling with this form of writing. Here I go:

With two daughters at the start of my PhD, and an extra son by then end, my PhD was always a balance between family and scholarship. I experienced tremendous joy, and personal development (and some sorrow), in both, and even more in their constant interaction.

I am thankful to my wife Aileen and my children—Cata, Juliana, and Manolo—for their patience through these four and a half years, and their unconditional support in this most self-centered endeavor. Supporting your husband to embark on a PhD must be one of those acts of 'true love' that mark a loving household. Thanks, guys! I am also thankful to my extended family—my parents, my brothers, my sisters in law (my brothers' wives), my other sisters in law (Aileen's sisters), my parents in law, and all my nieces and nephews. Each of you provided very unique support, encouragement, and company; always, much beyond what I could have wished for. Most days, the only hard part of being here was being so far away from you.

I am very thankful to Scott Stern, who was always generous with his time and patience in guiding me through this process. He is a wonderful friend and mentor. We had many meetings, at odd hours, in odd days, in which we collaborated on our research agenda continuously. I learned from his example what it meant to take research seriously, to seek for the truth, to eloquently present your results, and the importance of a strong work ethic in being a productive scholar. Scott—who has many talents—is an accomplished advisor, mentoring many successful students throughout the years. His greatest skill as an advisor is helping students find projects that build on their own skillsets, and making a non-heterogeneous background an asset, not a hindrance; I am thankful for his guidance in helping me leverage my computer science background into a unique research agenda. I look forward to our work together and continued friendship.

Six months ago, my family faced a crisis: our middle daughter Juliana was diagnosed with a brain tumor. She had to undergo multiple surgeries and subsequent chemotherapy. I am very thankful for MIT's unconditional and thoughtful support in this process. Particularly Ezra Zuckerman and Scott Stern, but also Pierre Azoulay, Catherine Fazio, Olenka Kacperczyk, and Hillary Ross. Their support (and flexibility) gave me the strength I needed on my back to face these personal challenges with the necessary energy to push through them. For all of this, my family and I will always be very grateful.

During these four and a half years at MIT, I have experienced tremendous intellectual and personal growth (or so I seem to believe). This was undoubtedly due to my continued interactions with faculty—Scott, Pierre, Olenka, Christian Catalini, Ramana Nanda, Don Lessard, Michael Cusumano, Roberto Rigobon, Alberto Cavallo, Mercedes Delgado, and Fiona Murray; students—Abhishek Nagaraj, Josh Krieger, Dan Fehder, Sam Zyontz, Danny Kim, Ankur Chavda, and HyeJun Kim; and others—Cathy Fazio, Pilar Iglesias, Sarah Jane Maxted, Georgina Campbell, Tetyana Pecherska, and Isabella DiMambro. I am very thankful to all of you for your friendship, and your interest in my research and personal development.

There is much more to say, and many more people to thank. And, while some might be missing from this page, I remember everybody in my heart.

Each day in Cambridge has been special, each friendship unique. All of them building blocks off of which a happy life is built. And I am sure that what I have lived over the last few years will form steady and fruitful ground on which the rest of my life will blossom. Thanks to everyone that was part of it, and that allowed me to be part of their own lives.

I'll see you around ☺

Jorge Guzman
Cambridge, Massachusetts
2017

TABLE OF CONTENTS

CHAPTER 1

MEASURING ENTREPRENEURIAL QUALITY IN ECONOMIC RESEARCH: AN OVERVIEW

This chapter provides an introduction to entrepreneurial quality. I begin by explaining the theoretical challenge of measuring firm potential, and the recent calls to account for it. Then, I overview the entrepreneurial quality approach and the implementation of Guzman and Stern (2016b), and explain three ways in which entrepreneurial quality measures have been used in research. In the second half, I outline the key choices to be done by researchers in choosing how to implement an entrepreneurial quality approach in empirical design. And, finally, I provide my perspective on what empirical quality measures tell us about 'true' underlying firm quality.

I. INTRODUCTION

While entrepreneurship creates economic growth[1], not all firms contribute equally to this growth. The growth rate and outcomes of startups are skewed, with a small proportion of firms accounting for most economic production[2]. While economics as a field has historically attributed this skewness to a process of 'random growth' on a log-normal distribution (Gibrat, 1931; Cabral and Matta, 2003), the view amongst entrepreneurship researchers has consistently been that this variation is not random, but reflects differences in the underlying 'quality' of firms[3]. In recent work, Schoar (2010) and Hurst and Pugsley (2011) make a compelling case on the importance of accounting for this heterogeneity in ex-ante firm potential: when entrepreneurial quality is not accounted for, "many policy interventions may have unintended consequences and may even have an adverse impact on the economy" (Schoar, 2010). Notwithstanding this call to action, systematic strategies to account for firm quality appear, to date, elusive[4]. How can differences in startup potential be measured?

The purpose of this paper is to review a new approach—the entrepreneurial quality approach—that offers a novel way for researchers to characterize differences in the at-birth

---

[1] For theory and evidence documenting a positive relationship of entrepreneurship and economic growth see Schumpeter (1942), Aghion and Howitt (1992), Davis and Haltiwanger (1992), Haltiwanger et al (2013), Glaeser et al (2014), Akcigit and Kerr (2010), and Klette and Kortum (2004).

[2] This skewness is documented in several strands of research, a few of the main contributions being Dunne, Roberts, and Samuelson (1988), Hall (1987), Kaplan and Lerner (2010), and Decker et al (2014).

[3] For example, as early as 1982, Boyan Jovanovic stated that "selection matters too" (Jovanovic, 1982).

[4] For example, as recently as 2014, Hathaway and Litan responded to calls to account for this heterogeneity by stating that "The problem is that it is very difficult, if not impossible, to know at the time of founding whether or not firms are likely to survive and/or grow. This is true even with venture-capital backed firms" (Hathaway and Litan, 2014).

potential of firms by estimating the likelihood of achieving a growth outcome given these characteristics. The key insight is that, in the process of creating their firms, entrepreneurs make observable decisions—such as filing for a patent or deciding whether to register as a corporation or LLC—that correlate to their intention for the firm and its underlying potential. These observable signals can be correlated to observable ex-post outcomes, measure the predictive capacity of each signal, and create estimates of the entrepreneurial quality of firms given these observables.

This chapter is intended as a survey of entrepreneurial quality for other academics. It is general enough to serve as an overview of the topic for those who wish to learn about it, but it is also detailed enough for those who seek to implement their own version of this algorithm. To achieve both goals, this paper proceeds in three steps:

First, I begin by providing a theoretical overview of the entrepreneurial quality approach, an approach to perform out of sample tests on the quality estimates, and some aggregate statistics that can be built with quality measures.

Second, I outline a specific implementation, built in Guzman and Stern (2016b) (Chapter II of this dissertation), for 15 US states—accounting for 51% of US GDP, and test the quality of these predictive estimates. The estimates of entrepreneurial quality turn out to be highly predictive (even when the predictive algorithm, a logit, is rather simple). In multiple out of sample tests, we find that around 70% of the firms that eventually growth are in the top 5% of the predicted quality distribution, and over 40% of them in the top 1%. I also explain three ways in which these estimates can be used in economic research—as aggregate economic statistics, to control for differences in firms, and to understand selection processes. Finally, I highlight specific examples in which each has been implemented and quickly summarize the findings from those papers.

Third, in the last two sections of this paper, I outline the considerations that researchers should keep in mind when implementing their own entrepreneurial quality approach. In the first section, I focus on practical considerations, such as how to choose the correct sample, the correct observables, the correct outcome variable, and an adequate predictive model. In the second section, I try to go a little deeper to understand what these quality estimates *really mean*. Both sections emphasize the importance of alignment between the way entrepreneurial quality is measured and the theoretical question of interest—there is no one-size-fits-all solution—as well as the importance of out of sample testing of quality estimates.

The entrepreneurial quality approach is a general-purpose tool, that is likely to be useful in multiple areas of research. This chapter—as well as this dissertation—already highlights examples in macroeconomics and national statistics (Guzman and Stern, 2016b) (Chapter II), economic geography (Guzman and Stern, 2016b; Guzman, 2017) (Chapters II and IV), gender and entrepreneurship (Guzman and Kacperczyk, 2016) (Chapter III), and entrepreneurial finance (Catalini et al, 2017) (not included here). These are only a few, and many others are likely to be developed through time.

The rest of this paper is structured as follows.

Section II provides the theoretical overview of entrepreneurial quality. Section III focuses on a specific implementation of this approach (from Guzman and Stern, 2016b) using business registration records for 15 US states, representing 51% of US GDP[5], from 1988 to 2014. Section IV, highlights specific uses of this approach across a range of papers. Section V, I outline some key choices the researcher should keep in mind when implementing an entrepreneurial quality approach. Finally, in Section VI I ask what do we learn from the 'true' quality distribution of firms from empirical estimates. In Section VII, I conclude.

II. THE ENTREPRENEURIAL QUALITY APPROACH

The approach to measure entrepreneurial quality—introduced in Guzman and Stern (2015) and refined in Guzman and Stern (2016a, 2016b)[6]—combines three interrelated insights. First, as the challenges to reach a growth outcome as a sole proprietorship are formidable, a practical requirement for any entrepreneur to achieve growth is business registration (as a corporation, partnership, or limited liability company). This practical requirement allows us to form a population sample of entrepreneurs "at risk" of growth at a similar (and foundational) stage of the entrepreneurial process. Second, we can potentially distinguish among business registrants by measuring characteristics related to entrepreneurial quality observable *at or close to the time of registration.* For example, we can measure start-up characteristics such as whether the founders name the firm after themselves (eponymy), whether the firm is organized in order to facilitate

[5] Using 2013 state GDP estimates of the Bureau of Economic Analysis.
[6] This section borrows heavily, often verbatim, from these three papers, particularly Guzman and Stern (2016b).

equity financing (e.g., registering as a corporation or in Delaware), or whether the firm seeks intellectual property protection (e.g., a patent or trademark). Third, we leverage the fact that, though rare, we observe meaningful growth outcomes for some firms.

Combining these insights, we measure entrepreneurial quality by estimating the relationship between observed growth outcomes and start-up characteristics using the population of at-risk firms. Specifically, for a firm i born in region r at time t, with start-up characteristics $H_{i,r,t}$, we observe growth outcome $g_{i,r,t+s}$ s years after founding and estimate:

$$\theta_{i,r,t} = P\left(g_{i,r,t+s} \middle| H_{i,r,t}\right) = f\left(\alpha + \beta H_{i,r,t}\right) \qquad (1)$$

This model allows us to *predict* quality as the probability of achieving a growth outcome given start-up characteristics at founding, and so estimate entrepreneurial quality as $\hat{\theta}_{i,r,t}$. As long as the process by which start-up characteristics map to growth remains reasonably stable over time, this mapping allows us to form an estimate of entrepreneurial quality for any business registrant within our sample.

Assessing the Merit of Entrepreneurial Quality Estimates. To assess the merit of the estimates—rather than simply accepting them—we also propose a simple testing procedure. We define the test statistic Ψ_p as the *out-of-sample* share of growth firms that are above the percentile p of quality. A higher value of Ψ_p indicates a more informative entrepreneurial quality estimate[7]. We then use a K-fold cross validation process is used to obtain confidence intervals on this test statistic.

This approach allows us to test the estimates without having to categorize any specific firm as 'growth' or 'non-growth', working only with their quality scores, and is effective even with rare outcomes such as IPOs.

Aggregate Indexes of Entrepreneurship. We use these estimates to propose three new entrepreneurship statistics:

- *Entrepreneurial Quality Index.* Capturing the level of entrepreneurial quality for a given population of start-ups and measured simply as the mean quality of firms born in a cohort or region.
- *Entrepreneurial Potential*[8]. The potential for growth entrepreneurship within a given region and start-up cohort estimated as the mean quality times the number of

[7] And, a value of Ψ_p higher than $1 - p$ indicates the quality approach is at least better than random.
[8] Also called Regional Entrepreneurship Cohort Potential Index (RECPI)

firms in a region or cohort. This measure has the feature of also being the number of *expected* growth events for a sample of firms, given their underlying quality.

- *Regional Ecosystem Acceleration Index (REAI).* Measuring the performance over time of a regional entrepreneurial ecosystem in realizing the potential performance of firms founded within a given location and time period, and estimated as the ratio of realized growth events for the sample, over expected growth events (the entrepreneurial potential).

III. AN IMPLEMENTATION OF ENTREPRENEURIAL QUALITY WITH BUSINESS REGISTRATION RECORDS FOR 15 US STATES

We implement this approach for 15 US states in Guzman and Stern (2016b)[9]. Our analysis draws on the complete population of firms satisfying one of the following conditions: (a) a for-profit firm in the local jurisdiction or (b) a for-profit firm whose jurisdiction is in Delaware but whose principal office address is in the local state. The resulting dataset contains 18,145,359 observations.[10] For each observation we construct variables related to: (a) a growth outcome for each start-up; (b) start-up characteristics based on business registration observables; and (c) start-up characteristics based on external observables that can be linked directly to the start-up. We briefly review each one in turn and provide a more detailed summary in our data appendix.

Growth. The growth outcome utilized in this paper, Growth, is a dummy variable equal to 1 if the start-up achieves an initial public offering (IPO) or is acquired at a meaningful positive valuation within 6 years of registration[11]. During the period of 1988 to 2008, we identify 5,187 firms that achieve growth, representing 0.04% of the total sample of firms in that period.

Start-Up Characteristics. We develop two types of measures of start-up characteristics: (a) those based measures based on business registration data observable in the registration record

[9] This section, also, borrows extensively (and often, verbatim), from Guzman and Stern (2016b), which in turn also borrows from Guzman and Stern (2016a) and Guzman and Stern (2015).

[10] The number of firms founded in our sample is substantially higher than the US Census Longitudinal Business Database (LBD), done from tax records. For example, for Massachusetts in the period 2003-2012, the LBD records an average of 9,450 new firms per year and we record an average of 24,066 firm registrations. We have yet to explore the reasons for this difference. However, we expect that it may be explained, in part by: (i) partnerships and LLCs that do not have income during the year do not file a tax returns and are thus not included in the LBD, and (ii) firms that have zero employees and thus are not included in the LBD.

[11] In our Data Appendix (Section III, Table A4) we investigate changes in this measure both in the threshold of growth (e.g. only IPOs) as well as the time to grow, all results are robust to these variations

itself, and (b) measures based on external indicators of start-up quality that are observable at or near the time of business registration.

Measures Based on Business Registration Observables. We construct twelve measures based on information observable in business registration records. We first create two binary measures that relate to how the firm is registered, *Corporation*, whether the firm is a corporation rather than an LLC or partnership, and *Delaware Jurisdiction*, whether the firm is registered in Delaware. We then create two additional measures based directly on the name of the firm. *Eponymy* is equal to 1 if the first, middle, or last name of the top managers is part of the name of the firm itself.[12] *Short Name,* is equal to one if the entire firm name has three or less words, and zero otherwise.[13]

We then create several measures based on how the firm name reflects the industry or sector within which the firm is operating, taking advantage of the industry categorization of the US Cluster Mapping Project ("US CMP") (Delgado, Porter, and Stern, 2016) and a text analysis approach. We develop eight such measures. The first three are associated with broad industry sectors and include whether a firm can be identified as local (*Local*), or traded (*Traded*), or traded within resource intensive industries (*Traded Resource Intensive*). The other five industry groups are narrowly defined high technology industries that could be expected to have high growth, including whether the firm is associated with biotechnology (*Biotech Sector*), e-commerce (*E-Commerce*), other information technology (*IT Sector*), medical devices (*Medical Dev. Sector*) or semiconductors (*Semiconductor Sector*).

Measures based on External Observables. We construct two measures related to start-up quality based on intellectual property data sources from the U.S. Patent and Trademark Office. *Patent* is equal to 1 if a firm holds a patent application within the first year and 0 otherwise. We include patents that are filed by the firm within the first year of registration and patents that are assigned to the firm within the first year from another entity (e.g., an inventor or another firm). *Trademark* is equal to 1 if a firm applies for a trademark within the first year of registration.

[12] We hypothesize that eponymous firms are likely to be associated with lower entrepreneurial quality. Belenzon, Chatterji, and Daley (2014) perform a more detailed analysis of the interaction between eponymy and firm performance.

[13] Based on our review of naming patterns of growth-oriented start-ups versus the full business registration database, a striking feature of growth-oriented firms is that the vast majority of their names are at most two words (plus perhaps one additional word to capture organizational form (e.g., "Inc.")). Companies such as Akamai or Biogen have sharp and distinctive names, whereas more traditional businesses often have long and descriptive names (e.g., "New England Commercial Realty Advisors, Inc.").

Table 1 in Guzman and Stern (2017) reports the summary statistics and the source of each of the measures.

Estimation of Entrepreneurial Quality. To estimate entrepreneurial quality for each firm in our sample, we regress *Growth* on the set of start-up characteristics observable either directly through the business registration records or otherwise related to the early-stage activities of growth-oriented start-ups.

In Table 1 of this paper, we show the headline model of Guzman and Stern (2017). Columns 1 through 3 are preliminary models, while Columns 4 and 5 develop the main predictive models by including the measures in prior models plus industry controls. Our first specification (Model 4) uses only business registration observables. Corporate structure measures are particularly informative even after including other covariates. Corporations are 4.6 times more likely to grow and firms registered under Delaware jurisdiction are 46 times more likely to grow. Our two industry agnostic name-based measures are informative as well. Firms with a short name are 2.9 times more likely to grow, and eponymous firms are 73% less likely to grow. Finally, industry controls associated to particular US CMP industry clusters are significant. Firms whose names indicate inclusion in a local industry (such as "restaurant", "realtor", etc) are 29% less likely to grow, firms associated with traded industries are 14% more likely to grow, and firms specifically associated with resource intensive traded industries are 29% more likely to grow. Names associated with specific high-technology sectors are also associated with growth: firms related to biotechnology are 3.1 times more likely to grow, firm associated with e-commerce are 26% more likely to grow, firms associated with IT 2.4 times, firms associated with semiconductors 3 times more likely to grow. The relationship with firms names related to medical devices, however, is insignificant. The regression also includes state fixed-effects to account for institutional differences that changes the marginal firm that registers in each one[14]. Generally, we find the magnitudes of these fixed effects small relative to the variation that can result from firm observables, suggesting high stability across inter-region quality estimates (i.e. firms are *much* closer in their quality within a type and across states, than within a state and across types).

We extend this specification in Model 5 to include observables associated with early-stage milestones related to intellectual property. The coefficients on the business registration

[14] The point here is not that *high growth* firms would be excluded, but that some very low quality firms might or might not make it into the sample due to, for example, differences registration costs or on the ease with which registration can be done such as the possibility of online submission.

observables are quite similar (though slightly reduced in magnitude), while each of the intellectual property observables is highly predictive. Given that Delaware and Patent are highly correlated, we separate the interaction including three different effects, firms with a patent and no Delaware jurisdiction, firms with a Delaware jurisdiction and no patent, and firms with both.[15] In particular, receiving a patent is associated with a 35 times increase in the likelihood of growth for non-Delaware firms, and the combination of Delaware registration and patenting is associated with a 196 times increase in the likelihood of growth (simply registering in Delaware without a patent is associated with only a 46X increase in the growth probability). Finally, firms successfully applying for a trademark in their first year after business registration are associated with a five times increase in the probability of growth.[16]

These two models offer a tradeoff. On the one hand, the "richer" specification (Model 5) involves an inherent lag in observability, since we are only able to observe early-stage milestones in the period after business registration (in the case of the patent applications, there is an additional 18-month lag due to the disclosure policies of the USPTO). While including a more informative set of regressors, Model 5 is not as timely as Model 4: specifications that rely exclusively on information encoded within the business registration record can be calculated on a near real-time basis, and so provide the most timely index for policymakers and other analysts.[17] We calculate indices based on both specifications; while our main historical analyses are based off the results from Model 5, Model 4 can be used to provide our best estimate of changes in the last few years. Building on recent work developing real-time statistics (Scott and Varian, 2015), we use the term *nowcasting* in referring to the estimates related to Model 4 and refer to Model 5 as the "full information" model.

Robustness and Predictive Merit. Finally, in Figure 1, we evaluate the predictive quality of our estimates by undertaking a tenfold cross-validation test as outlined in the prior section[18]. This

[15] An alternative way of presenting this would be to include only an interaction for both. The Delaware and Patent coefficients would stay the same, but the joint effect would require estimating $Delaware \times Patent$ interaction rather than providing the effect directly.

[16] It is worth noting that the coefficients in these two regressions are very similar to what we found in previous research in California (Guzman and Stern, 2015a) and Massachusetts (Guzman and Stern, 2015b).

[17] It is also worthwhile to note that we can compare the historical performance of indices based on each approach – as emphasized in Figure 2 and 4, aggregate indices have a high level of concordance during the period in which a comparison is feasible, giving us some confidence in the trends predicted by the nowcasting index in the last few years.

[18] Specifically, we divide our sample into 10 random subsamples, using the first subsample as a testing sample and use the other 9 to train the model. For the retained test sample, we compare realized performance with entrepreneurial quality estimates from the model resulting from the 9 training samples. We then repeat this process 9 additional times, using each subsample as the test sample exactly once.

approach allows us to estimate average out of sample performance, as well as the distribution of out of sample test statistics for our model specification. We then report the relationship between the out-of-sample realized growth outcomes and our estimates of initial entrepreneurial quality. The share of growth firms in the top 5% of our estimated growth probability distribution ranges from 65% to 72%, with an average of 69%. The share of growth firms in the top 1% ranges from 49% to 53%, with 52% on average. To be clear, growth is still a relatively rare event even among the elite: the average firm within the top 1% of estimated entrepreneurial quality has only a 2% chance of realizing a growth outcome.

IV. RESEARCH IMPLEMENTATIONS OF ENTREPRENEURIAL QUALITY: AN OVERVIEW OF THIS THESIS

What can researchers do with estimates of entrepreneurial quality? In this section I document three applications: using estimates of entrepreneurial quality as aggregate as economic statistics, using entrepreneurial quality to control for firm heterogeneity, and using entrepreneurial quality to understand firm selection processes.

Using Estimates of Entrepreneurial Quality as Aggregate Economic Statistics. One use of of entrepreneurial quality estimates focuses in analyzing the distribution of aggregate entrepreneurial quality at the city, ZIP Code, state, or national level. In our work (Guzman and Stern, 2015, 2016a, 2016b), the interest has been to move beyond aggregate measurement of counts of new firms to include differences in the quality of firms across regions.

Using measures of entrepreneurial quality, we find substantial variation in the quality of entrepreneurship across geography and time. Often, differences in quality appear to be more meaningful: what differentiates the area of Silicon Valley from others, for example, is its noticeably high quality of entrepreneurship, rather than high quantity (Guzman and Stern, 2015); high average quality is also present in other areas usually considered highly entrepreneurial such as Cambridge, Massachusetts (Guzman and Stern, 2016a).

In Guzman and Stern (2016b), we also use aggregates of entrepreneurial quality to document the overall 'potential' of the entrepreneurship created in the United States by cohort. Figure 2 shows the estimates of overall entrepreneurship production in the United States from 1988-2012, using our model. Our results show substantial variation across time, including an upward swing in the most recent years.

Guzman and Stern (2017) use these aggregates at the regional level to describe entrepreneurship across US regions, and Guzman and Stern (2016a) even use aggregates for each single address to visualize entrepreneurship of each address in the MIT Cambridge ecosystem.

While most of our analysis is centered around the United States, this need-not be the case. Appio et al (2017) show an implementation of this approach using Spanish data, yielding equally insightful conclusions.

Using Entrepreneurial Quality to Control for Firm Heterogeneity. A second use of entrepreneurial quality estimates is to control for firm heterogeneity. Entrepreneurship research has found controlling for firm heterogeneity challenging, and hence struggles to identify 'treatment' effects in econometric analysis. Entrepreneurial quality is a measure of the firm at the time of founding. As long as the observables used to build the measures are observed at founding, estimates can be performed on differences *conditional on the observable quality at-founding*[19], making the challenges of identification more empirically tractable.

In Guzman and Kacperczyk (2016) (Chapter III of this dissertation), we use this approach to document the entrepreneurial quality of women-led and men-led firms, and to try to estimate the 'gender gap' in securing venture capital financing. Using our approach to compare pairs of firms with similar observable potential, we find that two-thirds of the gap is explained by differences in the potential of firms, while one-third remains unexplained—and can potentially be adjudicated to gender itself.

Using Entrepreneurial Quality to Understand Firm Selection Processes. Finally, entrepreneurial quality measures can allow us to understand the heterogeneity of firms in firm selection process. Guzman (2017) (Chapter IV of this dissertation) provides a good example of this. In this paper, I use these measures to understand the role of the firm's own quality in rates of entrepreneurial migration across MSAs, and compare the quality of migration to both firms at home and firms in the destination region.

As a second example, in Catalini et al (2017), we document the incidence of venture capital financing across the entrepreneurial quality distribution, to document the entrepreneurial quality of venture-backed firms.

[19] That is, a researcher can estimate what is the impact of X given two firms that appear to have the same potential at birth.

V. THE KEY CHOICES TO IMPLEMENT AN ENTREPRENEURIAL QUALITY APPROACH

Entrepreneurial quality has valuable potential as a new tool for research, but it does not come without caveats. A researcher must undertake careful choices when implementing the approach. In this section, I overview the four most important choices: the choice of outcome measure, the choice of observables, the choice of a predictive model specification, and the choice of a 'quasi-population' sample of firms (such as business registration records).

The choice of an outcome measure. The entrepreneurial quality approach can be used to develop estimates of the likelihood of *any* outcome available to the researcher. This outcome can be continuous or discrete, bounded or un-bounded. The only requirement is that it is observed systematically for all firms in a non-biased sub-sample. This gives researchers tremendous flexibility in implementing entrepreneurial quality: researchers can choose an outcome variable that is relevant to the question they are studying, and use multiple outcome variables to perform comparisons of different measures of firm quality. For example, one might want to estimate the level of 'innovativeness' of a firm, instead of its likelihood of growth. In this case, it would be possible to use a measure of innovation—such as the filing of a patent—as an outcome and develop 'innovative quality' estimates following our approach.

There are, however, some useful guidelines in choosing an outcome variable. Conditional on data availability, the choice of the optimal outcome variable is a balance of three aspects:

First, the outcome variable must be relevant for the economic analysis at hand. Different economic analyses will require different outcome variables to make a successful study, and the best outcome in some scenarios might be a poor choice in others. In Guzman and Stern (2016b), we seek to understand high growth entrepreneurship and its potential welfare impact, and use specifications with both an employment growth outcome and an equity growth outcome. However, in Guzman and Kacperczyk (2016), we seek to estimate the value of a firm to venture capitalists. In this case, a firm's ability to achieve equity growth is probably the only appropriate perspective for VCs, and we do not include any estimates of employment.

Second, the entrepreneurial quality predictive model must perform well in out of sample tests. Given that the goal is to measure differences in the underlying quality of firms, it is not enough to develop a predictive model; the researcher must test that the predictions are, in fact, informative of this quality. This is the second criteria we use in Guzman and Stern (2016b) to also include equity growth outcomes, and not only employment growth—the models we are able to estimate with our data have a better fit in tests of equity growth, and hence statements about changes in this measure are more informative.

Third, the researcher must choose the right time lag after which to observe the outcome, such as, in our model, six years after founding. In choosing the correct lag, researchers face a trade-off: they would often prefer having a long lag that allows them to see most of the firms that eventually grow, but researchers can only include in regressions firms that have 'lived' long enough to allow for the outcome to occur. For example, in Guzman and Stern (2016b) we use firm cohorts up to 2008—which we observe in 2014—for our regression. Increasing the lag to, say, 10 years would imply using only cohorts up to 2004. Our choice of six years came from careful analysis of the age distribution of growth outcomes as well as careful calibration confirming that results were similar around neighboring ages.

Choosing firm observables. The choice of firm observables is also quite permissive. In principle, researchers can include any firm observables they are able to find. As long as they take care of any strong multicollinearity, a predictive model would then automatically tell the researchers the weight of each observable, and give those observables that are less useful low weights.

However, not all observables are good observables, and researchers must defend economically the choice of observables used. In particular, they must argue convincingly that the observables used relate to later potential through underlying quality. For example, at some point in our research, as an exercise, we tested the model of Guzman and Stern after including gender dummies and fixed effects for each municipality in the dataset. Both models had higher predictive capacity. Yet, in our view, these models are not capturing underlying quality anymore, but local and institutional differences that drive differences in potential outside of firm quality. These measures make poor observables in our implementation.

Finally, whatever measure of firm observables is captured, it is critical that is observed for all firms in the sample and reasonably close to the time of firm birth, such that they can be assumed to be part of the 'at-birth' characteristics of the firms. This condition is harder to meet than, perhaps, one might naively think when approaching this problem.

Choosing the best predictive model. In the process of developing a predictive model, many different options of which model to specify become available to researchers, including least-squares, logit regression, neural networks, decision trees, regression trees, and others. In fact, any model that can be used in a predictive algorithm can be used in estimating entrepreneurial quality. The model we use, a logit regression, is one of the simplest models possible, and other models might have better in predictive accuracy. The choice of model depends, once again, on the application a researcher will have for entrepreneurial quality. In our case, the decision to use a logit was driven by two reasons:

First, while logit models are not sophisticated in prediction, they are very easy to read, and we had a direct interest in understanding the magnitudes of the coefficients of the model itself. Neural networks or non-parametric models might provide better estimates, but we would stop being able to know *what* drives these estimates—i.e. which measures are most meaningful for actual firm quality.

Second, the logit model performed reasonably well in our out of sample tests. In particular, early in our agenda we experimented with several different models including OLS and neural networks. Logit performed substantially better than OLS in out of sample tests, and performed as well as a neural network with two hidden layers. We found the value of using a neural network, therefore, unappealing.

However, this does not mean that a logit model might be best for all our applications. Specifically, our goal so far has been to advance measurements of entrepreneurial quality for none to 'good', but some applications might require a higher level of accuracy in characterizing this quality to be as close to 'perfect' as possible. In this case, other approaches might be ideal. In very early work in Guzman and Stern (2016d) we use a non-parametric approach and a well-calibrated neural network to study the true distribution of the entrepreneurial quality of firms at birth. Logit is not a good approach in this application, as it imposes a specific functional form (log-normal) on the distribution of predicted outcomes making it impossible to know the 'true' functional forms without relying on less parametric approaches. Form 'agnostic' models are, in this case, ideal.

Choosing a quasi-population of firms. Finally, business registration records is only one of multiple sample of firms that can be used to develop entrepreneurial quality measures. In most cases, it is the best sample: it is comprehensive and non-controversial requirement that all firms that seek to grow are going to do so at a foundational point in their lifetime. But it might not be available for all cases, and research projects that require less comprehensive data (for example, focused on a specific industry) might be able to use focused lists of firms as the sample of firms likely to grow. The key is that it is a *non-biased* sample of all firms with growth intention, and that these firms enter the sample at an arguably similar point in their history. In Appio et al (2016), we use the Iberian Balance Sheet Analysis System (SABI)—a division of Dunn & Bradstreet—to perform entrepreneurial quality estimates of Spain with good results.

VI. WHAT DO WE LEARN ABOUT 'TRUE' QUALITY FROM EMPIRICAL ESTIMATES?

At a more theoretical level, researchers might also ask *what exactly* is learned from entrepreneurial quality estimates. Can a predictive estimate really tell us the true potential of a firm?[20] The results in this paper, however, provide some partial answers to this question.

The first step is to realize that it is too abstract to simply refer to 'the potential of a firm'. The potential *to do what*? The answer to this question will depend on the research question of each project, and it is this underlying potential to achieve something specific that the research design should seek to capture.

What the firm might achieve is determined by the researcher through the outcome measure chosen. Empirical quality estimates are simply the likelihood of a firm achieving an outcome. How much we learn about underlying 'true' potential will depend on the correct alignment between this measure, and the underlying quality of interest. If the outcome is equity growth, then the estimated quality tells us something about the potential of the firm to achieve an equity growth outcome; if the outcome is, survivorship, then the estimated quality of the firm is its likelihood of surviving. These two might be very different[21]. Equity growth is the right outcome when taking the perspective of firm quality of venture capitalists, entrepreneurs seeking wealth, and perhaps economic value creation, but survivorship might be a better outcome in the study of local businesses development and economic opportunity.

Second, in a way, the proof is in the pudding. The entrepreneurial quality approach does not blindly offer estimates of quality for the researcher to trust, but instead asks researchers to do

[20] Another question might be what is the relationship between entrepreneurial quality estimates, and the more abstract concept of 'quality' per se? I do not claim any ability to answer this question.

[21] For example, Nanda et al (2014) document how venture-backed firms are both likely to grow a lot more than non-venture-backed, and less likely to survive in general.

adequate out of sample testing to verify that the model used has high predictive capacity. In this paper, I offer a testing approach that looks at the incidence of growth events across the entrepreneurial quality distribution. This is a sensible approach, particularly with a skewed outcome such as equity growth. But the field of predictive analytics has a number of other tests that could also be performed. In general, empirical quality estimates tell us only as much about the true potential of the firm as their predictive ability is.

Finally, researchers might also ask about the differences between shape of the *distribution* of empirical quality estimates versus the true distribution of firm quality. The answers here are exploratory at this point, but can still offer some insight. First, in a parametric model, the shape of the distribution of quality depends on the functional form imposed. A logit model imposes a lognormal distribution of predicted probabilities. This appears sensible, as a large literature has highlighted the related distribution of firm sizes to also be lognormal. However, researchers can also estimate the functional form non-parametrically to compare. In Guzman and Stern (2016d), we estimate entrepreneurial quality non-parametrically, and find that, while it is generally well approximated by a lognormal, it is more skewed in the upper tail, suggesting that it is likely a Pareto distribution.

VI. CONCLUSION

Accounting for entrepreneurial quality has been an important concern in entrepreneurship research with little empirical traction. This paper provides a comprehensive overview of a new approach, the entrepreneurial quality approach, with the goal of allowing researchers to build on this implementation in their research designs. While it makes progress over prior work, many improvements are surely yet to come, as well as novel and creative ways to use measures of quality to answer research questions. The role of entrepreneurial quality in economic research is, hopefully, only beginning.

BIBLIOGRAPHY

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 60(2): 323–51.

Akcigit, Ufuk, and William R. Kerr. 2010. "Growth Through Heterogeneous Innovations." *NBER Working Paper* #16443.

Appio, Francesco, Jorge Guzman, Scott Stern, and Antonella Martini. 2016. *"How a golden carriage turned a pumpkin: the case of the Spanish entrepreneurship"* Working Paper.

Belenzon, Sharon, Aaron Chatterji, and Brendan Daley. 2014. "Eponymous Entrepreneurs." *Working Paper.*

Cabral, Luís M. B., and José Mata. 2003. "On the Evolution of the Firm Size Distribution in an African Economy." *American Economic Review* (9): 1075–90.

Catalini, Christian, Jorge Guzman, and Scott Stern. 2017. "Venture Capital and Entrepreneurial Quality" Mimeo.

Davis, SJ, and J Haltiwanger. 1992. "Gross Job Creation , Gross Job Destruction , and Employment Reallocation." *The Quarterly Journal of Economics* 107(3): 819–63.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." *Journal of Economic Perspectives* 28(3): 3–24.

Delgado, Mercedes, Michael E. Porter, and Scott Stern. 2014. "Clusters, Convergence, and Economic Performance." *Research Policy* 43(10): 1785–99.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1988. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries." *The RAND Journal of Economics* 19(4): 495–515.

Fazio, Catherine, Jorge Guzman, Fiona Murray, and Scott Stern. "A New View of the Skew: Quantitative Estimates of the Quality and Quantity of Entrepreneurship." *MIT Innovation Initiative. Policy Brief Series.*

Gibrat, Robert. 1931. *Les Inegalite 's Economiques. Ap- Plications: Aux Ine 'galite 's Des Richesses, a La Concentration Des Entreprises, Aux Popula- Tions Des Villes, Aux Statistiques Des Familles, Etc., D'une Loi Nouvelle: La Loi de L'effect Proportionnel.* Sirley. Paris.

Glaeser, Edward L, Sari Pekkala Kerr, and William R Kerr. 2014. "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines." *Review of Economics and Statistics* 97(2): 498–520.

Guzman, Jorge. 2017 "Entrepreneurial Migration". In *Three Essays on Entrepreneurial Quality. MIT PhD Dissertation.* Chapter 4.

Guzman, Jorge, and Aleksandra Kacperczyk. 2016. "Gender Gap in Entrepreneurship." *MIT Working Paper.*

Guzman, Jorge, and Scott Stern. 2015. "Where Is Silicon Valley?" *Science* 347(6222): 606–9.

Guzman, Jorge, and Scott Stern. 2016. "Nowcasting and Placecasting Entrepreneurial Quality and Performance." *NBER/CRIW Measuring Entrepreneurial Businesses: Current Knowledge and Challenges conference* (December): 67.

Guzman, Jorge, and Scott Stern. 2016. "The State of American Entrepreneurship: New Estimates of The Quantity and Quality of Entrepreneurship for 15 US States, 1988-2014." *NBER Working Paper #22095*

Guzman, Jorge, and Scott Stern. 2016. "The Quality and Quantity of Entrepreneurship in U.S. Regions" Mimeo.

Hall, Bronwyn H. 1987. "The Relationship Between Firm Size Adn Firm Growth in the US Manufacturing Sector." *The Journal of Industrial Economics* XXXV(4): 583–606.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large vs Young." *Review of Economics and Statistics* 95(2): 347–61.

Hurst, Erik, and Benjamin W Pugsley. 2011. "What Do Small Businesses Do?" *Brookings Papers on Economic Activity* (Fall): 73–128.

Jovanovic, Boyan. 1982. "Selection and The Evolution of Industry." *Econometrica* 50(3): 649–70.

Kaplan, Steven N, and Josh Lerner. 2010. "It Ain't Broke: The Past, Present, and Future of Venture Capital." *Journal of Applied Corporate Finance* 22(2): 36–47.

Klette, Tor Jakob, and Samuel Kortum. 2004. "Innovating Firms and Aggregate Innovation." *Journal of Political Economy* 112(5): 986–1018.

Schoar, Antoinette. 2010. "The Divide between Subsistence and Transformational Entrepreneurship." *Innovation Policy and the Economy* 10(1): 57–81.

Schumpeter, Joseph. 1942. *Capitalism, Socialism and Democracy*. New York: Harper.

Scott, Steven L, and Hal R Varian. 2015. "Bayesian Variable Selection for Nowcasting Economic Time Series." In *Economic Analysis of the Digital Economy.*, eds. Avi Glodfarb, Shane Greenstein, and Catherine Tucker. , 1–22.

TABLE 1

Growth Predictive Model - Logit Regression on IPO or Acquisition within 6 years

We estimate a logit model with *Growth* as the dependent variable. Growth is a binary indicator equal to 1 if a firm achieves IPO or acquisition within 6 years and 0 otherwise. This model forms the basis of our entrepreneurial quality estimates, which are the predicted values of the model. Incidence ratios reported; Robust standard errors in parenthesis.

		Preliminary Models		***Nowcasting Model (Estimated up to real-time)***	***Full Information Model (2 year lag)***
	(1)	(2)	(3)	(4)	(5)
Corporate Governance Measures					
Corporation	6.346***			4.565***	4.055***
	(0.268)			(0.191)	(0.171)
Delaware	51.14***			40.37***	
	(1.579)			(1.297)	
Name-Based Measures					
Short Name		3.160***		2.862***	2.478***
		(0.101)		(0.0939)	(0.0836)
Eponymous		0.161***		0.270***	0.298***
		(0.0160)		(0.0270)	(0.0298)
Intellectual Property Measures					
Patent			71.97***		
			(3.249)		
Trademark			10.94***		5.014***
			(0.888)		(0.335)
Patent - Delaware Interaction					
Delaware Only					44.70***
					(3.161)
Patent Only					35.34***
					(1.257)
Patent and Delaware					196.4***
					(10.66)
US CMP Cluster Dummies					
Local				0.705***	0.755***
				(0.0432)	(0.0468)
Traded Resource Intensive				1.292***	1.283***
				(0.0507)	(0.0512)
Traded				1.145***	1.256***
				(0.0380)	(0.0426)
US CMP High-Tech Clusters					
Biotechnology				3.139***	2.288***
				(0.280)	(0.221)
E-Commerce				1.255***	1.136*
				(0.0638)	(0.0591)
IT				2.401***	1.971***
				(0.123)	(0.104)
Medical Devices				1.100	0.886
				(0.0663)	(0.0551)
Semiconductors				3.025***	1.835***
				(0.480)	(0.313)
N	12162777	12162777	12162777	12162777	12162777
pseudo R-sq	0.210	0.060	0.130	0.235	0.272

FIGURE 1

10-Fold Test of Predictive Quality of Model*

Top 1% includes 51% of growth outcomes (range: [49%, 53%])
Top 5% includes 69% of growth outcomes (range: [65%, 72%])
Top 10% includes 75% of growth outcomes (range: [70%, 79%])

*10-Fold analysis of model separates the model into 10 random samples and then uses each of those sample as a test sample. We report the average value as well as minimum and maximum (range) of such.

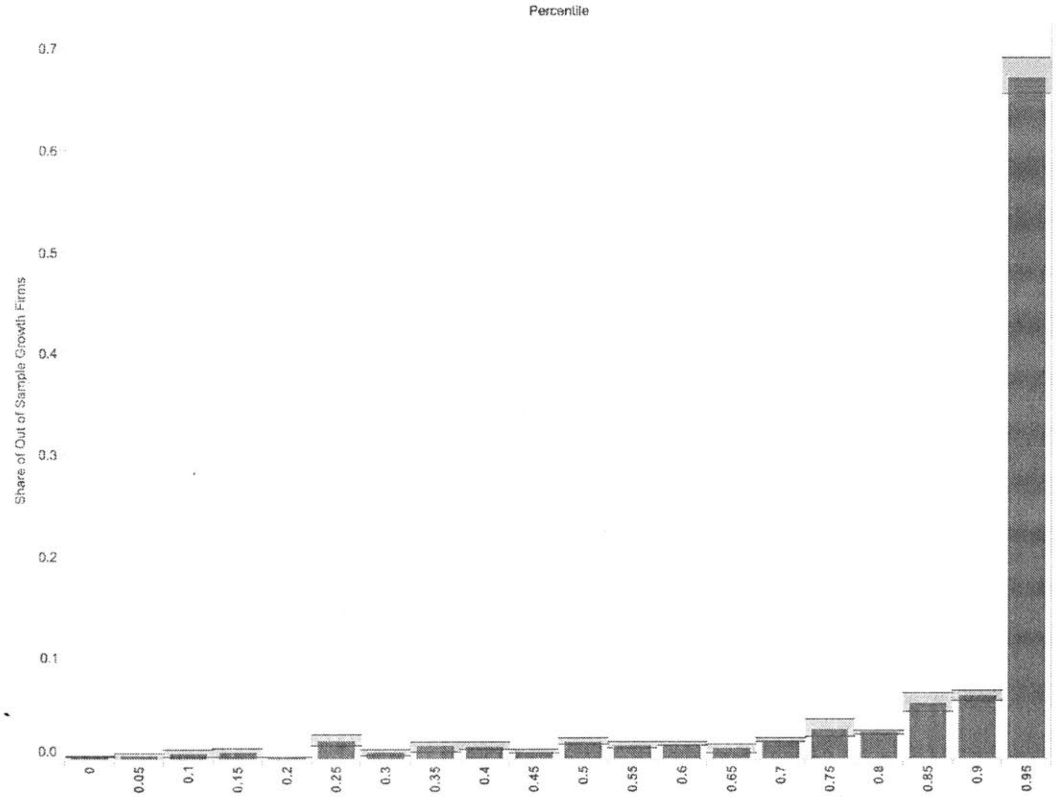
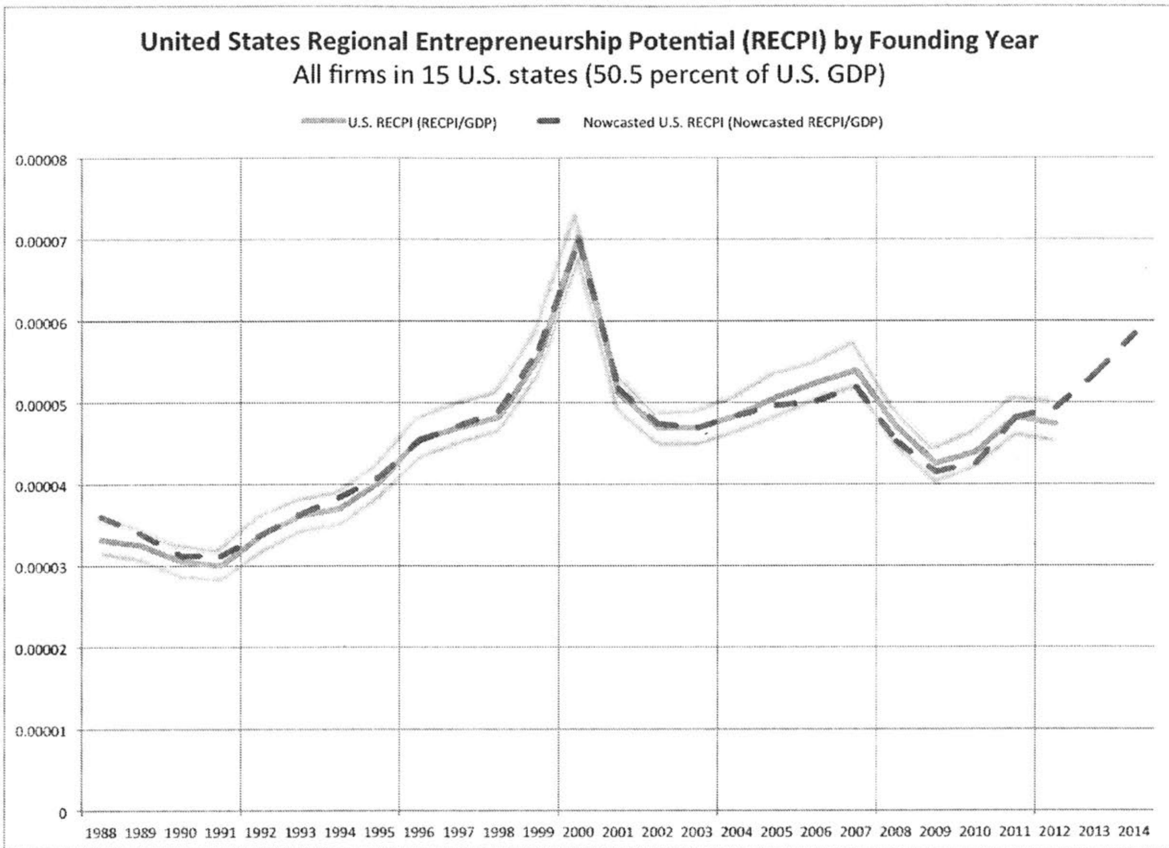


FIGURE 2



United States Regional Entrepreneurship Potential (RECPI) by Founding Year
All firms in 15 U.S. states (50.5 percent of U.S. GDP)
U.S. RECPI (RECPI/GDP)
Nowcasted U.S. RECPI (Nowcasted RECPI/GDP)
0.00008
0.00007
0.00006
0.00005
0.00004
0.00003
0.00002
0.00001
0
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

CHAPTER 2

THE STATE OF AMERICAN ENTREPRENEURSHIP: NEW ESTIMATES OF THE QUANTITY AND QUALITY OF ENTREPRENEURSHIP FOR 15 US STATES, 1988-2014

(co-authored with Scott Stern)

ABSTRACT

Assessing the state of American entrepreneurship requires not simply counting the number but also the initial quality of new ventures. Combining comprehensive business registries and predictive analytics, we present estimates of entrepreneurial quality from 1988-2014. In contrast to a long-term decline in business dynamism highlighted by prior research, our quality-adjusted statistics follow a cyclical pattern sensitive to economic and capital market conditions, register a sharp upward swing beginning in 2010, and indicate a decline over time in the propensity for high-potential firms to achieve a high-value exit. Regional variation in entrepreneurial quality is positively correlated with subsequent regional economic growth.

I. INTRODUCTION

> "There's too much entrepreneurship: Disruption running wild!" "There's too little entrepreneurship: Economy stalling out!"
>
> Marc Andreessen, Twitter, January 2015

Over the past two decades, economists have made significant progress in advancing the measurement of entrepreneurship. The pioneering studies of Haltiwanger and co-authors (Davis et al, 1996; Haltiwanger et al, 2013; Decker et al, 2014) moved attention away from simply counting the density of *small and medium sized firms* towards the measurement of the prevalence (and growth dynamics) of *young firms* (i.e., start-ups). These studies established that a disproportionate share of new job creation has historically been linked to new firms, and economic growth is grounded in measures of business dynamics (the process of firm entry, expansion, contraction and exit). A separate stream of research focusing on more selective samples of firms (e.g., high-performance entrepreneurial ventures) and the institutions (like venture capital) that surround them reinforce this perspective: for example, Kortum and Lerner (2000) find that venture capital is associated with higher levels of innovation, and Samila and Sorenson (2011) find a robust positive effect of venture capital on aggregate income, employment, and rates of new establishments.

Despite these advances, a sharp divide has emerged between systematic population-level indices of entrepreneurial activity (such as the Business Dynamics Statistics database, hereafter BDS) and measures based on the financing and activities of start-up firms, particularly in hotspots such as Silicon Valley or Cambridge. On the one hand, Hathaway and Litan (2014a; 2014b; 2014c) use the BDS to document a secular decline in the rate of business dynamism and the "aging" of US private sector establishments, a theme echoed in work emphasizing job growth dynamics such as Decker, et al (2014). This stagnation has become a key piece of evidence emphasized by those concerned with the prospects for long-term economic growth (Gordon, 2016). At the same time, a practitioner literature emphasizes the recent "explosion" of start-up activity over the past half decade, including levels of venture capital investment not observed since the late 1990s (PricewaterhouseCoopers, 2016). Not simply a matter of financing, recent research documents a striking shift in the propensity for elite undergraduate engineering students (based on a population sample of MIT graduates) to join startup firms upon graduation (Roberts, Murray, and Kim, 2015). As aptly summarized by venture capitalist Marc Andreesen, there seems to be a disconnect between population measurement of entrepreneurship and the founding of start-up firms with significant ambitions for growth at founding (Andreesen, 2015).

To put these differences in perspective, it is useful to consider the historical gap between these divergent views. In Figure 1A, we compare (for 15 US states which will form the basis for our analysis) the rate (relative to GDP) of firm births per year as measured by the Business Dynamics Statistics versus the rate (relative to GDP) of successful growth firms founded in a particular year (i.e., the number of firms founded in a given year that achieved an IPO or significant acquisition within six years of initial business registration).[1] While the BDS shows a slow and steady decline of approximately 40% (consistent with Hathaway and Litan (2014a)), the realization of growth experienced a much sharper up-and-down cycle, with 1996 representing the most successful start-up cohort in US history, followed by a relatively stable level from 2001 to 2008. This divergence is reinforced by comparing BDS firm births and economic growth. Figure 1B compares BDS firm births / GDP per year with GDP growth in the five years

[1] Though Figure 1 is based on data for only the 15 states that we use in our overall analysis, the broad patterns documented in Figure 1 are qualitatively similar if we contrast the BDS birth rate, the incidence rate of entrepreneurial growth outcomes based on cohort founding dates, and overall economic growth for the entire United States.

following each observation year. Relative to the BDS, GDP growth exhibits a sharp up-and-down pattern, with a high point beginning in 1995 (i.e., growth from 1995 to 2000).

How can we resolve this puzzle? How can we assess the State of American Entrepreneurship? Building on Guzman and Stern (2015a; 2015b), this paper breaks through this impasse by focusing not only on the quantity of entrepreneurship nor on highly selective measures of the rate of successful entrepreneurs but instead focus on the role of entrepreneurial "quality." While it has long been known that the growth consequences of start-up activity are concentrated in the outcomes associated with a very small fraction of the most successful firms (Cochrane, 2005; Kerr, Nanda, and Rhodes-Kropf, 2014), prior attempts to use population-level data to characterize the rate of entrepreneurship have largely abstracted away from initial differences across firms in the ambitions of their founders or their inherent growth potential. As emphasized by Hathaway and Litan, the challenge in directly incorporating heterogeneity is a fundamental measurement problem: "The problem is that it is very difficult, if not impossible, to know at the time of founding whether or not firms are likely to survive and/or grow." (Hathaway and Litan, 2014b). Overcoming this measurement problem in a systematic way holds promise for multiple areas of economics research. In addition to characterizing how the potential for growth entrepreneurship varies across time and place (which we emphasize in this paper), characterizing initial differences among firms can allow for novel tests of alternative theories of the firm-size distribution (e.g., testing the relative importance of initial conditions versus firm dynamics), evaluate alternative potential entrants in the context of industrial organization models, and also test for the role of factors such as regulation and financing constraints on the growth process for firms with different initial levels of underlying potential.

Our approach to measuring entrepreneurial quality combines three interrelated insights.[2] First, a practical requirement for any growth-oriented entrepreneur is business registration (as a

[2] In our earlier work, we undertook preliminary explorations of the approach that we develop in this paper. In Guzman and Stern (2015a), we introduced the overall methodology in an exploratory way by examining regional clusters of entrepreneurship such as Silicon Valley at a given point in time. We then focused on a single US state (Massachusetts) to see if it was feasible to estimate entrepreneurial quality over time on a near real-time basis (Guzman and Stern, 2015b). This paper builds on these earlier exercises to develop an analysis for 15 "representative" US states (comprising more than 50% of overall GDP) over a 30-year period, introduce new economic statistics that allow for the characterization of entrepreneurial quantity and quality over time and place, consider the relationship between alternative metrics of entrepreneurship and measures of economic performance, and consider the changing nature of regional entrepreneurship for selected metropolitan areas. Passages of text describing our methodology and approach, as well as the Data Appendix, draw upon these earlier papers (with significant revision for clarity and concision as appropriate).

corporation, partnership, or limited liability company). These public documents allow us to observe a "population" sample of entrepreneurs observed at a similar (and foundational) stage of the entrepreneurial process (in this paper, from fifteen US states comprising ~ 51% of total US economic activity over a 25-year period). Second, moving beyond simple counts of business registrants (Klapper, Amit, and Guillen, 2010), we are able to measure characteristics related to entrepreneurial quality *at or close to the time of registration.* These characteristics include how the firm is organized (e.g., as a corporation, partnership, or LLC, and whether the company is registered in Delaware), how it is named (e.g., whether the owners name the firm eponymously after themselves), and how the idea behind the business is protected (e.g., through an early patent or trademark application). These start-up characteristics may reflect choices by founders who perceive their venture to have high potential. As a result, though observed start-up characteristics are not causal drivers of start-up performance, they may nonetheless represent early-stage "digital signatures" of high-quality ventures. Third, we leverage the fact that, though rare, we observe meaningful growth outcomes for some firms (e.g., those that achieve an IPO or high-value acquisition within six years of founding), and are therefore able to estimate the relationship between these growth outcomes and start-up characteristics. This mapping allows us to form an estimate of entrepreneurial quality for any business registrant within our sample (even those in recent cohorts where a growth outcome (or not) has not yet had time to be observed).

We use this predictive analytics approach to propose three new statistics for the measurement of growth entrepreneurship: the Entrepreneurship Quality Index (EQI), the Regional Entrepreneurship Cohort Potential Index (RECPI), and the Regional Entrepreneurial Acceleration Index (REAI). EQI is a measure of *average quality* within any given group of firms, and allows for the calculation of the probability of a growth outcome for a firm within a specified population of start-ups. RECPI multiples EQI and the number of start-ups within a given geographical region (e.g., from a zip code or town to the entire five-state coverage of our sample). Whereas EQI compares entrepreneurial quality across different groups (and so facilitates apples-to-apples comparisons across groups of different sizes), RECPI allows the direct calculation of the expected number of growth outcomes from a given start-up cohort within a given regional boundary. As such, we will use RECPI relative to GDP (or "U.S.

RECPI") as our primary measure of the potential for growth entrepreneurship for a given start-up cohort. REAI, on the other hand, measures the ratio between the realized number of growth events for a given start-up cohort and the expected number of growth events for that cohort (i.e.., RECPI). REAI offers a measure of whether the "ecosystem" in which a start-up grows is conducive to growth (or not), and allows variation in ecosystem performance across time and at an arbitrary level of geographic granularity.

We calculate these measures on an annual basis for the fifteen states included in our sample for the period from 1988-2014, documenting several key findings.[3] First, in contrast to the secular and steady decline observed in the BDS, U.S. RECPI has followed a cyclical pattern that seems sensitive to the capital market environment and overall economic conditions. Second, while the peak value of U.S. RECPI is recorded in 2000, the overall level during the first decade of the 2000s is actually *higher* than the level observed between 1990 and 1995, and we additionally observe a sharp upward swing beginning in 2010. Even after controlling for change in the overall size of the economy, the third highest level of entrepreneurial growth potential is registered in 2014. Finally, there is striking variation over time in the likelihood of start-up firms for a given quality level to realize their potential (REAI): REAI declined sharply in the late 1990s, and did not recover through 2008. Though preliminary projections show some improvement after 2009, whether the most recent cohorts are able to realize their potential at rates similar to those achieved during the mid-1990s is yet to be seen.

Relative to quantity-based measures of entrepreneurship, regional variation in entrepreneurial quality appears to hold a stronger relationship to economic growth. Once one controls for the initial level of GDP, MSA-level GDP growth between 2003 and 2014 is uncorrelated with the baseline quantity of entrepreneurship but has a statistically and quantitatively significant relationship with the baseline level of entrepreneurial quality.

Finally, there is striking variation across regions (and over time) in entrepreneurial potential. Consistent with Guzman and Stern (2015a), we document an extremely high and persistent level of entrepreneurial quality in regions such as Silicon Valley (and San Francisco

[3] We use a "nowcasting" index for the most recent cohorts which only use start-up characteristics available within the business registration data, and compare that index to an "enriched" index which captures events that might occur early within the life of a start-up such as the initial receipt of intellectual property

over time) as well as the Boston region, while other regions such as Miami with a high quantity of entrepreneurship have yet to realize a meaningful level of persistent entrepreneurial quality.

Before turning to more general interpretations, we emphasize that our approach, though promising, does come with important limitations and caveats. First, and most importantly, we strongly caution against a causal interpretation of the regressors we employ for our predictive analytics – while factors such as eponymy and business registration form are a "digital signature" that allows us to differentiate among firms in the aggregate, these are not meant to be interpreted as causal factors that lead to growth per se (i.e., simply registering your firm in Delaware is not going to directly enhance an individual firm's underlying growth potential). And, while we are encouraged by the robustness of our core approach across multiple states and time periods, we can easily imagine (and are actively working on identifying) additional firm-level measures (such as founder characteristics) which might allow for even more differentiation in quality, or accounting directly for changing patterns over time and space in the "drivers" of growth. Finally, while we focus here on equity growth outcomes, we do not provide any direct measure of the potential of firms in terms of employment growth (while these are likely highly correlated, it may be the case that a much more diverse range of start-ups contributes to employment growth relative to the highly skewed nature of equity growth outcomes).

Keeping in mind these caveats, our findings nonetheless do offer a new perspective on the state of American entrepreneurship. Most importantly, our results highlight that the recent shift in attention towards young firms (pioneered by Haltiwanger and co-authors) is enriched by directly accounting for initial heterogeneity among new firms. Even within the same industry, there is significant heterogeneity among new firms in their ambition and inherent potential for growth. Policies that implicitly treat all firms as equally likely candidates for growth are likely to expect "too much" from the vast majority of firms with relatively low growth potential, and might be focusing on a lever that is only weakly related to the economic growth they often seek. Second, the striking decline in REAI after the boom period of the 1990s is the first independent evidence for an often-cited concern of practitioners – even as the number of new ideas and potential for innovation is increasing, there seems to be a reduction in the ability of companies to scale in a meaningful and systematic way. Whether this is primarily a challenge for capital markets, or reflects systematic reductions in various aspects of ecosystem efficiency remains an important challenge for future research. Finally, our results highlight that the regional variation in start-up performance reflects significant regional differences in both the underlying quality of ventures started in different locations (Silicon Valley has by far the highest EQI in the nation) and in the ability of these entrepreneurial ecosystems to nurture the scaling of high-potential companies. Systematic and real-time measurement of both of these dimensions – entrepreneurial quality and ecosystem performance – can serve as tools for policymakers and stakeholders seeking to understand the impact of entrepreneurship on economic and social progress.

The rest of this paper is organized as follows. Section II provides an overview of entrepreneurial quality in economics and briefly outlines the theoretical intuition for our approach. Section III explains our methodology. In section IV we explain our dataset and estimate entrepreneurial quality for our sample. Section V describes the geographic and time variation of entrepreneurship in the United States since 1988. Section VI compares the potential of cohorts to their performance to estimate the performance of the US entrepreneurship ecosystem in helping firms scale. In Section VII, we study the correlation between our index and future economic growth. And Section VIII studies variation of entrepreneurial quality and potential for the regions of Silicon Valley, Boston, and Miami. Section IX concludes.

II. ENTREPRENEURIAL QUALITY: DO INITIAL DIFFERENCES MATTER?

Ever since Gibrat (1931), economists have sought to understand the role of firm-specific characteristics in industry dynamics. In establishing the Law of Proportional Growth (more commonly referred to as Gibrat's Law),[4] Gibrat provided a framework in which the primary factor determining firm dynamics at a moment in time is the state of the firm at that moment in time. In other words, firm dynamics are governed by a random process (Ijiri and Simon, 1977).[5] Despite broad patterns consistent with Gibrat's Law, a large literature beginning with Mansfield (1962) instead emphasizes deviations from proportional growth. In its initial formulation, this literature emphasized that smaller firms had both higher growth rates and lower probabilities of survival (Mansfield, 1962; Acs and Audretsch (1988), among others); over time, additional research emphasized that younger firms also had high average growth rates and lower

[4] Formally, Gibrat's Law states that the growth rate of firms is independent of firm size (Gibrat's Law for Means) and that variance of the growth rate is independent of firm size (Gibrat's Law for Variances) (see Sutton, 1997 for a review).

[5] Gibrat's Law serves as the foundation for key theoretical models across multiple fields within economics (see, for example, Lucas and Prescott, 1971; Lucas, 1978; Kortum and Klette, 2004; and Luttmer, 2007).

probabilities of survival (Evans, 1987; Dunne, Roberts, and Samuelson, 1988).[6]

Davis and Haltiwanger (1992) clarified this empirical debate by considering both the role of size and age at the same time using a population-level sample from the US Census of Manufacturers. Importantly, this line of research developed a systematic empirical case that virtually all net job creation was in fact due to younger firms (which are small because they are young) rather than smaller firms per se (Davis, Haltiwanger, and Shuh, 1996). Over the last several years, population-level studies of (essentially) all US establishments have reinforced these findings, and provided new and important insight into the sources and dynamics of net new job creation (Jarmin, Haltiwanger, and Miranda, 2013). Building on these studies, Decker et al (2014) further uses this approach to document an overall decline in the rate of new business formation (with at least one employee), which the authors characterize as a reduction in the rate of business dynamism. In addition to its direct insight for our understanding of entrepreneurial dynamics, these studies have been invoked as crucial pieces of evidence in entrepreneurial policy analyses emphasizing the importance of a "shots on goal" approach that would focus on reinvigorating the overall quantity of entrepreneurship in the US economy (Hathaway and Litan, 2014a).

However, the role of young firms in shaping job creation is not homogenous across the population of new firms. The vast majority of new firms are associated with no net new job growth, and consequently a very small fraction of new firms is disproportionately responsible for net new job growth (Decker, Haltiwanger, Jarmin and Miranda, 2015). In other words, for many questions for economics research and policy, a central difficulty is being able to systematically account for "the skew": the fact that the overall ability of entrepreneurship to facilitate American economic prosperity depends disproportionately on the realized performance of a very small number of new firms. Using surveys and aggregate economic comparisons, some have suggested that these differences in growth are accounted for by underlying differences in the firms themselves (Hurst and Pugsley, 2011; Kaplan and Lerner, 2010; Schoar, 2009). Yet, systematic studies of firm dynamics have not been able to incorporate underlying differences and still consider this variation unexplained (Angelini and Generale, 2008). But how do we identify whether the economy at a given point in time is nurturing startups that have the potential for such

[6] Not simply a set of empirical regularities, these findings formed the foundations for important theoretical work, notably Jovanovic (1982) and subsequent formal model of firm and industry dynamics (Ericson and Pakes, 19995; Klepper, 1996; Hopenhayn, 1992).

growth?

Accounting for the skew requires confronting a measurement quandary: at the time that a company is founded, one cannot observe whether that particular firm will experience explosive growth (or not). On the one hand, this challenge is fundamental, since by its nature entrepreneurship involves a high level of uncertainty and luck. And, some outsized successes certainly result from unlikely origins. Ben & Jerry's, for example, was founded with the intention to be a one-store, home-made ice-cream shop.[7] With that said, there are many startups that aspire to a specific level of performance and then achieve it, including startups that we refer to as innovation-driven enterprises (IDEs), and more traditional small and medium size enterprises (SMEs) (Aulet and Murray, 2013). Across all new business starts, firms span a wide gamut in terms of their founders' ambitions and potential for growth. A very large number of new businesses aim to offer successful local services (such as a neighborhood handyman striving to build a steady book of regular clients), while others have aspirations to be the next Google or Facebook (classic IDEs). To the extent that the new firms that ultimately contribute to the skew are disproportionately drawn from IDEs with significant growth ambitions and underlying potential at their time of founding, mapping the skew requires accounting for these initial differences in a systematic way.

To accomplish this task, we take advantage of the fact that entrepreneurs themselves likely have information about their underlying idea and ambition, and make choices at the time of founding consistent with their objectives and potential for growth. In Appendix A, we develop a simple model outlining the logic of our approach. Essentially, we relate the ultimate performance of start-ups to initial early-stage choices by the entrepreneur that are also observable at or around the time of founding as a "digital signature" for each firm. By mapping the relationship between growth outcomes and these digital signatures, we are able to form an estimate of initial entrepreneurial quality. To see the intuition behind this, consider a model where all new firms have an underlying quality level q (e.g., the underlying quality of the idea or the ambition and capabilities of the founder) that is observable to the entrepreneur but not to the econometrician. Firms with a higher level of q are more likely to realize a meaningful growth

[7] As Ben Cohen of Ben & Jerry's fondly recalls: "[W]e took a $5 correspondence course in ice-cream technology and started making ice-cream in our kitchen ... When we first started, it was just a lark. We never expected to have anything more than that one home-made ice-cream shop ..." How We Met: Ben Cohen And Jerry Greenfield, Interviews by Ronna Greenstreet, INDEPENDENT, May 27, 1995. *Available at* http://www.independent.co.uk/arts-entertainment/how-we-met-ben-cohen-and-jerry-greenfield-1621559.html.

outcome g (for simplicity, we consider a binary growth outcome such as an IPO or meaningful acquisition within a given number of years after founding). In addition, all entrepreneurs face a set of binary corporate governance and strategy choices $H = \{h_1, \dots, h_N\}$, such as how to register the firm (e.g., as an LLC or corporation), what to name the firm (e.g., whether to name the firm after the founders) and how to protect their underlying idea (e.g., whether to apply for either a patent or trademark). Suppose further that while the cost of each corporate governance choice h is independent of the quality of the idea (but might vary idiosyncratically across entrepreneurs), the expected value of each of these choices is increasing in underlying quality (i.e., firms with a higher q receive a higher marginal return to each element of H). Finally, suppose that while the econometrician cannot observe underling quality, she is able to observe both the corporate governance choice bundle H^* as well as growth outcomes g. As we show in the Appendix, a mapping between g and H allows us to form a consistent estimate of the underlying probability of growth conditional on initial conditions H (we refer to this estimate as θ) and moreover show that this mapping is a monotonically increasing function of the underlying level of q.

III. THE MEASUREMENT OF ENTREPRENEURIAL QUALITY AND ECOSYSTEM PERFORMANCE INDICES

Building on this discussion, we now develop our empirical strategy. Our goal is to estimate the relationship between a growth outcome, g, and early firm choices, H^*, in order to form an estimate of the probability of growth (a θ) for all firms at their time of founding. This approach (and our discussion) builds directly on Guzman and Stern (2015a; 2015b).

We combine three interrelated insights. First, as the challenges to reach a growth outcome as a sole proprietorship are formidable, a practical requirement for any entrepreneur to achieve growth is business registration (as a corporation, partnership, or limited liability company). This practical requirement allows us to form a population sample of entrepreneurs "at risk" of growth at a similar (and foundational) stage of the entrepreneurial process. Second, we are able to potentially distinguish among business registrants through the measurement of characteristics related to entrepreneurial quality observable *at or close to the time of registration.* For example, we can measure start-up characteristics (which result from the initial entrepreneurial choices in our model) such as whether the founders name the firm after themselves (eponymy), whether the firm is organized in order to facilitate equity financing (e.g., registering as a corporation or in Delaware), or whether the firm seeks intellectual property protection (e.g., a patent or trademark). Third, we leverage the fact that, though rare, we observe meaningful growth outcomes for some firms (e.g., those that achieve an IPO or high-value acquisition within six years of founding). Combining these insights, we measure entrepreneurial quality by estimating the relationship between observed growth outcomes and start-up characteristics using the population of at-risk firms. Specifically, for a firm i born in region r at time t, with start-up characteristics $H_{i,r,t}$, we observe growth outcome $g_{i,r,t+s}$ s years after founding and estimate:

$$\theta_{i,r,t} = P(g_{i,r,t+s}|H_{i,r,t}) = f(\alpha + \beta H_{i,r,t}) \qquad (1)$$

This model allows us to *predict* quality as the probability of achieving a growth outcome given start-up characteristics at founding, and so estimate entrepreneurial quality as $\theta_{i,r,t}$. As long as the process by which start-up characteristics map to growth remain stable over time (an assumption which is itself testable), this mapping allows us to form an estimate of entrepreneurial quality for any business registrant within our sample (even those in recent cohorts where a growth outcome (or not) has not yet had time to be observed).[8]

We use these estimates to propose three new entrepreneurship statistics capturing the level of entrepreneurial quality for a given population of start-ups, the potential for growth entrepreneurship within a given region and start-up cohort, and the performance over time of a regional entrepreneurial ecosystem in realizing the potential performance of firms founded within a given location and time period.

The Entrepreneurial Quality Index (EQI). To create an index of entrepreneurial quality for any group of firms (e.g., all the firms within a particular cohort or a group of firms satisfying a particular condition), we simply take the *average* quality within that group. Specifically, in our regional analysis, we define the *Entrepreneurial Quality Index* (EQI) as an aggregate of quality at the region-year level by simply estimating the average of $\theta_{i,r,t}$ over that region:

[8] The practical requirement for estimating entrepreneurial quality in recent cohorts is the timeliness of observing the start-up characteristics, H. As in Guzman and Stern (2015b), we consider two different indices – a real-time "nowcasting" index that only includes information directly observable from the business registration form (and so can be calculated for firms as they register), and an informationally richer index that includes early-stage start-up milestones such as the acquisition or grant of a patent within the first year after founding, the granting of a trademark in the first year after founding, or mention in local media or news in the first year after founding. When one aggregates individual firm results in to aggregate indices, there is a very high level of concordance between indices based on these two approaches.

$$EQI_{r,t} = \frac{1}{N_{r,t}} \sum_{i \in \{I_{r,t}\}} \theta_{i,r,t} \quad (2)$$

where $\{I_{r,t}\}$ represents the set of all firms in region r and year t, and $N_{r,t}$ represents the number of firms in that region-year. To ensure that our estimate of entrepreneurial quality for region r reflects the quality of start-ups in that location rather than simply assuming that start-ups from a given location are associated with a given level of quality, we exclude any location-specific measures $H_{r,t}$ from the vector of observable start-up characteristics.

The Regional Entrepreneurship Cohort Potential Index (RECPI). From the perspective of a given region, the overall inherent potential for a cohort of start-ups combines both the quality of entrepreneurship in a region and the number of firms in such region (a measure of quantity). To do so, we define *RECPI* as simply $EQI_{r,t}$ multiplied by the number of firms in that region-year:

$$RECPI_{r,t} = EQI_{r,t} \times N_{r,t} \tag{3}$$

Since our index multiplies the *average* probability of a firm in a region-year to achieve growth (quality) by the number of firms, it is, by definition, the expected number of growth events from a region-year given the start-up characteristics of a cohort at birth. This measure of course abstracts away from the ability of a region to realize the performance of start-ups founded within a given cohort (i.e., its ecosystem performance), and instead can be interpreted as a measure of the "potential" of a region given the "intrinsic" quality of firms at birth, which can then be affected by the impact of the entrepreneurial ecosystem, or shocks to the economy and the cohort between the time of founding and a growth outcome.

The Regional Ecosystem Acceleration Index (REAI). While RECPI estimates the *expected* number of growth events for a given group of firms, over time we can observe the *realized* number of growth events from that cohort. This difference can be interpreted as the relative ability of firms within a given region to grow, conditional on their initial entrepreneurial quality. Variation in ecosystem performance could result from differences across regional ecosystems in their ability to nurture the growth of start-up firms, or changes over time due to financing cycles or economic conditions. We define REAI as the ratio of realized growth events to expected growth events:

$$REAI_{r,t} = \frac{\Sigma g_{i,r,t}}{RECPI_{r,t}} \tag{4}$$

A value of REAI above one indicates a region-cohort that realizes a greater than expected number of growth events (and a value below one indicates under-performance relative to expectations). REAI is a measure of a regional performance premium: the rate at which the regional business ecosystem supports high potential firms in the process of becoming growth firms.

Together, EQI, RECPI, and REAI offer researchers and regional stakeholders the ability to undertake detailed evaluations (over time, and at different levels of geographic and sectorial granularity) of entrepreneurial quality and ecosystem performance.

IV. DATA AND ENTREPRENEURIAL QUALITY ESTIMATION

Our analysis leverages business registration records, a potentially rich and systematic data for the study of entrepreneurship. Business registration records are public records created endogenously when an individual register a new business as a corporation, LLC or partnership. Section II of the data appendix in this paper provides a rich and detailed overview of this data set, as do the data appendixes in our prior work (Guzman and Stern, 2015a; 2015b).

We focus on the fifteen states of Alaska, California, Florida, Georgia, Idaho, Massachusetts, Missouri, Michigan, New York, Oklahoma, Oregon, Texas, Vermont, Washington, and Wyoming, from 1988-2014. While it is possible to found a new business without business registration (e.g., a sole proprietorship), the benefits of registration are substantial, and include limited liability, various tax benefits, the ability to issue and trade ownership shares, and credibility with potential customers. Furthermore, all corporations, partnerships, and limited liability companies must register with a Secretary of State in order to take advantage of these benefits: the act of *registering* the firm triggers the legal creation of the company. As such, these records reflect the population of businesses that take a form that is a practical prerequisite for growth.[9]

Concretely, our analysis draws on the complete population of firms satisfying one of the following conditions: (a) a for-profit firm in the local jurisdiction or (b) a for-profit firm whose

[9] This section draws on Guzman and Stern (2015a, 2015b), where we introduce the use of business registration records in the context of entrepreneurial quality estimation.

jurisdiction is in Delaware but whose principal office address is in the local state. In other words, our analysis excludes non-profit organizations as well as companies whose primary location is not in the state. Thed resulting dataset contains 18,145,359 observations.[10] For each observation we construct variables related to: (a) a growth outcome for each start-up; (b) start-up characteristics based on business registration observables; and (c) start-up characteristics based on external observables that can be linked directly to the start-up. We briefly review each one in turn and provide a more detailed summary in our data appendix.

Growth. The growth outcome utilized in this paper, Growth, is a dummy variable equal to 1 if the start-up achieves an initial public offering (IPO) or is acquired at a meaningful positive valuation within 6 years of registration[11]. During the period of 1988 to 2008, we identify 5,187 firms that achieve growth, representing 0.04% of the total sample of firms in that period.

Start-Up Characteristics. At the center of our analysis is an empirical approach to map growth outcomes to observable characteristics of start-ups at or near the time of business registration. We develop two types of measures of start-up characteristics: (a) those based measures based on business registration data observable in the registration record itself, and (b) measures based on external indicators of start-up quality that are observable at or near the time of business registration.

Measures Based on Business Registration Observables. We construct ten measures based on information observable in business registration records. We first create two binary measures that relate to how the firm is registered, *Corporation*, whether the firm is a corporation rather than an LLC or partnership, and *Delaware Jurisdiction*, whether the firm is registered in Delaware. We then create five additional measures based directly on the name of the firm. *Eponymy* is equal to 1 if the first, middle, or last name of the top managers is part of the name of the firm itself.[12] We hypothesize that eponymous firms are likely to be associated with lower

[10] The number of firms founded in our sample is substantially higher than the US Census Longitudinal Business Database (LBD), done from tax records. For example, for Massachusetts in the period 2003-2012, the LBD records an average of 9,450 new firms per year and we record an average of 24,066 firm registrations. We have yet to explore the reasons for this difference. However, we expect that it may be explained, in part by: (i) partnerships and LLCs that do not have income during the year do not file a tax returns and are thus not included in the LBD, and (ii) firms that have zero employees and thus are not included in the LBD.

[11] In our Data Appendix (Section III, Table A4) we investigate changes in this measure both in the threshold of growth (e.g. only IPOs) as well as the time to grow, all results are robust to these variations

[12] Belenzon, Chatterji, and Daley (2014) perform a more detailed analysis of the interaction between eponymy and firm

entrepreneurial quality. Our last measure relates to the structure of the firm name. Based on our review of naming patterns of growth-oriented start-ups versus the full business registration database, a striking feature of growth-oriented firms is that the vast majority of their names are at most two words (plus perhaps one additional word to capture organizational form (e.g., "Inc.")). We define *Short Name* to be equal to one if the entire firm name has three or less words, and zero otherwise.[13]

We then create several measures based on how the firm name reflects the industry or sector within which the firm is operating, taking advantage of the industry categorization of the US Cluster Mapping Project ("US CMP") (Delgado, Porter, and Stern, 2016) and a text analysis approach. We develop eight such measures. The first three are associated with broad industry sectors and include whether a firm can be identified as local (*Local*), or traded (*Traded*), or traded within resource intensive industries (*Traded Resource Intensive*). The other five industry groups are narrowly defined high technology industries that could be expected to have high growth, including whether the firm is associated with biotechnology (*Biotech Sector*), e-commerce (*E-Commerce*), other information technology (*IT Sector*), medical devices (*Medical Dev. Sector*) or semiconductors (*Semiconductor Sector*).

Measures based on External Observables. We construct two measures related to start-up quality based on intellectual property data sources from the U.S. Patent and Trademark Office. *Patent* is equal to 1 if a firm holds a patent application within the first year and 0 otherwise. We include patents that are filed by the firm within the first year of registration and patents that are assigned to the firm within the first year from another entity (e.g., an inventor or another firm). Our second measure, *Trademark*, is equal to 1 if a firm applies for a trademark within the first year of registration.

Table 1 reports the summary statistics and the source of each of the measures. A detailed description of all variables as well as the specific set of US CMP clusters used to develop each industry classification are provided in the Data Appendix.

---

performance.

[13] Companies such as Akamai or Biogen have sharp and distinctive names, whereas more traditional businesses often have long and descriptive names (e.g., "New England Commercial Realty Advisors, Inc.").

Estimation of Entrepreneurial Quality. To estimate entrepreneurial quality for each firm in our sample, we regress *Growth* on the set of start-up characteristics observable either directly through the business registration records or otherwise related to the early-stage activities of growth-oriented start-ups.

In Table 2, we present a series of univariate logit regressions of *Growth* on each of these start-up characteristics. All regressions are run on the full sample of firms from 1988 to 2008. To facilitate the interpretation of our results, we present the results in terms of the odds-ratio coefficient and include the McFadden pseudo R^2. In all our models, we use logit rather than OLS for our predictions for two reasons. First, a large literature documents firm sizes and growth rates as much closer to log-normal than linear (Gibrat, 1931; Axtell, 2001). While we stress that entrepreneurial quality is a distinct measure from firm size, it is still more natural to use a functional form that best fits the known regularities of the data.[14] Second, while OLS is known to perform better than logit in estimating marginal effects (see Angrist and Pischke, 2008), logit performs better than OLS in prediction (Pohlman and Leitner, 2003), consistent with the objective of this paper.

Our univariate results are suggestive, and highlight a relationship between early firm choices and later growth. Measures based on the firm name are statistically significant and inform variation in entrepreneurial outcomes. Having a short name is associated a 3.6X increase in the probability of growth, and having an eponymous name with an 82% *lower* probability of growth. Corporate form measures are also significant. Corporations are 3.9 times more likely to grow and firms registered under Delaware jurisdiction (instead of the local jurisdiction) are 47 times more likely to grow. These magnitudes are economically important and have strong explanatory power – the pseudo-R^2 of a Delaware binary measure alone is 0.16 – indicating a potential role of firm governance choices as a screening mechanism for entrepreneurial quality. Intellectual property measures have the highest magnitude of all groups. Firms with a patent close to their birth are 143 times more likely to grow, while firms with a trademark are 77 times

[14] While it is also possible to estimate quality non-parametrically, it leads to a "curse of dimensionality" for predictive purposes. The 14 observables we use can combine in $2^{14} = 16{,}384$ ways, not all of which have a robust number of growth firms to estimate a value. In Guzman and Stern (mimeo) we investigate the non-parametric distribution of entrepreneurial quality outside of prediction, and its implications for firm performance. We have found preliminary evidence that quality is best approximated by a Pareto distribution, rather than log-normal. We consider this an important topic for future research.

more likely to grow. Finally, the set of US CMP Cluster Dummies, implied from firm name, are also informative. Firms whose name is associated with local industries (e.g. "Taqueria") are 74% less likely to grow, while firms whose name associated with traded industries are 1.4 times more likely to grow, as are firms with names associated in specific resource intensive sectors (e.g. Oil and Gas). Firms associated with the biotechnology sector are 16 times more likely to grow, firms associated with ecommerce 1.9 times, associated to IT 6 times, medical devices 3 times, and 21 times for firms with name associated to semiconductor. These coefficients are large and highlight the value of early firm name choices as an indicator of firm intentions and signals of a firm's relationship to an industry.

It is of course important to emphasize that each of these coefficients must be interpreted with care. While we are capturing start-up characteristics that are associated with growth, we are neither claiming (or even implying) a causal relationship between the two: if a firm with low growth potential changes its legal jurisdiction to Delaware, this decision need not have any impact on its overall growth prospects.[15] Instead, Delaware registration is an informative signal, based on the fact that external investors often prefer to invest in firms governed under Delaware law, of the ambition and potential of the start-up at the time of business registration.

In Table 3, we turn to a more systematic regression analysis to evaluate these relationships. In models 1 to 3, we begin by evaluating the joint role of small groups of measures, which we then combine in models 4 and 5, which we then use as our core specifications in the estimation of entrepreneurial quality. We include state fixed effects in each of the models to account for idiosyncratic differences in corporate registration offices in each state. While it is a reasonable assumption to expect business registration records to include all firms with high quality (i.e. all firms with growth potential), it is not clear a-priori if the quality of the marginal registering firm (which is of low quality) in each state is exactly the same. In

[15] It is of course possible that use of this approach might change firm incentives if they try to "game" the algorithm by selecting into signals of high-quality (e.g., changing their name). Though real, this incentive is bounded by the objectives of the founders. For example, it is unlikely that a founder with no intention to grow would incur the significant yearly expense require to keep a registration in Delaware (which we estimate around $1000). And, firms that signal in their name that they are meant to serve a local customer base (e.g. "Taqueria") are unlikely to change their names in ways that affect their ability to attract customers. Finally, we also note that any effects from "gaming" would be short-lived since, as low quality firms select into a specific measure the correlation between such measure and growth – and therefore the weight our prediction model would assign to it – would weaken (i.e., the gaming hypothesis is testable over time).

almost all cases, however, the magnitude of fixed effects is small relative to the coefficients of our firm measures, suggesting large similarities across state registries.[16]

Columns 1-3 investigate the joint role of different groups of measures after including state fixed effects. Column 1 investigates corporate governance measures, corporations are 6.3 times more likely to grow and Delaware firms are 51 times more likely to grow. Since these are incidence-rate ratios (odds-ratios), the joint coefficients can be interpreted multiplicatively: Delaware corporations are 321 times more likely to grow ($51 \times 6.3 = 321$). Interestingly, both of these coefficients are actually larger than their respective coefficient in the univariate analysis. In column 2, we study the relationship of name-based measures to Growth. Firms with a short name are 3 times more likely to grow while eponymous firms are 84% *less* likely to grow. Finally, in column 3, we study the role of intellectual property measures to Growth. Firms with a patent are 72 times more likely to grow and firms with a trademark are 11 times more likely to grow.

In columns 4 and 5 we develop predictive models by including the measures in prior models plus industry controls. Our first specification (Model 4) uses only business registration observables. Corporate structure measures continue to be particularly informative even after including other covariates. Corporations are 4.6 times more likely to grow and firms registered under Delaware jurisdiction are 46 times more likely to grow. Our two industry agnostic name-based measures are informative as well. Firms with a short name are 2.9 times more likely to grow, and eponymous firms are 73% less likely to grow. Finally, industry controls indicating association to particular US CMP industry clusters are significant. Firms whose names indicate inclusion in a local industry (such as "restaurant", "realtor", etc) are 29% less likely to grow, firms associated with traded industries are 14% more likely to grow, and firms specifically associated with resource intensive traded industries are 29% more likely to grow. Names associated with specific high-technology sectors are also associated with growth: firms related to biotechnology are 3.1 times more likely to grow, firm associated with e-commerce are 26% more likely to grow, firms associated with IT 2.4 times, firms associated with semiconductors 3

[16] The only coefficient of an important difference in magnitude appears to be Vermont. Relative to Washington State (the excluded category), firms registered in Vermont are 90% less likely to grow. We view this result as indicative of other elements generally associated with Vermont, which is largely recognized as a highly innovative state (with the highest level of patents per capita) yet having relatively low entrepreneurial performance.

times more likely to grow. The relationship with firms names related to medical devices, however, is insignificant. Finally, the state fixed-effects show that there exists some variation in state-level corporate registration regimes, where the marginal firm to register (one that has all the negative observables and no positive ones), has different quality depending on the state. The marginal firm in California (the highest fixed-effect value) is 2.7 times higher quality than that in Washington (the reference category), while the marginal firm in Vermont (the lowest value) is 90% lower quality and Wyoming (the second lowest) is 57% lower quality. Generally, we find the magnitudes of these fixed effects small relative to the variation that can result from firm observables, suggesting high stability across inter-region quality estimates (i.e. firms are *much* closer in their quality within a type and across states, than within a state and across types).

We extend this specification in Model 5 to include observables associated with early-stage milestones related to intellectual property. The coefficients on the business registration observables are quite similar (though slightly reduced in magnitude), while each of the intellectual property observables is highly predictive. Given that Delaware and Patent are highly correlated, we separate the interaction including three different effects, firms with a patent and no Delaware jurisdiction, firms with a Delaware jurisdiction and no patent, and firms with both.[17] In particular, receiving a patent is associated with a 35 times increase in the likelihood of growth for non-Delaware firms, and the combination of Delaware registration and patenting is associated with a 196 times increase in the likelihood of growth (simply registering in Delaware without a patent is associated with only a 46X increase in the growth probability). Finally, firms successfully applying for a trademark in their first year after business registration are associated with a five times increase in the probability of growth.[18]

These two models offer a tradeoff. On the one hand, the "richer" specification (Model 5) involves an inherent lag in observability, since we are only able to observe early-stage milestones in the period after business registration (in the case of the patent applications, there is an additional 18-month lag due to the disclosure policies of the USPTO). While including a more informative set of regressors, Model 5 is not as timely as Model 4. Indeed, specifications

[17] An alternative way of presenting this would be to include only an interaction for both. The Delaware and Patent coefficients would stay the same, but the joint effect would require estimating $Delaware \times Patent$ interaction rather than providing the effect directly.

[18] It is worth noting that the coefficients in these two regressions are very similar to what we found in previous research in California (Guzman and Stern, 2015a) and Massachusetts (Guzman and Stern, 2015b).

that rely exclusively on information encoded within the business registration record can be calculated on a near real-time basis, and so provide the most timely index for policymakers and other analysts.[19] We will calculate indices based on both specifications; while our main historical analyses will be based off the results from Model 5, Model 4 can be used to provide our best estimate of changes in the last few years. Building on recent work developing real-time statistics (Scott and Varian, 2015), we use the term *nowcasting* in refererring to the estimates related to Model 4 and refer to Model 5 as the "full information" model.

Robustness and Predictive Quality. In Table 4, we repeat our nowcasting and full information models with a series of robustness tests. Since this paper uses the models to estimate quality through time and region, our main interest is to verify that the magnitudes in our model are not driven by variation across years or states. In columns 1 and 2, we repeat our models but also include year fixed effects (note that these cannot be included in our predictive model as we would not know the fixed-effect value for future years); in columns 3 and 4, we include year fixed effects and state-specific time trends. While there is some variation in the magnitude of our coefficients, the changes are relatively small, providing us confidence that our estimates are not driven by changes across years or within year and states.

Further, in Figure 2, we evaluate the predictive quality of our estimates by undertaking a tenfold cross-validation test (Witten and Frank, 2005). Specifically, we divide our sample into 10 random subsamples, using the first subsample as a testing sample and use the other 9 to train the model. For the retained test sample, we compare realized performance with entrepreneurial quality estimates from the model resulting from the 9 training samples. We then repeat this process 9 additional times, using each subsample as the test sample exactly once. This approach allows us to estimate average out of sample performance, as well as the distribution of out of sample test statistics for our model specification. We then report in Figure 2 the relationship between the out-of-sample realized growth outcomes and our estimates of initial entrepreneurial quality. The results are striking. The share of growth firms in the top 5% of our estimated growth probability distribution ranges from 65% to 72%, with an average of 69%. The share of growth firms in the top 1% ranges from 49% to 53%, with 52% on average (interestingly, these

[19] It is also worthwhile to note that we can compare the historical performance of indices based on each approach – as emphasized in Figure 2 and 4, aggregate indices have a high level of concordance during the period in which a comparison is feasible, giving us some confidence in the trends predicted by the nowcasting index in the last few years.

results are extremely similar to the findings for California from Guzman and Stern (2015)). To be clear, growth is still a relatively rare event even among the elite: the average firm within the top 1% of estimated entrepreneurial quality has only a 2% chance of realizing a growth outcome.

V. THE STATE OF AMERICAN ENTREPRENEURSHIP

With this analysis in hand, we are able to move to the centerpiece of our analysis: evaluating trends in entrepreneurial quality (EQI), entrepreneurial potential (RECPI), and regional economic performance (REAI) in the United States over time and space.

We begin by studying the trends in US entrepreneurial potential (RECPI) from 1988 to 2014. We estimate two RECPI indexes, a full information index based on (3-5) using information in intellectual property and business registration records which we simply call RECPI, and a nowcasting index that uses only business registration records (3-4), which we call Nowcasted RECPI. U.S. RECPI, reported in Figure 3, is RECPI adjusted by the aggregate yearly GDP of our sample of fifteen states (Alaska, California, Florida, Georgia, Massachusetts, Michigan, New York, Oregon, Texas, Vermont, Washington, and Wyoming). Finally, we also include a confidence interval estimated through a Monte Carlo process repeating our procedure for 100 bootstrapped random samples (i.e. with replacement) of the same size as our original sample. Before analyzing trends in the indexes, we note that both U.S. RECPI and Nowcasted U.S. RECPI move very close to each other and that the confidence interval of U.S. RECPI is narrow.

Both indexes indicate a rise of entrepreneurial potential in the 1990s through the year 2000, with a rapid drop between 2000 and 2002. However, the level observed through 2008 during the 2000s is consistently higher than the level observed during the first half of the 1990s. After a decline during the Great Recession (2008 and 2009), we observe a sharp upward spring starting in 2010.[20] Interestingly, Nowcasted U.S. RECPI is observed at its third highest level in 2014. Relative to quantity-based measures of entrepreneurship such as the BDS, these estimates seem to reflect broad patterns in the environment for growth entrepreneurship, such as capturing

[20] These broad patterns closely accord with the patterns we found for Massachusetts in Guzman and Stern (2015b).

the dot-com boom and bust of the late 1990s and early 2000s, and capturing the rise of high-growth start-up over the early years of this decade.

Our index of entrepreneurial potential does show gaps relative to realized entrepreneurial performance. Though the statistics of GDP Growth in Figure 1B as well as the number of growth firms in Figure 1A peak in the years 1995 and 1996 (respectively), U.S. RECPI instead peaks in the year 2000. This offers insight into the potential sensitivity of entrepreneurial potential to credit market cycles. While the 1996 cohort may have had lower initial potential, those firms were able to take advantage of the robust financing environment during the early years of their growth; in contrast, the peak U.S. RECPI start-up cohorts of 1999 and 2000 may have been limited in their ability to reach their potential due to the "financial guillotine" that followed the crash of the dot-com bubble (Nanda and Rhodes-Kropf, 2013, 2014).

U.S. RECPI offers a new perspective on the "state" of entrepreneurship (at least for these fifteen states). Specifically, our Nowcasting index suggests that there has been a steep rise in entrepreneurial potential over the last several years, and 2014 is the first year to begin to reach the peaks of the dot-com boom. Indeed, it is useful to recall that our measure is *relative to GDP*: on an absolute scale, U.S. RECPI 2014 is at the highest level ever registered (327 in 2014 versus the previous peak of 312 in 2000). Finally, we emphasize that, though there are small deviations, both the nowcasted and full information indexes have a very high concordance.

Geographic Variation in Entrepreneurial Quality. We also study the geographic variation in entrepreneurial quality for our 15 states. Figure 3 shows our estimate of quality in 2012 (the last year for which we have full data) by ZIP Code, with the size of each point representing to the number of firms in that ZIP Code and the color capturing its average quality (EQI) (with darker coloring indicating a higher level of entrepreneurial quality). Starting from the southwest region of the contiguous 48 states, entrepreneurship potential is clearly high in California, and is particularly high around the Bay Area. Potential drops quickly once we move into Oregon, except for a cluster of entrepreneurial quality around Portland and a smaller one around Eugene. Washington has an overall high level of quality (we are unable to estimate ZIP Code level scores as we lack addresses for our firms in Washington). Idaho and Wyoming show much less density and generally lower entrepreneurship, through there is still a small pocket of quality around

Boise (albeit much lower than the West Coast cities), and a high level of quantity (though not quality) around Cheyenne in Wyoming. Texas shows important clusters of high mass of entrepreneurship potential around Dallas and Houston, followed by Austin (a much smaller city, but of high quality). The area around San Antonio and the Rio Grande Valley shows a high number of firms but mostly low quality and the areas of El Paso and the Southern Plains (which houses important oil investments) have a smaller but visible mass of entrepreneurship potential. In Oklahoma and Missouri, it is possible to see Oklahoma City, Springfield, St. Louis, Kansas City and Columbia, all of which have low quality except for a small pocket in Columbia (where the University of Missouri is housed). In the Midwest, Michigan has small clusters of high quality around Detroit and Ann Arbor. In the Southeast, there is substantial entrepreneurship in both Florida and Georgia, thought the quality appears to be low, except, perhaps, for a slightly higher quality area around Atlanta, GA. In the Northeast, New York has a medium level of quality and we are once again unable to study micro-geography in this state as we do not have the ZIP Code of each individual firm. It is possible to appreciate the important mass of entrepreneurship potential around the Boston area, with a smaller but still visible mass around Central Massachusetts. For Vermont, there is little indication of high entrepreneurial quality across the state. Finally, Alaska shows virtually no entrepreneurship except for a very small pocked of high quality around Juneau and another south of Anchorage.

Overall, this evidence supports three interrelated conclusions. First, relative to a perspective emphasizing a worrisome secular decline in "shots on goal" (Hathaway and Litan, 2014b), our approach and evidence suggest that there has been a more variable pattern of entrepreneurship over the last 25 years, and that the last five years has been associated with an accumulation of entrepreneurial potential similar to that which marked the late 1990s. Second, this variation in potential has a clear relationship with later entrepreneurship performance of such cohorts using both measures of number of realized growth firms as well as market value created by firms in those cohorts. Finally, given the more gently sloped level of the entrepreneurial boom of recent years, it may be the case that this accumulation of entrepreneurial potential is more sustainable than earlier periods.

VI. TRENDS IN THE EFFECT OF THE US ENTREPRENEURIAL ECOSYSTEM (REAI)

Entrepreneurship performance depends on more than simply founding new enterprises, but also scaling those enterprises in a way that is economically meaningful. This insight motivates our second set of findings where we examine "ecosystem" performance across the United States, as measured by the Regional Ecosystem Acceleration Index (REAI). REAI captures the relative ability of a given start-up cohort to realize its potential, relative to the expectation for growth events as measured by RECPI (i.e., REAI = Growth Events / RECPI). A value of 1 in the index indicates no ecosystem effect. A value above 1 indicates a positive ecosystem effect, and a value under 1 indicates a negative effect. In contrast to RECPI, this index reflects the impact of the economic and entrepreneurial environment in which a start-up cohort participates (i.e., the "ecosystem" in which it participates). This ecosystem will include the location in which the firm is founded (e.g., Silicon Valley versus Miami) as well as the environment for funding and growth at the time of founding. In Figure 5, to examine the changing environment for entrepreneurship in the United States (i.e., change in the US Ecosystem, as reflected in the twelve states for which we have data), we plot REAI over time from 1988-2008, and developed a projected measure of REAI for years 2009-2012. [21]

Three distinct periods stand out. The early portion of our sample saw a significant increase in REAI from a slight negative level to a peak of 1.98 for the 1996 cohort. This is consistent with our evidence from Figure 1, in which the 1996 start-up cohort was indeed the most "successful." This peak was followed by a steady decline through 2000, in which, conditional on the estimated quality of a given start-up, the probability of growth was declining as the result of the environment (i.e., time) in which that start-up was trying to grow. From 2001-2008, there is a period of stagnation, with REAI going slowly form 0.7 down to 0.52. These differences are economically meaningful: a start-up for a given quality level is estimated to be 4 times more likely to experience a growth event in the six years after founding if they were founded in 1996 rather than in 2005. Finally, though still a preliminary estimate, we observe a weak resurgence the first increase in REAI for cohorts in 2009 to 2011, highlighting a potential improvement in the entrepreneurial ecosystem in recent years in parallel with the boom

[21] Because our approach requires that we observe the *realized* growth firms we can only measure our index with a 6 year lag, thus, up to 2008. For years 2009 to 2012, we estimate our model with a varying lag of $n = 2014 - year$ and calculate RECPI using such lag.

in the availability of entrepreneurial finance. While this rise is economically important, its realization once all growth outcomes realize is still to be seen.

This pattern is both striking and worrisome. Over the past years, there has been increasing understanding of the role that successful entrepreneurship plays as an engine for economic progress, and increased public involvement in supporting start-up activity and nurturing regional entrepreneurial ecosystems. Yet, despite that attention, the emergence from the Great Recession seems to have not been driven by (nor helped) the start-up cohorts founded in the late 2000s. Preliminary evidence shows that more recent cohorts experience a more favorable set of outcomes, but how favorable still remains an open question, and understanding the factors that facilitate more favorable outcomes for a given level of RECPI are an important agenda for future research.

VII. DO CHANGES IN ENTREPRENEURIAL QUALITY CORRELATE TO FUTURE ECONOMIC GROWTH?

We now shift our focus to the relationship between entrepreneurial quantity and quality and measures of subsequent economic performance. To do so, we build an MSA-level dataset of measures of the total quantity of entrepreneurship (OBS), EQI, as well as MSA GDP measures obtained from the Bureau of Economic Analysis. We focus on the 63 largest MSAs, each of which register more than 1000 yearly firm births on average (we include all MSAs in our geographic coverage in the robustness checks). Our core specification is a simple "long differences" analysis, in which examine the relationship between growth between 2003 and 2014 as a function of the initial level of GDP (average between 2001-2003), as well as the initial quantity and quality of entrepreneurship (both measured as an average between 2001-2003 for OBS and EQI).

Figure 6 shows the scatterplot and correlation between log GDP growth and our two entrepreneurship measures, ln (EQI) (Panel A) and $\ln(Obs)$ (Panel B). The relationship between EQI and GDP growth is positive, with a slope of .08, and significant at the 1 percent level. The relationship between quantity and GDP growth, though noisier and lower in magnitude, is also positive, with a slope of .038, and significance at the 5 percent level.

In Table 5 we measure this relationship in a regression framework. Columns 1 and 2 repeat the relationships represented graphically in Figure 6. Columns 3 and 4 include the level of GDP (ln $(GDP_{2003-2001})$) as a control. Once one accounts for initial GDP level, there is no relationship between GDP growth and the quantity of entrepreneurship.

Column 5 is our main specification, including initial levels of GDP, OBS, and EQI at the same time.[22] The results are striking. While the initial level of GDP and OBS have no relationship to subsequent GDP growth, there is a strong relationship with our measure of initial entrepreneurial quality: a doubling of entrepreneurial quality predicts an increase of 6.8% in GDP 11 years in the future. Given the skewed nature of entrepreneurial quality by region (moving a region from the 5th to the 95th percentile represents an 11X increase in quality), moving from the bottom to the top of the distribution of initial entrepreneurial quality is associated with a 75% increase in GDP growth.

Finally, in Column 6 we include all cities as a robustness test. The overall pattern is basically the same. Though the results are noisier and the coefficient for EQI slightly lower (.049 rather than .068), the coefficient is still significant at the 5% level while quantity is not distinguishable from zero.

We emphasize that these results are not causal estimates. Entrepreneurial quality (and quantity) are themselves endogenous outcomes resulting from the underlying strength and environment in a given region, and so a causal analysis would focus on whether factors shifting the environment for entrepreneurship (and resulting in an increase in OBS or EQI) could then be linked over time to overall changes in regional economic performance. With that said, these measures do provide some new insight into the relationship between entrepeneurship and economic growth. If entrepreneurial quality correlates to later economic growth, then measures of quality can serve as a useful leading indicator of the economic performance of regions. Policymakers for example can use quality-adjusted entrepreneurship index to gauge whether a particular region is encouraging the type of entrepreneurship that might yield significant economic dividends. The analysis also highlights the role of alternative indices for evaluating the role of entrepreneurship: given the focus of entrepreneurship as a pathway to economic

[22] Notably, this result also nests the relationship of RECPI and GDP. Since RECPI is defined as the product of EQI and quantity, regressing $\ln(RECPI)$ implies regressing $\ln(EQI) + \ln(Obs)$ on GDP. An unreported regression including RECPI instead of EQI in column 5 results in the exact same elasticity between RECPI and GDP than that of EQI and GDP.

performance, our analysis suggests that measures that explicitly incorporate quality are likely to accord more closely with certain types of economic phenomena.

VIII. ENTREPRENEURIAL QUALITY ACROSS METROPOLITAN AREAS

RECPI Silicon Valley: A Case Study. While our results so far have focused on the aggregate experience across fifteen (relatively diverse) US states, many questions about the state of entrepreneurship are particularly concerned with specific regional ecosystems, perhaps none more so than Silicon Valley. We therefore calculate RECPI over time solely for the combined counties of Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma, and plot the results (on an absolute scale) in Figure 7.[23] The overall pattern of results is quite similar to that of the aggregate RECPI in Figure 2, with a sharp increase in RECPI Silicon Valley during the dot-com boom, an equally sharp drop from 2000-2002, a higher but constant level through 2010, followed by a sharp increase over the last few years. While the overall directional shifts are the same, the levels are quite different. In particular, the boom in RECPI since the bottom of the Great Recession has been as steep (if not steeper) than during the late 1990s, and Nowcasted RECPI Silicon Valley is more than 50% higher than was ever realized during the dot-com boom (indeed, RECPI Silicon Valley has exceeded its dot-com peak every year since 2011). Of course, the very rapid increase in recent years may indeed be cause for concern (suggesting a bubble that, like the 1990s, cannot be sustained).

The Micro-Geography of Entrepreneurial Quality. As a final piece of analysis, we look at the changing nature of the micro-spatial distribution of *average* entrepreneurial quality (EQI) for a few key geographic areas in our sample. Figures 8A-8C show maps of EQI at the ZIP Code level, for five areas across 4 different years: the Boston metropolitan area, the San Francisco Bay area, the City of San Francisco, and the Miami metropolitan area. Each map represents a snapshot of entrepreneurial quality during the year in question. Looking across snapshots of quality for a particular city gives a sense of the evolution of the ecosystem. While one might expect each region to follow a similar pattern, we see important heterogeneity in changes in entrepreneurial quality across regions and time periods.

---

[23] While a full analysis of economic impact would properly "deflate" RECPI by the overall size of the economy (as we did in Figure 2), it is useful to consider the absolute numbers to capture the perspective of individual observers of a regional ecosystem who may be benchmarking their experience against an earlier time period.

Figure 8A shows the Boston metropolitan area. In 1988, we find entrepreneurial quality concentrated around the Route 128 corridor, a pattern documented in the detailed analyses of Massachusetts growth entrepreneurship by Saxenian (1992) and Roberts (1991). As the Boston area moves into the dot-com boom, the amount of entrepreneurial quality increases in both the central and neighboring districts while continuing to be centered around Route 128. However, over the past decade, the center of high-quality entrepreneurship has shifted. There is still high quality entrepreneurship around Route 128, but Cambridge (particularly Kendall Square) and areas of Boston (such as the Innovation District) have emerged as the leading areas in terms of intensive entrepreneurial quality in the Boston region.[24]

Figures 8B looks at the San Francisco Bay Area. First, the initial state of entrepreneurial quality in 1988 is relatively modest, with a narrow set of areas near San Jose and Sunnyvale accounting for the entirety of a "Silicon Valley" effect. The 1990s saw both an upgrade of entrepreneurial quality in the South Bay, with a boom particularly around Stanford and Berkeley. Consistent with Figure 3, the drop-off in entrepreneurial quality was much more muted after the dot-com crash than in many other places, with a particular striking rise in overall quality by 2012. More importantly, we see a shift over the past decade in the rise of entrepreneurial quality in San Francisco, extending beyond a few districts (as in 2000); by 2012, more than half of the zip codes in San Francisco registered a level of entrepreneurial quality that places them in the top 5% of the distribution of all zip codes throughout the 25-year sample period.

Beyond these hotspots, Figures 8C documents the pattern of a regions that has yet to experience the type of entrepreneurial ecosystem development as Boston or the Bay Area: Miami and its surrounding metropolitan area. In the entrepreneurial ecosystem of Miami, even during the height of the dot-com boom, there was relatively little shift in the overall entrepreneurial quality of any region, and over time, there has been an erosion of relative quality in this region. By 2012, most of the Miami area has low entrepreneurial quality (outside the top quartile). This result stands in sharp contrast to previous results that have found this same area to have the

[24] In Guzman and Stern (2015b) we have also documented this pattern of migration from Route 128 to Cambridge by estimating yearly average quality for both regions, We also document micro-geographical patterns at the level of individual addresses, highlighting the heterogeneity that exists around the "MIT Ecosystem" (e.g., comparing buildings around Kendall Square from the more retail entrepreneurship around Central Square and Cambridgeport.

highest level of self-employment (e.g. Glaeser, 2007),[25] thus highlighting the importance of focusing on quality rather than intensity of new firm formation in analyses of entrepreneurial ecosystems.

IX. CONCLUSION

Using a quality-based approach with business registration records for fifteen states, we focus on the systematic measurement of entrepreneurial quality to create synthetic entrepreneurship indexes at the national level. Not simply a matter of data, a focus on entrepreneurial quality allows us to focus on a more rigorous examination of variation over time and across places in the potential from a given start-up cohort (RECPI) and the ability of an entrepreneurial ecosystem to realize that potential over time (REAI).

This approach presents a different view into the state of American entrepreneurship, highlighting several interrelated patterns:

- The expected number of growth outcomes in the United States has followed a cyclical pattern that appears sensitive to the capital market environment and overall market conditions. U.S. RECPI reflects broad and well-known changes in the environment for startups, such as the dotcom boom and bust of the late 1990s and early 2000s.
- While the expected number of high-growth startups peaked in 2000 and then fell dramatically with the dot-com bust, starting in 2010 there is a sharp, upward swing in the expected number of successful startups formed and the accumulation of entrepreneurial potential for growth (even after controlling for the change in the overall size of the economy).
- Notwithstanding the cyclical nature of U.S. RECPI trends, U.S. RECPI has exhibited an overarching *upward* trend across the full time-series of our sample (Figure 3). The rate of expected successful startups fell to its lowest point in 1991 at a level which has not been approached again. U.S. RECPI downturns in the wake of the dotcom burst (from 2000-2004) and Great Recession (from 2007-2009) ebbed at levels significantly above its

[25] Specifically, Gleaser (2007) finds that the top three MSAs (using the 2000 Census definitions) in the United States by rates of self-employment are West Palm Beach-Boca Raton-Delray Beach, FL, Miami-Hialeah, FL, and Fort Lauderdale-Hollywood-Pompano Beach, FL. Here we use the updated 2012 MSA definitions and present the Miami-Fort Lauderdale-West Palm Beach, FL MSA, which is (basically) the same area.

1991 nadir. U.S. RECPI thus provides a strong signal that the State of American Entrepreneurship is not imperiled by a lack of formation of high-growth potential startups, but instead by other dynamics or ecosystem.

- There is striking variation in entrepreneurial potential for growth (EQI) across regions and over time. There are extremely high and persistent levels of entrepreneurial quality in areas such as Silicon Valley and Boston, while other regions with high rates of self-employment such as Miami have yet to achieve a high measured level of entrepreneurial quality.
- REAI—the likelihood of startups to reach their potential—declined sharply in the late 1990s and did not recover through at least 2008. During this time period (which preceded the Great Recession), the American ecosystem for entrepreneurship was *not* conducive to startup growth. For example, conditional on the same estimated potential, a 1996 startup was 4 times more likely to achieve a growth event in 6 years than a startup founded in 2005.
- Relative to quantity-based measures of entrepreneurship, regional variation in entrepreneurial quality appears to hold a stronger relationship to economic growth. Once one controls for the initial level of GDP, MSA-level GDP growth between 2003 and 2014 is uncorrelated with the baseline quantity of entrepreneurship but has a statistically and quantitatively significant relationship with the baseline level of entrepreneurial quality.

Our analysis thus indicates that *both* changes in entrepreneurial potential and ecosystem effects are economically important in US entrepreneurial performance. Relative to the 1990s (without the dot-com boom and bust of 1998-2002), we observe a three to four-fold drop in the US ecosystem performance while observing very little drop in overall entrepreneurial potential. Changes in both entrepreneurial potential and ecosystem effects are important for understanding the state of American entrepreneurship. While the supply of new high-potential-growth startups appears to be growing, the ability of U.S. high-growth-potential startups to commercialize and scale seems to be facing continuing stagnation.

Entrepreneurship is often identified as a key factor driving long-term economic performance, with significant policy attention and investment. To date, most entrepreneurship policy has emphasized an increase in "shots on goal" and abstracted away from significant differences across firms at founding (except for sectoral differences). However, to the extent that heterogeneity across firms matters, policy interventions to enhance the process of scale-up may be more impactful than those that simply aim to increase shots on goal. More generally, our analysis suggests that directly taking a quantitative approach to the measurement of entrepreneurial quality can yield new economic statistics to help provide a more granular analysis of entrepreneurial ecosystems and the impact of entrepreneurship on economic and social progress.

REFERENCES

Angelini, Paolo & Andrea Generale. 2008. "On the Evolution of Firm Size Distributions," *American Economic Review*, American Economic Association, vol. 98(1), pages 426-38, March.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Axtell, Robert L. 2001. "Zipf distribution of US firm sizes." *Science* 293.5536: 1818-1820.

Belenzon, Sharon, Chatterji, Aaron and Brendan Daley. 2014. "Eponymous Entrepreneurs" *Working Paper*

Cochrane, John H. (2005) "The risk and return of venture capital" *Journal of Financial Economics*. 75. 3-52

Davis, Steven, and John Haltiwanger. 1992. "Gross Job Creation, Gross Job Destruction, and Employment Reallocation". *The Quarterly Journal of Economics*. 107 (3): 819-862

Davis, Steven, Haltiwanger, John, and Scott Shuh. 1996. "Job Creation and Destruction". *MIT Press*.

Decker, Ryan, Haltiwanger, John, Jarmin, Ron and Javier Miranda. 2014. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism". *Journal of Economic Perspectives*. 28(3): 3-24

Decker, Ryan, John Haltiwanger, Ron Jarmin and Javier Miranda. 2015. "Where Has All The Skewness Gone? The Decline in High-Growth (Young) Firms". *NBER Working Paper #21776*.

Delgado, Mercedes, Michael E. Porter and Scott Stern. 2016. "Defining Clusters of Related Industries" *Journal of Economic Geography*. 16 (1): 1-38

Dunne, Timothy, Mark J. Roberts amd Larry Samuelson, 1988. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries," *RAND Journal of Economics*, The RAND Corporation, vol. 19(4), pages 495-515, Winter.

Evans, David S. 1987. "The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries". *The Journal of Industrial Economics* 35 (4). Wiley: 567–81. doi:10.2307/2098588.

Gibrat, Robert. 1931. *Les Inégalités Économiques*, Paris, Librairie du Recueil Sirey

Glaeser, Edward. 2007. "Entrepreneurship and the City". *NBER Working Paper #13551.*

Glaeser, Edward, Sari Pekkala Kerr, and William R. Kerr. 2012. "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines". *NBER Working Paper #18333*

Guzman, Jorge, and Scott Stern. 2015a. "Where is Silicon Valley?" *Science*. Vol. 347. Issue #6222.

Guzman, Jorge, and Scott Stern. 2015b. "Nowcasting and Placecasting Entrepreneurial Quality and Performance" *NBER Working Paper #20954*

Guzman, Jorge, and Scott Stern. 2016. "Zipf's Law in Entrepreneurial Quality" *mimeo*

Haltiwanger, John, Ron Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young" *The Review of Economics and Statistics*. 95 (2). 347-361.

Hathaway, Ian, and Robert E. Litan. 2014a. "Declining Business Dynamism in the United States: A Look at States and Metros". *Economic Studies at Brookings Series.* May, 2014.

Hathaway, Ian, and Robert E. Litan. 2014b. "Declining Business Dynamism: It's For Real," *Economic Studies at Brookings Series.* May, 2014.

Hathaway, Ian, and Robert E. Litan. 2014c. "The Other Aging of America: The Increasing Dominance of Older Firms," *Economic Studies at Brookings Series.* July, 2014

Holtz-Eakin, Douglas, Whitney Newey and Harvey S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data" *Econometrica*. Vol. 56, No. 6 pp. 1371-1395

Hurst, Erik & Benjamin Wild Pugsley, 2011. "What do Small Businesses Do?," *Brookings Papers on Economic Activity, Economic Studies Program,* The Brookings Institution, vol. 43(2 (Fall)), pages 73-142

Ijiri, Yuri, and Herbert Simon. 1977. *Skew Distributions and the Sizes of Business Firms.* 231 pages.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry", *Econometrica*, 50, issue 3, p. 649-70

Kaplan, Steven, and Josh Lerner. 2010. "It Ain't Broke: the Past, Present, and Future of Venture Capital". *Journal of Applied Corporate Finance.* 22 (2). 36-47

Kerr, William, Nanda, Ramana, and Matthew Rhodes-Kropf. 2014. "Entrepreneurship as Experimentation". *Journal of Economic Perspectives.* 28 (3): 25-48.

Kortum, Samuel, and Josh Lerner. 2000. "Assessing the contribution of venture capital to innovation" *RAND Journal of Economics.* 31 (4). 674-692

Lerner, Josh. 2009. *"Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failed--and What to Do About It"* Princeton University Press. 240 pages.

Lucas Jr, Robert E., and Edward C. Prescott. 1971. "Investment under uncertainty." *Econometrica: Journal of the Econometric Society*: 659-681.

Lucas Jr, Robert E. 1978. "On the size distribution of business firms." *The Bell Journal of Economics*: 508-523.

Luttmer, Erzo G. J. 2007. "Selection, Growth, and the Size Distribution of Firms" *The Quarterly Journal of Economics* 122 (3): 1103-1144.doi: 10.1162/qjec.122.3.1103

Mansfield, Edwin. 1962. "Entry, Gibrat's Law, Innovation, and the Growth of Firms." *American Economic Review*. 52(5), pp. 1023–51.

McFadden, Daniel. 1974. "Conditional logit analysis of qualitative choice behavior". *Frontiers in Econometrics. Chapter 4*. Academic Press. p. 105-142.

Nanda, Ramana, and Matthew Rhodes-Kropf. 2014. "Financing Risk and Innovation". HBS Working Paper 11-013.

PricewaterhouseCoopers. 2016. "PwC/NVCA Money Tree™ Report". Retrieved from https://www.pwcmoneytree.com/HistoricTrends/CustomQueryHistoricTrend on March 2, 2016.

Reif, Rafael. 2015. "A better way to deliver innovation to the world". *The Washington Post*. Opinions. May 22, 2015.

Roberts, Edward. 1991. *Entrepreneurs in High Technology: Lessons from MIT and Beyond*. Oxford University Press. 1st Edition.

Sutton, John. 1997. "Gibrat's Legacy" *Journal of Economic Literature*. Vol. XXXV March pp. 40-59

Samila, Sampsa, and Olav Sorenson. 2011. "Venture Capital, Entrepreneurship, and Economic Growth". *The Review of Economics and Statistics*. 93(1). 338-349.

Saxenian, Anna Lee. 1992. "Regional Advantage: Culture and Competition in Silicon Valley and Route 128". Harvard University Press.

Schoar, Antoinette. 2010. "The Divide between Subsistence and Transformational Entrepreneurship" *Chapter in NBER book Innovation Policy and the Economy, Volume 10*. Edited by Josh Lerner and Scott Stern. p. 57 - 81

Witten, Ian H., and Eibe Frank. 2005. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

UPSTART Business Journal. 2014. "The great VC bubble fight of 2014: Tim Draper says yes, Marc Andreessen says no". January 13, 2014. *Accessed on June 5, 2015*. http://upstart.bizjournals.com/money/loot/2014/01/13/tim-draper-says-bubble-marc.html

TABLE 1

Variable Definition and Summary Statistics (1988-2014) (1)

(1) All variables are dummy variables with values of 0 or 1. A detailed description of how each measure is built is available in our data appendix, as well as in the main paper (in a less detailed manner).
(2) US CMP Cluster Dummies are estimated by using a sample of 10M firms and comparing the incidence of each word in the name within and outside a cluster, then selecting the words that have the highest relative incidence as informative of a cluster. Firms get a value of 1 if they have any of those words in their name. The procedure is explained in detail in our Data Appendix.
(3) Note that there are also firms that we cannot associate with local nor traded industries.
(4) All values for mean and standard deviation are presented as percentage values for ease of exposition.

	Definition	Source	Mean (4)	Std Dev
Outcome Variable				
Growth	1 if a firm achieves an equity growth outcome (IPO or acquisition) within 6 years or less, 0 otherwise.	SDC Platinum	0.0003	0.0177
Corporate Form Observables				
Corporation	1 if a firm is registered as corporation, 0 if registered as LLC, or partnership.	Bus. Reg. Records	0.5318	0.4990
Delaware	1 if registered under Delaware jurisdiction, 0 if registered under local (focal state) jurisdiction	Bus. Reg. Records	0.0281	0.1652
Name Observables				
Short Name	1 if the firm name is two words or less, 0 otherwise.	Bus. Reg. Records	0.4598	0.4984
Eponymous	1 if first or last name of top manager (president, CEO, partner) is part of firm name, 0 otherwise.	Bus. Reg. Records	0.0981	0.2975
Intellectual Property Observables				
Patent	1 if firm obtains a patent within a year of founding (either application of new patent or assignment of existing patent), 0 otherwise.	USPTO	0.0018	0.0420
Trademark	1 if firm obtains a trademark within a year of founding, 0 otherwise.	USPTO	0.0012	0.0350
US CMP Cluster Dummies (2)				
Local	1 if firm name is associated to local industries, 0 otherwise.	Estimated from name	0.1877	0.3905
Traded (3)	1 if firm name is associated to traded industries, 0 otherwise.	Estimated from name	0.5451	0.4980
Traded Resource Int.	1 if firm name is associated to resource intensive industries, 0 otherwise.	Estimated from name	0.1374	0.3443
Biotech Sector	1 if firm name is associated to industries in the biotechnology sector, 0 otherwise.	Estimated from name	0.0020	0.0443
Ecommerce Sector	1 if firm name is associated to industries in the ecommerce sector, 0 otherwise.	Estimated from name	0.0491	0.2160
IT Sector	1 if firm name is associated to industries in the IT sector, 0 otherwise.	Estimated from name	0.0221	0.1470
Medical Dev. Sector	1 if firm name is associated to industries in the medical devices sector, 0 otherwise.	Estimated from name	0.0288	0.1673
Semiconductor Sector	1 if firm name is associated to industries in the semiconductor sector, 0 otherwise.	Estimated from name	0.0005	0.0215
Observations			18,145,359	

TABLE 2

Logit Univariate Regressions

Logit univariate regressions of *Growth* (IPO or Acquisition within 6 years) with each of the observables we develop for our dataset. Incidence rate ratios reported; Standard errors in parentheses * p<0.05 ** p<0.01 *** p<0.001

Firm Name Measures:

Variable	*Univariate Coefficient*	*Pseudo R2*
Short Name	3.608***	0.021
	(0.116)	
Eponymous	0.179***	0.006
	(0.0177)	

Corporate Form Measures:

Variable	*Univariate Coefficient*	*Pseudo R2*
Corporation	3.933***	0.017
	(0.162)	
Delaware	46.93***	0.157
	(1.318)	

IP Measures:

Variable	*Univariate Coefficient*	*Pseudo R2*
Patent	142.7***	0.093
	(4.926)	
Trademark	76.41***	0.030
	(3.968)	
Observations	12162777	

US CMP Cluster Dummies:

Variable	*Univariate Coefficient*	*Pseudo R2*
Local	0.261***	0.008
	(0.0157)	
Traded Resource Intensive	1.321***	0.001
	(0.0478)	
Traded	1.428***	0.002
	(0.0412)	
Biotech Sector	16.16***	0.006
	(1.331)	
Ecommerce Sector	1.896***	0.002
	(0.0899)	
IT Sector	5.988***	0.013
	(0.248)	
Medical Dev. Sector	3.017***	0.004
	(0.150)	
Semiconductor Sector	20.74***	0.002
	(2.932)	
Observations	12162777	

TABLE 3

Growth Predictive Model - Logit Regression on IPO or Acquisition within 6 years

We estimate a logit model with *Growth* as the dependent variable. Growth is a binary indicator equal to 1 if a firm achieves IPO or acquisition within 6 years and 0 otherwise. This model forms the basis of our entrepreneurial quality estimates, which are the predicted values of the model. Incidence ratios reported; Robust standard errors in parenthesis.

				Nowcasting Model (Estimated up to real-time)	***Full Information Model (2 year lag)***
		Preliminary Models			
	(1)	(2)	(3)	(4)	(5)
Corporate Governance Measures					
Corporation	6.346***			4.565***	4.055***
	(0.268)			(0.191)	(0.171)
Delaware	51.14***			40.37***	
	(1.579)			(1.297)	
Name-Based Measures					
Short Name		3.160***		2.862***	2.478***
		(0.101)		(0.0939)	(0.0836)
Eponymous		0.161***		0.270***	0.298***
		(0.0160)		(0.0270)	(0.0298)
Intellectual Property Measures					
Patent			71.97***		
			(3.249)		
Trademark			10.94***		5.014***
			(0.888)		(0.335)
Patent - Delaware Interaction					
Delaware Only					44.70***
					(3.161)
Patent Only					35.34***
					(1.257)
Patent and Delaware					196.4***
					(10.66)
US CMP Cluster Dummies					
Local				0.705***	0.755***
				(0.0432)	(0.0468)
Traded Resource Intensive				1.292***	1.283***
				(0.0507)	(0.0512)
Traded				1.145***	1.256***
				(0.0380)	(0.0426)
US CMP High-Tech Clusters					
Biotechnology				3.139***	2.288***
				(0.280)	(0.221)
E-Commerce				1.255***	1.136*
				(0.0638)	(0.0591)
IT				2.401***	1.971***
				(0.123)	(0.104)
Medical Devices				1.100	0.886
				(0.0663)	(0.0551)
Semiconductors				3.025***	1.835***
				(0.480)	(0.313)

(Continues on next page)

(Table 3: Continued from prior page)

State Fixed Effects					
Alaska	0.465	0.179	0.220	0.461	0.481
	(0.465)	(0.179)	(0.221)	(0.461)	(0.481)
California	2.854***	2.937***	2.668***	2.652***	2.320***
	(0.217)	(0.222)	(0.204)	(0.203)	(0.179)
Florida	0.642***	0.392***	0.447***	0.685***	0.706***
	(0.0574)	(0.0339)	(0.0390)	(0.0613)	(0.0636)
Georgia	1.229*	0.669***	0.754**	1.282*	1.263*
	(0.125)	(0.0665)	(0.0756)	(0.130)	(0.129)
Idaho	0.965	0.333***	0.394***	0.832	0.741
	(0.245)	(0.0842)	(0.0997)	(0.212)	(0.190)
Massachusetts	2.226***	2.970***	2.520***	1.999***	1.763***
	(0.194)	(0.257)	(0.224)	(0.175)	(0.158)
Michigan	0.503***	0.388***	0.466***	0.483***	0.513***
	(0.0562)	(0.0432)	(0.0522)	(0.0541)	(0.0577)
Missouri	0.917	0.435***	0.531***	0.855	0.850
	(0.124)	(0.0583)	(0.0714)	(0.116)	(0.116)
New York	0.744***	0.793**	0.940	0.741***	0.777**
	(0.0637)	(0.0675)	(0.0808)	(0.0638)	(0.0673)
Oklahoma	1.614***	0.651**	0.828	1.470**	1.461**
	(0.228)	(0.0905)	(0.116)	(0.208)	(0.208)
Oregon	1.608***	0.730*	0.791	1.565**	1.424**
	(0.218)	(0.0976)	(0.106)	(0.213)	(0.195)
Texas	2.525***	1.757***	1.795***	2.404***	2.289***
	(0.204)	(0.140)	(0.145)	(0.194)	(0.187)
Vermont	0.0901***	0.294**	0.292**	0.0950***	0.110***
	(0.0374)	(0.122)	(0.121)	(0.0395)	(0.0458)
Washington	1	1	1	1	1
	(.)	(.)	(.)	(.)	(.)
Wyoming	0.420*	0.975	1.121	0.430*	0.492
	(0.175)	(0.405)	(0.468)	(0.180)	(0.206)
N	12162777	12162777	12162777	12162777	12162777
pseudo R-sq	0.210	0.060	0.130	0.235	0.272

TABLE 4

Regression Model Robustness Tests

We repeat the regression model of Table 3 but include year fixed effects (columns 1 and 2), and year fixed effects with state-specific time-trends (columns 3 and 4), both on top of the state fixed effects already included. Our goal is to evaluate whether changes across time might be driving our results. Given how close our coefficients are in magnitude to those in Table 3, we find little evidence of such. We perform other tests on the performance of our predictive model in our appendix.

	Nowcasting Model	*Full Information Model*	*Nowcasting Model*	*Full Information Model*
	(1)	(2)	(3)	(4)
Corporate Governance Measures				
Corporation	3.293***	2.828***	3.382***	2.915***
	(0.144)	(0.125)	(0.148)	(0.129)
Delaware	41.44***		42.05***	
	(1.314)		(1.336)	
Name-Based Measures				
Short Name	2.942***	2.541***	2.956***	2.551***
	(0.0966)	(0.0858)	(0.0972)	(0.0862)
Eponymous	0.265***	0.291***	0.267***	0.293***
	(0.0265)	(0.0291)	(0.0267)	(0.0293)
Intellectual Property Measures				
Patent				
Trademark		5.200***		5.179***
		(0.357)		(0.358)
Patent - Delaware Interaction				
Delaware Only		35.93***		36.42***
		(1.266)		(1.286)
Patent Only		40.72***		40.10***
		(2.917)		(2.883)
Patent and Delaware		234.7***		241.2***
		(12.84)		(13.25)
US CMP Cluster Dummies				
Local	0.708***	0.758***	0.709***	0.759***
	(0.0433)	(0.0470)	(0.0434)	(0.0470)
Traded Resource Intensive	1.242***	1.236***	1.243***	1.237***
	(0.0491)	(0.0498)	(0.0491)	(0.0499)
Traded	1.116***	1.219***	1.113**	1.216***
	(0.0371)	(0.0413)	(0.0370)	(0.0413)
US CMP High-Tech Clusters				
Biotechnology	3.501***	2.474***	3.529***	2.490***
	(0.317)	(0.248)	(0.320)	(0.251)
E-Commerce	1.191***	1.064	1.184***	1.055
	(0.0608)	(0.0561)	(0.0606)	(0.0559)
IT	2.369***	1.921***	2.358***	1.904***
	(0.120)	(0.101)	(0.119)	(0.100)
Medical Devices	1.104	0.885	1.103	0.883*
	(0.0666)	(0.0554)	(0.0666)	(0.0554)
Semiconductors	3.112***	1.803***	3.134***	1.813***
	(0.494)	(0.308)	(0.497)	(0.310)

State Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	No	No	Yes	Yes
State-Specific Time Trends	No	No	Yes	Yes
N	12162777	12162777	12162777	12162777
Pseudo-R2	0.248	0.287	0.250	0.289

TABLE 5

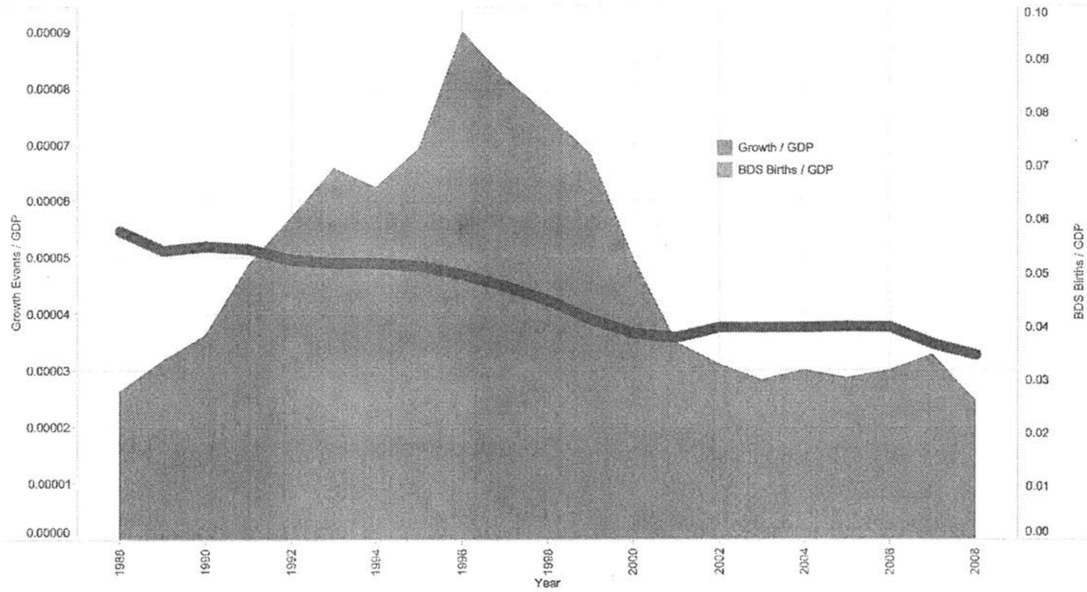
Regression of GDP Growth at MSA Level.
Dependent Variable: $ln(GDP_{2012-2014}) - ln(GDP_{2001-2003})$

	Large Cities[(1)]					*All Cities*
	(1)	(2)	(3)	(4)	(5)	(6)
$Ln(Obs_{2001-2003})$	0.0384*		-0.0109		0.0170	-0.00884
	(0.0154)		(0.0221)		(0.0294)	(0.0130)
$Ln(EQI_{2001-2003})$		0.0803***		0.0641**	0.0684**	0.0494*
		(0.0191)		(0.0218)	(0.0252)	(0.0201)
$Ln(GDP_{2001-2003})$			0.0509**	0.0238	0.0102	0.0145
			(0.0173)	(0.0120)	(0.0285)	(0.0182)
Constant	-0.218	0.809**	-0.288*	0.426	0.446	0.496
	(0.141)	(0.165)	(0.138)	(0.266)	(0.285)	(0.257)
N	63	63	63	63	63	150
R^2	0.093	0.240	0.158	0.278	0.283	0.062

Robust standard errors in parentheses * $p<0.05$ ** $p<0.01$.
(1) Large cities are defined as those with 1000 new firms or more on average per year.

FIGURE 1

Panel A. Firm Births in Business Dynamics Statistics vs. Number of Growth Events per Cohort
Fifteen U.S. states (50.5 percent of 2013 U.S. GDP)



Panel B. Firm Births in Business Dynamics Statistics vs. Yearly Growth in GDP
Fifteen U.S. states (50.5 percent of 2013 U.S. GDP)

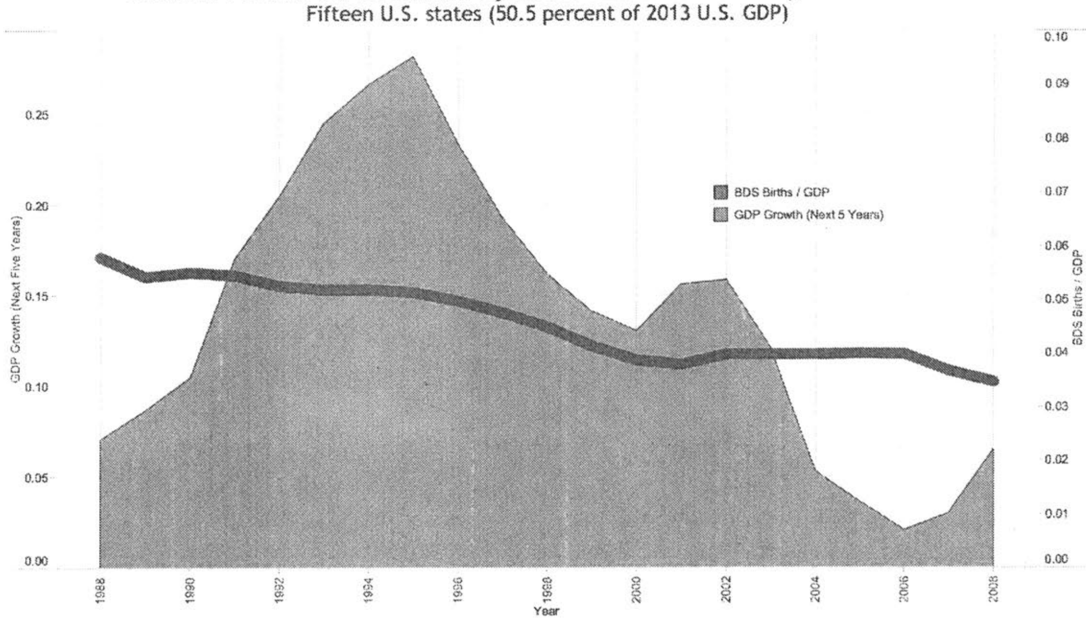


FIGURE 2

10-Fold Test of Predictive Quality of Model*

Top 1% includes 51% of growth outcomes (range: [49%, 53%])
Top 5% includes 69% of growth outcomes (range: [65%, 72%])
Top 10% includes 75% of growth outcomes (range: [70%, 79%])

*10-Fold analysis of model separates the model into 10 random samples and then uses each of those sample as a test sample. We report the average value as well as minimum and maximum (range) of such.

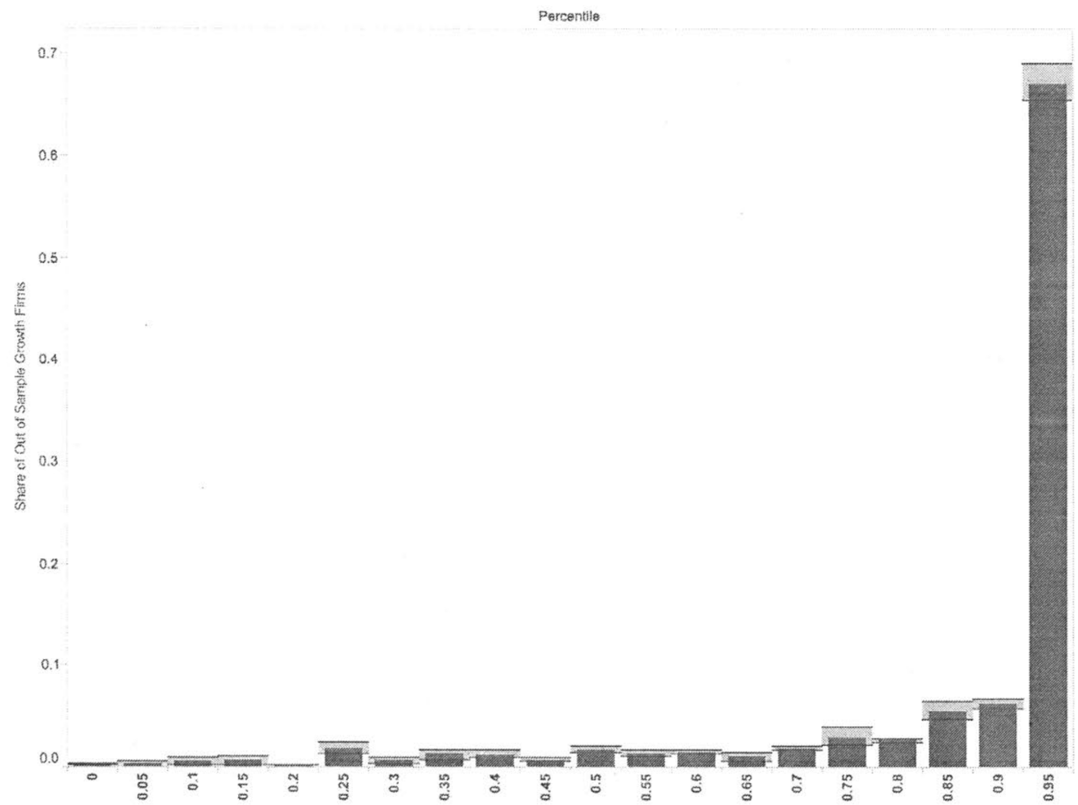


FIGURE 3

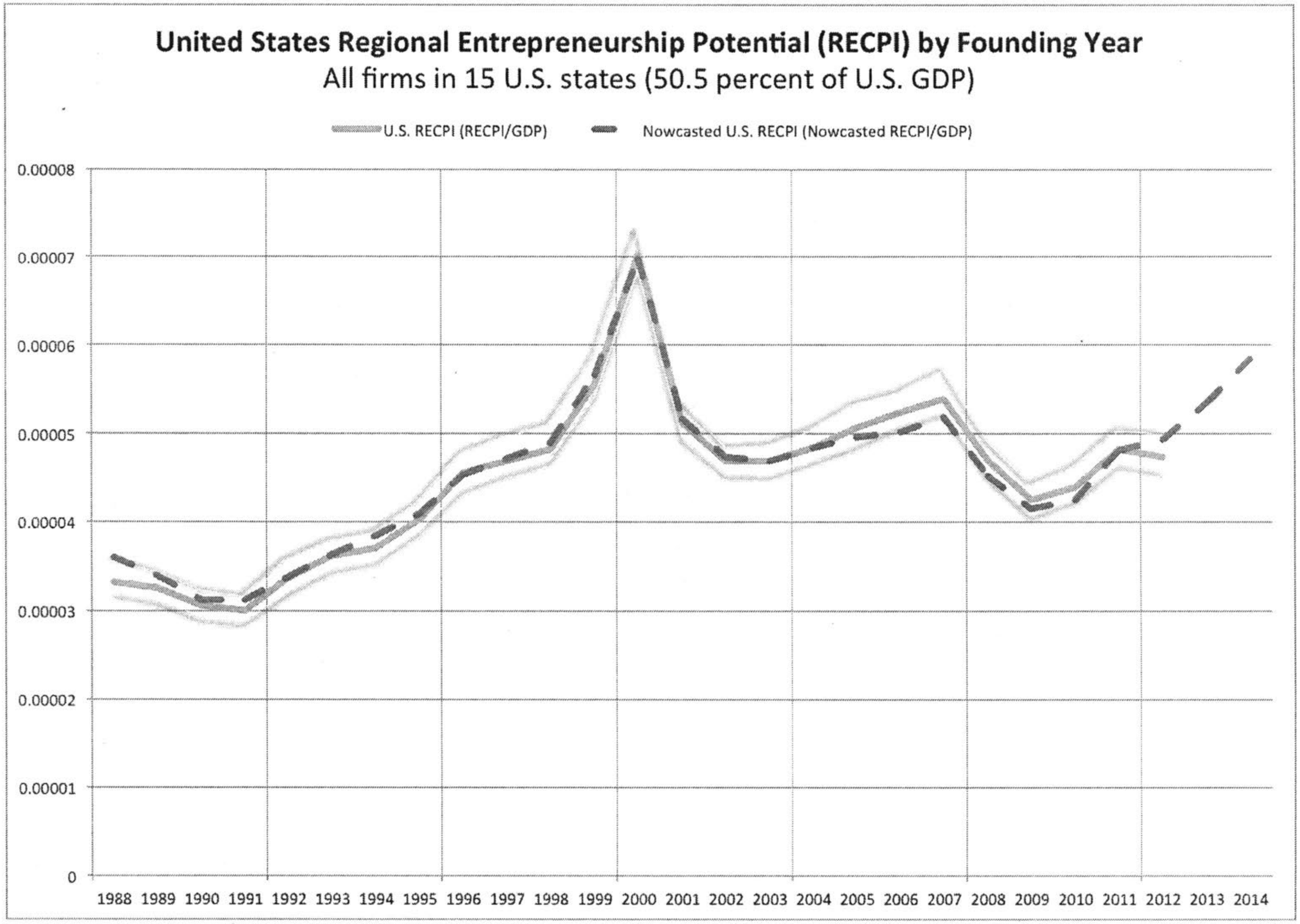

United States Regional Entrepreneurship Potential (RECPI) by Founding Year
All firms in 15 U.S. states (50.5 percent of U.S. GDP)
U.S. RECPI (RECPI/GDP)
Nowcasted U.S. RECPI (Nowcasted RECPI/GDP)
0.00008
0.00007
0.00006
0.00005
0.00004
0.00003
0.00002
0.00001
0
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

FIGURE 4

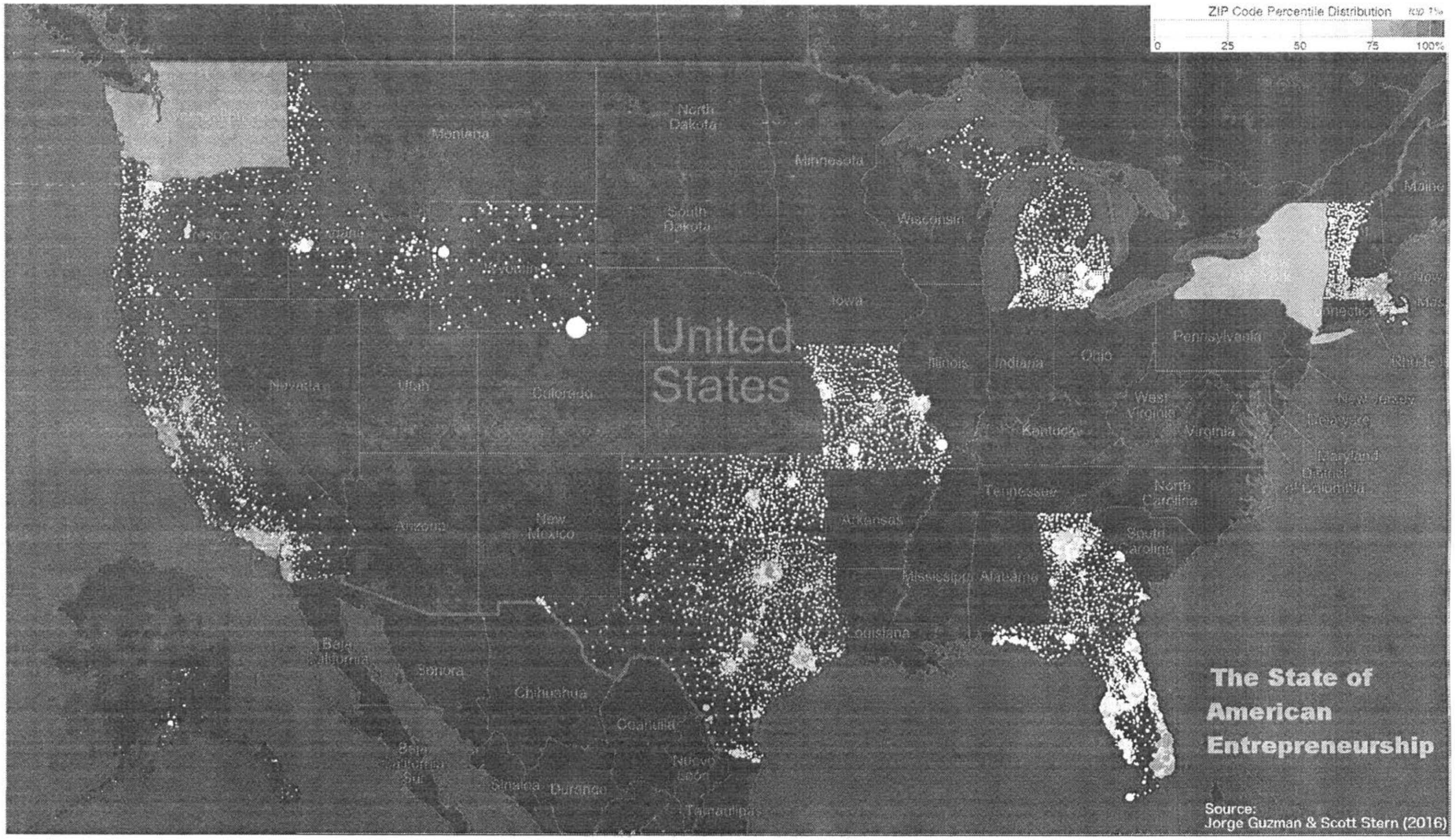

ZIP Code Percentile Distribution
0
25
50
75
100%
United States
Montana
North Dakota
South Dakota
Minnesota
Wisconsin
Iowa
Illinois
Indiana
Ohio
Pennsylvania
Nevada
Utah
Colorado
Arizona
New Mexico
Kentucky
West Virginia
Virginia
Tennessee
North Carolina
South Carolina
Arkansas
Mississippi
Alabama
Louisiana
New Jersey
Delaware
Maryland
District of Columbia
Maine
Sonora
Chihuahua
Coahuila
Sinaloa
Durango
Tamaulipas
The State of American Entrepreneurship
Source:
Jorge Guzman & Scott Stern (2016)

FIGURE 5

Regional Ecosystem Acceleration Index (REAI)

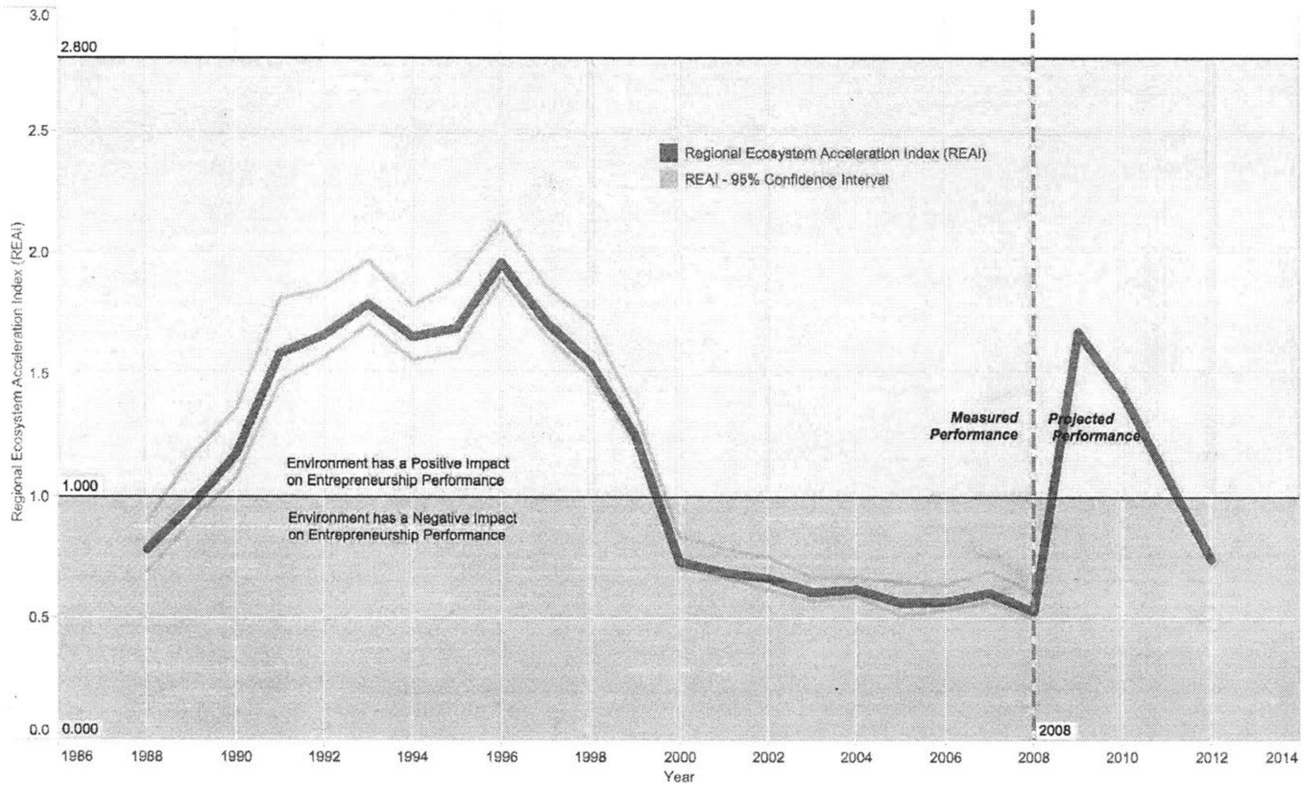
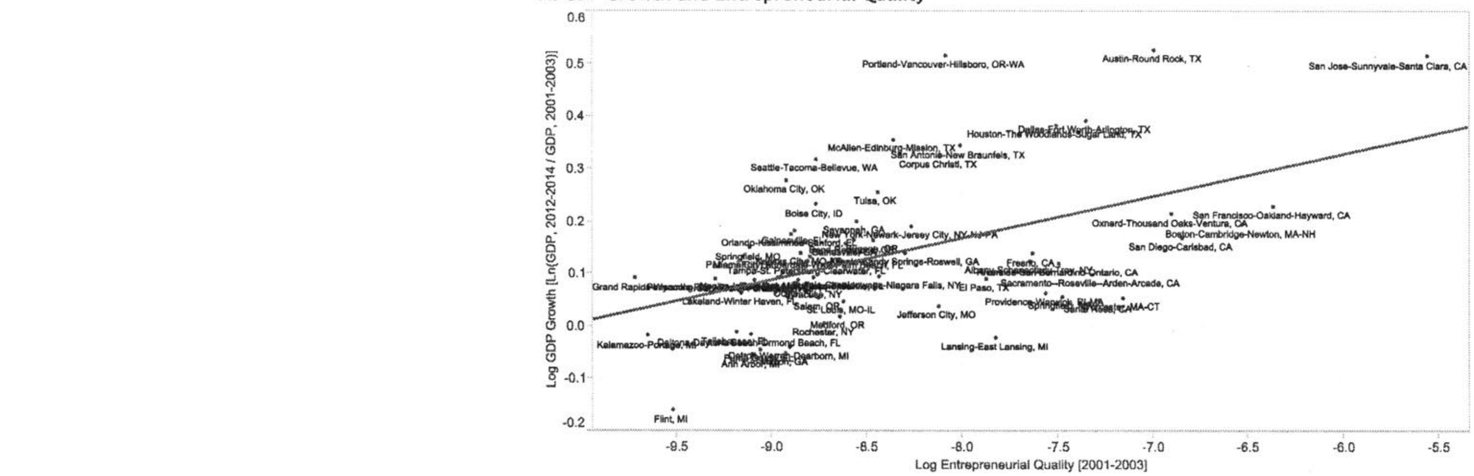


FIGURE 6

Entrepreneurial Quality and Quantity and GDP Growth

A. GDP Growth and Entrepreneurial Quality



B. GDP Growth and Entrepreneurial Quantity

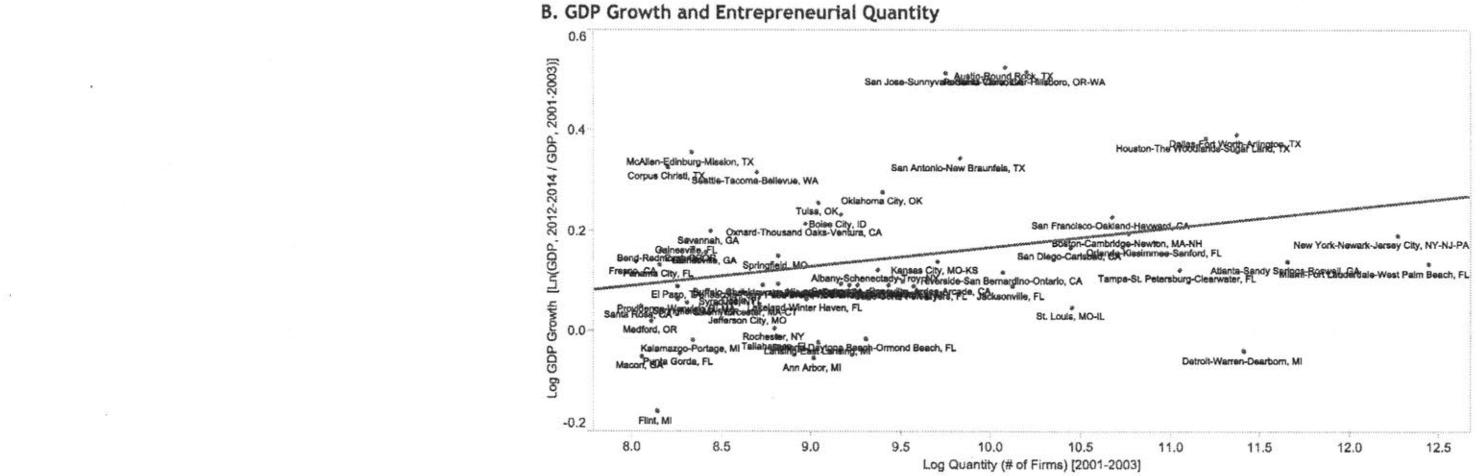
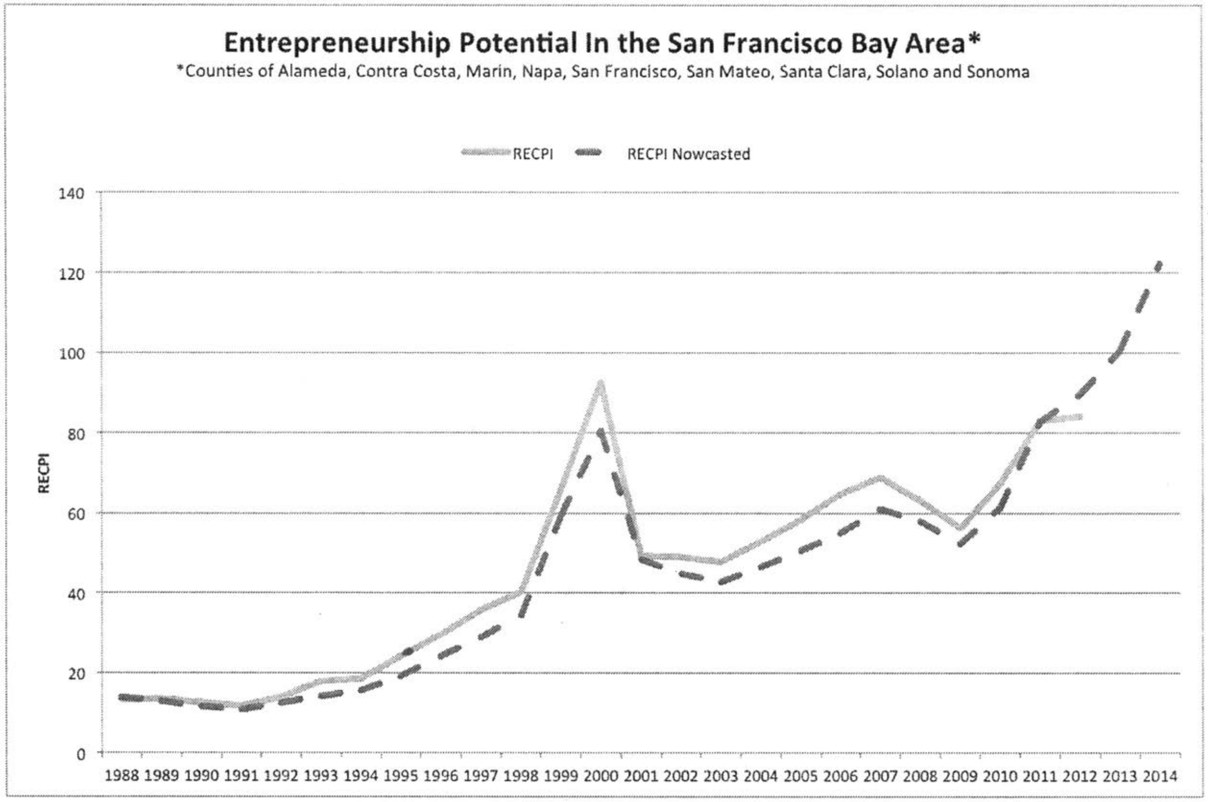


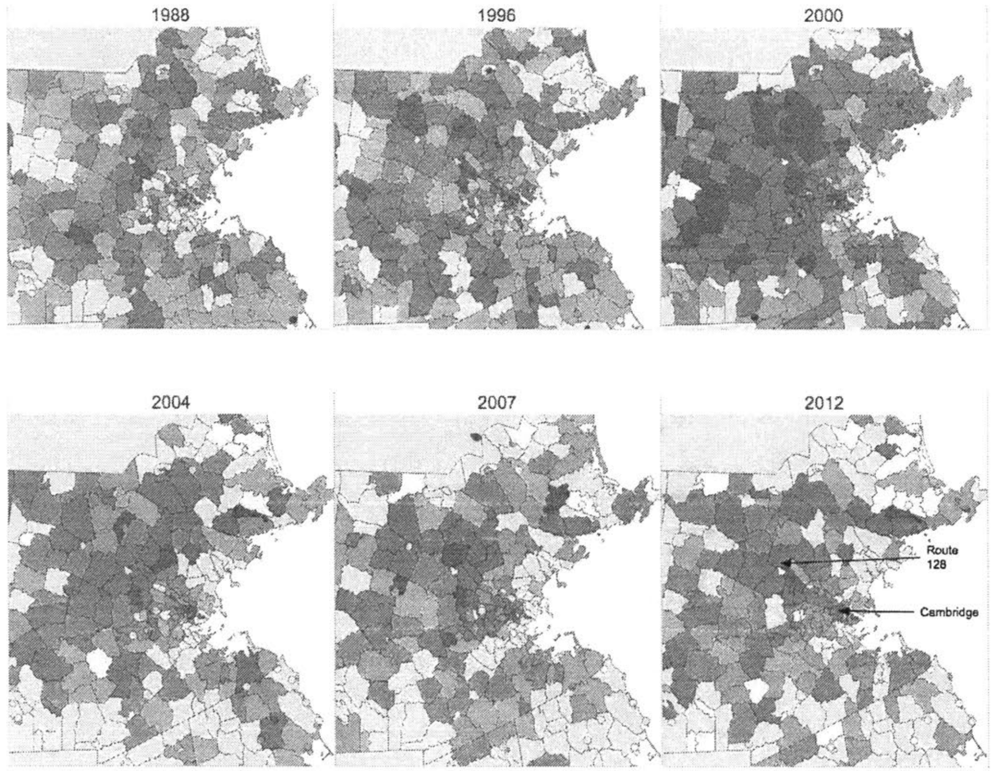
FIGURE 7



Entrepreneurship Potential In the San Francisco Bay Area*
*Counties of Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano and Sonoma
RECPI
RECPI Nowcasted
RECPI
140
120
100
80
60
40
20
0
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

FIGURE 8A

Entreprenurial Quality in Boston. 1988-2012



Entrepreneurial Quality around the Boston Area. Map of entrepreneurial quality for all ZIP codes around the Boston area. In 1988, we find entrepreneurial quality concentrated around the Route 128 corridor, a pattern already documented in the detailed analyses of Massachusetts growth entrepreneurship by Saxenian (1992) and Roberts (1991). As the Boston area moves into the dot-com boom, the amount of entrepreneurial quality increases in both the central and neighboring districts while continuing to be centered around Route 128. However, as the region moves to 2012, the amount of high quality ZIP Codes appears to be about the same as the early 1990s, but the spatial location of such quality has changed. While there is still high quality in Route 128 the central cities of Cambridge and areas of Boston have emerged as hotspots of entrepreneurship.

ZIP Code entrepreneurial quality is the average estimated quality of all firms registered in that year-ZIP Code. Firm quality is estimated using the predictive method outlined in Guzman and Stern (2015a).

Source: Guzman, Jorge and Scott Stern (2016) "The State of American Entrepreneurship"

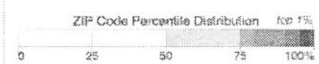
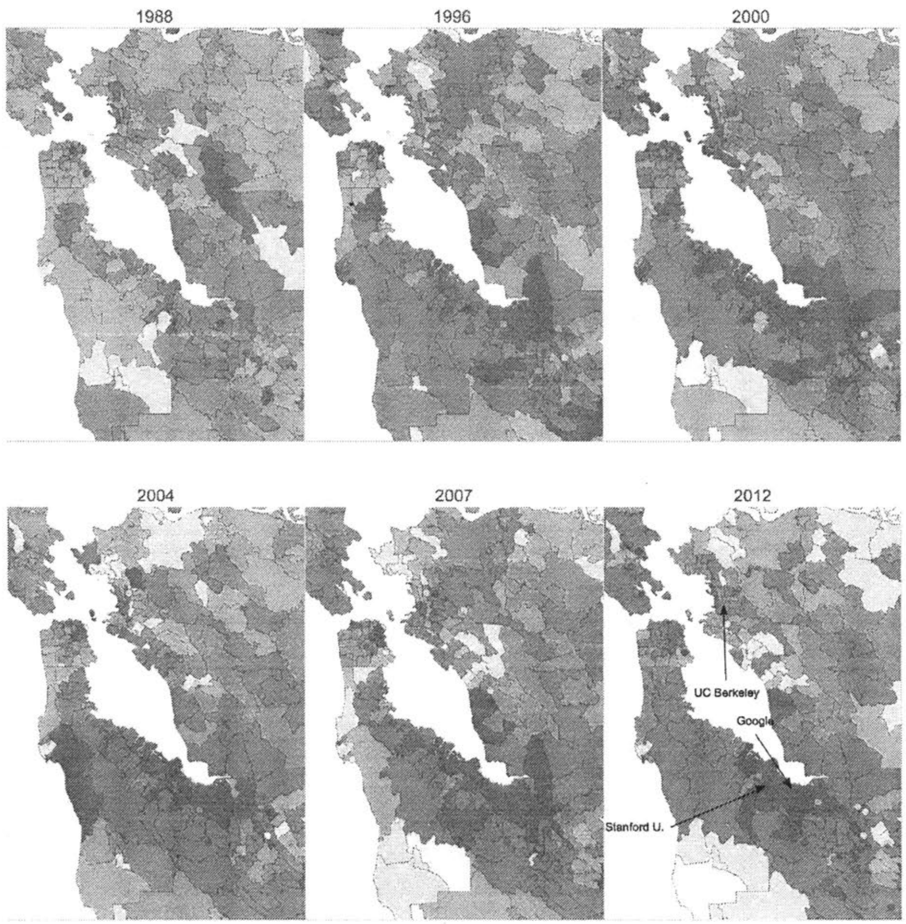


FIGURE 8B

Entreprenurial Quality in Silicon Valley. 1988-2012



Entrepreneurial Quality of Silicon Valley.

ZIP Code entrepreneurial quality is the average estimated quality of all firms registered in that year-ZIP Code. Firm quality is estimated using the predictive method outlined in Guzman and Stern (2016a).

Source: Guzman, Jorge and Scott Stern (2016) "The State of American Entrepreneurship"

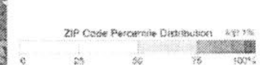
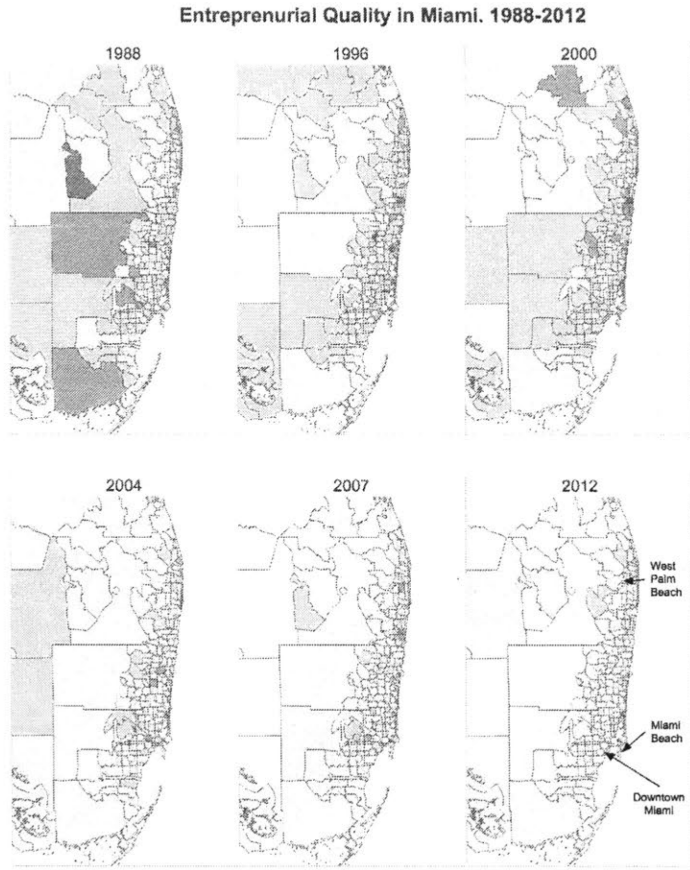


FIGURE 8C

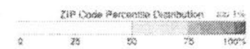


Entrepreneurial Quality of Miami.

ZIP Code entrepreneurial quality is the average estimated quality of all firms registered in that year-ZIP Code. Firm quality is estimated using the predictive method outlined in Guzman and Stern (2015e).

Though the Miami area boasts one of the highest levels of entrepreneurship as measured by self-employment (Glaeser, 2007) entrepreneurial quality in the region is notably low and has continued to drop over the last 20 years. The stark differences between these two measures highlights the importance of using

Source: Guzman, Jorge and Scott Stern (2016) "The State of American Entrepreneurship"



CHAPTER 3

GENDER GAP IN ENTREPRENEURSHIP

(co-authored with Aleksandra (Olenka) Kacperczyk)

ABSTRACT

Using data on all businesses registered in the state of California and Massachusetts between 1995 and 2011, we examine gender inequality in entrepreneurship. We propose and empirically show that female entrepreneurs are more likely than male entrepreneurs to pursue lower-quality ventures. We also find that female entrepreneurs are less likely than male entrepreneurs to obtain VC funding, even for comparable levels of entrepreneurial quality. Our findings further indicate that the gender disadvantage in entrepreneurship decreases dramatically with increases with entrepreneurial quality. Consistent with the notion of statistical discrimination, we also show that gender disadvantage is amplified for amateur evaluators, more likely to rely on stereotypes when assessing new-venture potential. More generally, the study emphasizes the need to take the distribution of entrepreneurial quality into consideration when examining the mechanisms behind gender gap in entrepreneurship.

I. INTRODUCTION

Transition to entrepreneurship is one of the most important features of the today's economy. For individuals, the decision to launch a new venture may be an important means of career attainment (Sorensen and Sharkey, 2014) and value capture (Carnahan, Agarwal, and Campbell, 2011). Yet, the entrepreneurial economy appears to be particularly disadvantageous for women. Though women have made considerable inroads into domains traditionally dominated by men (Rosin, 2012), they are significantly less likely than men to succeed as entrepreneurs (U.S. Department of Commerce, Economics and Statistics Administration, 2010; Canning, Haque, and Wang, 2012), and this imbalance is starkest among high-growth ventures, where female founders usually represent a small proportion of entrepreneurs (Robb, Coleman, and Stangler 2014).

In explaining this gender-based gap in entrepreneurship, an increased attention has focused on the role of venture capitalists (VC) in systematically excluding female-run ventures from the benefits of the venture-capital economy. Even though VC funding has become the engine behind commercializing innovations over the past several decades (Kortum and Lerner, 2000; Nanda and Rhodes-Kropf, 2013), the participation of women remains disproportionately low (Canning, Haque, and Wang, 2012; Greene at al., 2003; Gatewood, 2003). Accordingly, the extant accounts have concluded that a powerful solution to achieve more gender-equitable outcomes hinges on mitigating VC's biases against female entrepreneurs.[1]

Yet, despite the well-established finding that female-run ventures are less likely to secure venture-capital funding than male-run ventures, there remains substantial ambiguity concerning the precise drivers of this gap. Broadly speaking, two theoretical lenses might shed light on this empirical pattern. On the one hand, sociological and psychological theories of homophily, in-group preference and discrimination imply that evaluators, such as venture capitalists, tend to hold biases and preferences that systematically disadvantage female-led ventures (Brooks et al. 2014; Thébaud 2015). From this perspective, the primary source of gender gap in high-growth entrepreneurship lies in the biased preferences of these resource holders. On the other hand, resource-based explanations rooted in economics and strategy (Barney, 1991; Lippman and Rumelt, 2003) imply that the probability of securing entrepreneurial investment increases when resources are valuable, rare, and difficult-to-imitate. From this perspective, the key source of gender gap in high-growth entrepreneurship lies in resource disparities that female and male-led ventures might exhibit.

Although designing solutions to alleviate gender differences in entrepreneurship hinges on our ability to disentangle these two mechanisms, this poses a substantial empirical challenge. A major difficulty pertains to adequately accounting for the heterogeneity in new-venture resources, which determine the entrepreneurial quality and growth potential of a new venture (Guzman and Stern, 2015a; 2015b; 2016). Collecting appropriate data to measure such differences is difficult for several reasons. First, many data sources fail to measure new-venture quality from the beginning of the process – that is, when a new firm is founded. Yet, when data on new ventures are collected in the later stages of the development (such as during the receipt of venture capital), sample selection bias might arise because lower-quality ventures are likely to fail long before they are recorded. Moreover, extant accounts have often relied on surveys to

[1] For example, a common implication has been that, in order to succeed to accessing early-stage capital, women need to mitigating cognitive biases that venture capitalists might hold by "pitching more like men" (Brush et al., 2014).

measure venture's quality or initial resources (Reynolds, 2000), but these data may suffer from important biases, as entrepreneurs are asked to reconstruct events, attitudes, and motivations that took place in the past (Kepler and Shane, 2007). Finally, though a number of past studies have taken industry differences into consideration (Carter and Shaw 2006; Coleman and Robb 2009), new-venture quality tends to vary even within an industry (Guzman and Stern, 2015a; 2015b; 2016), suggesting the need for more granular performance measures at the firm level. More generally, measuring new-venture quality with precision requires developing objective and timely quality indicators that leverage a population of new firms.

In this study, we analyze and compare gender-based differences in entrepreneurial outcomes for the population of entrepreneurs for whom differences in entrepreneurial quality can be readily observed and measured directly. We leverage a novel approach that estimates entrepreneurial quality for each start-up, using publicly available business registration records. Following Guzman and Stern (2015a), we take advantage of business registration records as well as other publicly available data sources to develop a measure of entrepreneurial quality for new firms, which relies on differences in key start-up characteristics, shown to predict key growth outcomes, such as IPOs and high-value acquisitions (Guzman and Stern, 2015a; 2015b). In doing so, we construct a novel dataset containing *all* California and Massachusetts for-profit start-up corporations, limited liability companies, and partnerships from 1995 to 2011.

In the resulting, comprehensive empirical analysis, we find that discrimination processes play a more limited role than has been thought previously in generating gender-based gap in entrepreneurship. First, we confirm prior findings that a female-led venture is associated with a 63% lower likelihood of raising venture capital. However, we also show that as much as 65% of this gap can be explained by differences in entrepreneurial quality at the time of the venture founding. Specifically, our results indicate that female-run ventures are much less likely to exhibit the kinds of characteristics that predict growth outcomes. Once we account for differences in entrepreneurial quality, the remaining gap depends on the position of women in a given quality distribution. Most dramatically, our findings show that the gap disappears for ventures in higher "positions" in the estimated quality distribution, suggesting that women-led ventures are equally likely as male-led ventures to secure VC funding, at higher levels of entrepreneurial quality. Overall, our findings indicate that stark differences in new-venture quality play a more crucial role than previously assumed in driving the observed gender differences in entrepreneurship.

Our results extend the literature on the existence of gender-based entrepreneurship gap, with an application to growth-based entrepreneurship. We show that female entrepreneurs are less likely to obtain VC funding than their male counterparts, but that the majority of this gap reflects differences in new-venture-quality at founding, as well as the specific position of a start-up within an estimated quality distribution. More generally, our findings imply that equalizing access to resources at the earliest stage of a venture's founding is most critical in reducing gender inequality in high-growth entrepreneurship.

II. THEORY AND HYPOTHESES

Gender inequality is a persistent feature in entrepreneurship outcomes, with women being less likely to become entrepreneurs than men (Aldrich, 2005; Ruef, Aldrich, and Carter, 2003) and less likely to outperform once a new venture is founded (Kim, Aldrich, and Keister, 2006; Yang and Aldrich, 2014). In examining the drivers of such gap, extant accounts have argued that venture capitalists play a key role by systematically excluding women from access to venture-capital funding (Canning, Haque, and Wang, 2012; Brush et al., 2014; 2015), on which high-growth entrepreneurship depends (Kortum and Lerner, 2000; Nanda and Rhodes-Kropf, 2013). But despite the mounting evidence, the precise mechanisms that generate it are far less well understood. The available empirical evidence is consistent with two distinct mechanisms, derived directly from two distinct theoretical traditions, as discussed below.

Gender Discrimination in Access to VC Funding

First, there is a rationale to expect that gender gap in high-growth entrepreneurship arises because evaluators such as VCs hold systematic preferences for male-led ventures. Such preferences may stem from homophily, similarity-attraction, in-group preference, and gender stereotypes. Specifically, theories of homophily, similarity-attraction, and in-group preference posit that group members tend to informally associate with colleagues who share salient demographic characteristics (e.g. McPhereson, Smith-Lovin, and Cook, 2001). Therefore, when male venture capitalists evaluate male-led ventures, demographic similarity may lead to a higher degree of intergroup (e.g. VC/entrepreneur) interaction. The latter might further lead to increased liking and attraction (e.g., Tsui and O'Reilly, 1989), therefore increasing the probability of an entrepreneurial investment. In line with this view, there is evidence that venture capitalists are predominantly male (Brush et al., 2015), suggesting that, simply on the basis of homophily and in-group biases (e.g. Tajfel and Turner, 1979), male venture capitalists may show preferences for demographically similar start-ups – i.e., those led by men. Consequently, it may be harder for females to gain attention for their "deal."

Another reason why VCs may drive gender gap in high-growth entrepreneurship are negative biases about gender. Theories of discrimination suggest that gender inequalities arise because female endure disparate treatment from key resource holders due to discrimination and negative stereotypes about gender (Ridgeway and Correll, 2004; Castilla, 2008). These mechanisms are particularly likely to apply to the entrepreneurial setting because job-related schemas and stereotypes associated with entrepreneurship likely trigger systematic biases against individuals who are not males. There is evidence that entrepreneurship is a male-typed activity (Yang and Aldrich, 2014; Cavalluzzo, Cavalluzzo and Wolken 2002; Shane et al. 2012) and that consequently resource holders tend to discount the competence of female entrepreneurs and the investment-worthiness of their enterprises. Moreover, since female-led ventures are relatively rare, women may appear as less usual or natural fit for these positions. Regardless of the actual performance of female-led start-ups, the stereotypes associated with entrepreneurship might trigger bias against female-led ventures, leading to lower rates of venture-capital deals for start-ups founded or run by women. Overall, these accounts locate the source of inequality in the characteristics, preferences, and biases of the evaluators (i.e., venture capitalists), and not in resources of new ventures.

But though anecdotal evidence abounds with these kinds of arguments, the conclusion that venture capitalists prevent female success in high-growth entrepreneurship by discriminating against female-led ventures rests on fragile foundations. First, gender discrimination is often difficult for researchers to observe (Petersen and Saporta, 2004) and therefore providing direct evidence for such is challenging. Moreover, not all empirical studies confirm that resource holders perceive female entrepreneurs as less competent. For example, numerous studies on access to debt capital have failed to confirm such funding disadvantage for women-led firms (e.g., Blanchflower, Levine and Zimmerman 2003; Carter et al. 2007; Cavalluzo et al., 2002). And many other studies have not found that investors or lenders use different criteria to evaluate female-led ventures (e.g., Haines, Orser, and Riding 1999; Orser et al., 2006). Altogether, these studies suggest that the existing accounts remain vulnerable to the possibility that the well-established gender gap in high-growth entrepreneurship might not necessarily reflect gender biases held by investors but rather be driven by an alternative process.

Entrepreneurial Quality and Access to VC Funding

An alternative theoretical perspective suggests that, rather than reflecting VC biases, gender gap in entrepreneurship might arise due to unequal distribution of resources at the founding stage. The resource-based view of the firm implies that the presence of valuable, rare, and difficult-to-imitate resources is the key driver of firm performance and that such resources are thus an indicator of firm quality (Barney, 1991; Lippman and Rumelt, 2003). Heterogeneity in resources is particularly pertinent in the context of entrepreneurship because wide distribution of entrepreneurial resources has been shown to determine start-up growth potential (e.g., Alvarez and Lowell Busenitz, 2001; Mosakowski, 1997), driving entrepreneurial milestones, such as access to VC funding and IPO (Guzman and Stern, 2015b). For example, studies of venture-capital funding have found that objective markers of new venture quality drive success rates in access to VC funding and that venture capitalists attend to these objective quality indicators (Kaplan and Lerner 2010; Puri and Zerutskie, 2012). Related research in entrepreneurship has suggested that there exists considerable heterogeneity in entrepreneurial quality (Hurst and Pugsley, 2010) that persists even within a single industry (Guzman and Stern, 2015a; 2015b; 2016) and should thus be accounted for by examining the underlying differences in the firms themselves (Hurst and Pugsley, 2011; Kaplan and Lerner, 2010; Schoar, 2010). Still others have argued that it is the high-quality start-ups that drive economic growth such that growth is concentrated in a relatively small share of successful start-ups (Kaplan and Lerner, 2010; Kerr, Nanda, and Rhodes-Kropf, 2015). In numerous studies, Guzman and Stern (2015a; 2015b; 2016), find that over seventy percent of realized growth outcomes occur in the top five percent of our estimated quality distribution (and nearly fifty percent in the top one percent of the estimated quality distribution). These studies imply that initial quality differences between female- and male-led ventures, rather than biased preferences of venture capitalists at the investment stage, might drive gender gap in entrepreneurship.

There is a rationale to expect that new-ventures' resources might be unequally distributed across female and male-led start-ups. First, we expect that women will found and run less growth-oriented businesses because structural gender inequalities in occupation status are likely to carry over to entrepreneurship and hinder women's willingness and ability to pursue high-quality start-ups. It has been well established that women tend to be concentrated in low-profitability industries and positions that are low-status (Fernandez and Sosa, 2005). The sorting of women into less attractive industries and occupations affects the kinds of ventures women might found because women's occupancy of lower-status positions limits their ability to identify lucrative opportunities for new ventures. For example, there is evidence that women tend to start new ventures in unattractive industries with limited opportunities in part because they tend to occupy jobs in those industries (Loscocco et al., 1991; Kalleberg and Leicht, 1991).[2] A related line of research suggests that women may pursue ventures of lower-growth potential due to differences in human capital and educational background. Consistent with this claim, studies in economics and sociology have shown that levels of skills and education rather than gender account for the well-known differences in wages across women and men (Tomaskovic-Devey 1993, Kilbourne et al. 1994, Petersen and Morgan 1995). Other work has similarly found that women are underrepresented in engineering or science and STEM disciplines and that they are less likely to patent (Brush et al., 2006; Ding et al. 2006). Collectively, these studies imply that women will be less likely to access resources and skills necessary to start businesses that demand technical skills. More generally, if women occupy less attractive positions in less attractive industries, they will be less likely to start and lead high-quality, resource-attractive ventures.

Second, women and men have been shown to hold different motivations and preferences for starting new ventures (Brush et al., 2006; Carter et al., 1997; Sexton and Bowman-Upton, 1990). The notion that differences in preferences and attitudes partly shape women's assessment of the kinds of careers suitable has been well established (Correll, 2004). Women's preferences are influenced by disproportionate work-life demands that women face due to childrearing and household chores, which tend to fall to a greater extent on women and to generate an acute conflict for female workers (e.g., Brett and Stroh 2003; Rothbard, 2001). Such demands shape the perceived attractiveness of the occupational paths available to women (Aldrich and Cliff, 2003; Barbulescu and Bidwell, 2013), encouraging them to choose self-employment (Carter et al., 2003; Birley 1989) or part-time jobs to allow for accommodating family needs (Ginther and Kahn 2006). A number of studies have shown empirically that these kinds of preferences push women into entrepreneurship in hope of developing more flexible work schedule, balancing

[2] Studies have shown that women found ventures in consumer-oriented and personal services, retail and trade (Anna et al., 2000; Brush et al., 2006).

work and family demands (Georgellis and Wall, 2004; Lombard 2001) or reducing the cost of childcare (Connelly 1992; Presser and Baldwin, 1980; Thebaud, 2015). Others have similarly found that the probability of self-employment increases when women become a parent (Boden, 1996) and the needs for flexibility and work-family balance increase (Boden, 1999; Carter et al., 2003; Connelly 1992). Finally, since women face more competing demands for their time (e.g., Brett and Stroh 2003), we expect them to devote less time and effort to the process of forming and developing a new venture. These differences in preferences imply that women might choose to found and lead the kinds of ventures that exhibit less attractive resources and hence have lower entrepreneurial quality.

Finally, women may be involved in ventures of lower-growth potential because of stark differences in network structure and composition (Aldrich, 1989; Cromie and Birley, 1992). Although social ties are the most important resource for entrepreneurs (Stewart, 1990), women have generally less valuable networks (Moore, 1990; Smith, 2000) and this tendency has been found to carry over into entrepreneurship (Aldrich, Reese, and Dubini, 1989; Renzulli, 1998; Ruef, Aldrich, and Carter, 2003). For example, women overinvest in strong ties (Fischer and Oliker, 1983), or develop relatively small and homogenous networks (e.g., Ruef, Aldrich, and Carter, 2003; Renzulli, 1998), all of which tend to limit access to diverse information and instrumental support central to entrepreneurial entry (Aldrich, 1989). Such profound differences in network structure might limit women's access to opportunities and resources, reducing the probability that they are engaged in founding and managing high-quality ventures.

Overall, growing empirical evidence suggests that women tend to be excluded from high-growth entrepreneurship and that such imbalance could possibly arise due to unequal access to valuable, rare, and difficult-to-imitate resources on which entrepreneurial success depends. However, the available empirical evidence is consistent with at least two distinct mechanisms, each of which locate the source of gender gap at different stages of the entrepreneurial process. Progress in assessing the merits of the different mechanisms therefore requires identifying a setting in which the two processes can be disentangled. In what follows below, we discuss the analytic strategy to separate the two effects.

III. METHODOLOGY

Empirical Strategy

To decompose gender gap along different stages of the entrepreneurial process, we track new ventures since the time of their founding using detailed data on start-ups prior to the investment stage. However, measuring entrepreneurial quality at the time of founding across firms is challenging (Hathaway and Litan, 2014; Guzman and Stern, 2015a), because (…) "it is very difficult, if not impossible, to know at the founding whether or not firms are likely to survive and/or grow" (Hathaway and Litan, 2014). Appropriate data are difficult to assemble and rarely available to researchers; consequently, only few studies have attempted to account for entrepreneurial quality, in general, and considered gender differences in quality, in particular.

In this paper, we build on a novel empirical approach developed by Guzman and Stern (2015a; 2015b) to address these challenges, and to systematically track variation in quality of start-ups over time. Rather than measuring firm quality directly (or attempting to define it precisely), we build on the central insight of Guzman and Stern (2015a, 2015b) that a relationship between firm observables at the time of founding and ex-post measures of performance can be used to assess entrepreneurial quality.[3]

The methodology we use estimates entrepreneurial quality for each start-up at the firm level at the time of the firm founding. To identify firm founding, we use publicly available business registration records as well as other publicly available data sources to observe a set of "start-up characteristics." We rank the population of registered ventures based on their characteristics and subsequently construct a novel estimate of entrepreneurial quality for new firms, by creating a mapping between a growth outcome (observed of course with a lag) and the characteristics observable at or near the time of founding. More precisely, for a firm i born in region r at time t, with start-up characteristics $X_{i,r,t}$, we observe a growth outcome $g_{i,r,t+s}$, s years after founding and estimate:

[3] This approach can be illustrated with the following example. Assume entrepreneurs and start-up leaders are presented with choices regarding the firm (e.g. registering intellectual property, corporate form, jurisdiction, etc.), whose value varies depending on the firm's goals. For example, registering as a corporation is necessary to sell stock to investors but an LLC – which avoids double taxation – is preferable when ones seeks to hold the company herself. These strategic choices reflect intentions of growth and the market potential of a product or idea and can be usefully examined as indicators of firm type. In adopting this approach, we can sidestep multiple exogenous issues that have vexed the literature, such as education, social upbringing, family composition, and skills. While these covariates might be difficult to systematically observe on a large scale, they will be reflected in early choices the start-up makes, in as much as entrepreneurs and start-up leaders internalize these choices in their decision to start a firm.

$$\theta_{i,r,t} = 1000 \times P(g_{i,r,t+s}|X_{i,r,t}) = 1000 \times f(\alpha + \beta X_{i,r,t})$$

Using this predictive model, we are able to *predict* quality as the probability of achieving a growth outcome given start-up characteristics at birth, and estimate entrepreneurial quality as $\hat{\theta}_{i,r,t}$. As long as the process by which start-up characteristics map to growth remain stable over time, an assumption validated in Guzman and Stern (2016), we are able to then develop an estimate for entrepreneurial quality even for very recent cohorts.[4]

Data Sources and Sample Selection

Business registration records are public records created when individuals register a business. Since business registration is a requirement for growth, the sample of business registrants in a given time period composes a cohort of start-ups for which one could evaluate quantity (the number of business registrants, or the number of business registrants of a certain type) as well as quality (by assessing the underlying quality of each business registrant in a standardized way).

Our sample consists of all for-profit start-up business registrants in the states of California and Massachusetts in the period between 1995 and 2011. These states are particularly suitable for our purpose because more than 50% of VC market is located in California and 10% of the VC market is located in Massachusetts (by dollars invested in 2014 (NVCA, 2015)). During the period covered by our sample, it was possible to register several types of businesses: corporations, limited liability companies, limited liability partnerships, and general partnerships. All corporations, partnerships, and limited liability firms must register with the state in order to take advantage of important benefits: hence, the act of registering the firm triggers the legal creation of the company. As such, these records form the *population* of California and Massachusetts businesses that take a form that is a practical prerequisite for growth. Accordingly, our analysis draws on the complete population of firms satisfying one of the following conditions: (a) a for-profit firm whose jurisdiction is in California or in Massachusetts; (b) a for-profit firm whose jurisdiction is in Delaware but whose principal office address is in

[4] Our approach to measure quality offers two other benefits over alternative approaches. First, by using a measure of entrepreneurial quality based on characteristics at birth rather than later firm quality, we are able to separate the intrinsic quality of a new firm without confounding it with the effect of its ecosystem on subsequent growth. Second, by including all firms that achieve the basic and trivial requirement of registration, rather than some select set of firms that have achieved an intermediary outcome (e.g. being admitted to an accelerator), we reduce any potential bias on the selection process of the list itself (e.g. bias in the screening processes of accelerators).

California or in Massachusetts. In doing so, we exclude non-profit organizations as well as companies whose primary location is external to California or Massachusetts. Finally, we merge this database with VentureXpert data, which contains detailed information on venture capital funding. All venture investments in VentureXpert are matched by exact name with start-ups registered in California and Massachusetts. These selection criteria yield a sample of 1,875,087 start-up.

Dependent Variable

Our primary dependent variable is the reception of VC funding. We focus on access to VC funding as the main outcome for several reasons. First, venture capital has been a central source of finance for commercializing innovations in the US economy over the past several decades (Kortum and Lerner, 2000; Samila and Sorenson, 2011). Moreover, though venture-backed startups represent only a very small fraction of all new firms (about 1/6th of 1%), over 60% of IPOs since 1999 have been venture-backed (Kaplan and Lerner, 2010). The main dependent variable is a dummy equal to 1 if a start-up receives venture capital funding within 2 years since the founding date. We consider a two-year window to control any potential time heterogeneity but the results are also robust to different time frames. Seventy percent of VC events occur within 2 years[5]. For robustness, we consider the total amount of capital raised, conditional of VC investment, and find similar results (available upon request).

Independent Variables

Female-led start-up. The main independent variable is a dummy equal to 1 if a start-up is female-run. The identification of female-led start-ups relies on two conditions: (a) gender could be identified for at least one of the main managers of the firm (i.e., the President or the CEO of the corporation); (b) if gender is identified for the management team, all members for which we can identify gender, need to be female. We construct a measure of gender based on first names provided by the business registration records for individuals in the above-mentioned positions. To do so, we use the Social Security Administration list of names registered at least five times in a year from 1950 to 2000. To handle ambiguous names (e.g. Taylor) we use only names that are

[5] This result is consistent with other samples of business registration records containing more states (Catalini, Guzman, and Stern, 2016) and samples matching the receipt of VC to the U.S. Census Longitudinal Business Database (LBD) (Puri and Zerutskie, 2012).

give times more common in female than male (or vice-versa). With this procedure, we are able to confidently identify gender for 84 percent of firms in our sample.

Entrepreneurial Quality Indicators. We utilize quality indicators established by Guzman and Stern (2015a; 2015b) as markers of entrepreneurial growth potential. These measures incorporate start-up characteristics based on business registration observables and start-up characteristics based on external observables. We construct a number of measures based on information available in the business registration records. We first construct two binary measures that relate to how the firm is registered, *Corporation*, whether the firm is a corporation rather than an LLC or partnership, and *Delaware Jurisdiction*, whether the firm is registered in Delaware. *Corporation* is an indicator equal to 1 if the firm is registered as a corporation and 0 if it is registered either as an LLC or partnership. *Delaware jurisdiction* is equal to 1 if the firm is registered under Delaware, but has its main office in California (all other foreign firms are dropped before analysis). We then construct two additional measures based directly on the name of the firm. *Eponymy* is equal to 1 if the first, middle, or last name of the top managers is part of the name of the firm itself. Our second measure relates to the length of the firm name. Based on our review of naming patterns of growth-oriented start-ups versus the full business registration database, a striking feature of growth-oriented firms is that the vast majority of their names consist of two words. We define *Short Name* to be equal to one if the entire firm name has three or fewer words, and zero otherwise. Based on findings of Guzman and Stern (2015a), we additionally examine the type of traded cluster a firm is associated with, focusing on whether the firm is in a high-technology cluster or a cluster associated with resource-intensive industries.[6] Finally, an important quality predictor evidenced is presence of patents or trademarks. These measures are constructed using a name-matching algorithm that connects the firms in the business registration data to external data sources. We include patents that are filed by the firm within the first year of registration and patents that are assigned to the firm within the first year from another entity (e.g., an inventor or another firm). Our second measure, *Trademark*, is equal to 1 if a firm applies for a trademark within the first year of registration.

Estimating Entrepreneurial Quality

[6] For our high technology cluster group (*Traded High Technology*), we draw on firm names from industries in ten USCMP clusters: Aerospace Vehicles, Analytical Instruments, Biopharmaceuticals, Downstream Chemical, Information Technology, Medical Devices, Metalworking Technology, Plastics, Production Technology and Heavy Machinery, and Upstream Chemical.

Following the methodology in Guzman and Stern (2015a), we estimate the firm-level probability of achieving a growth outcome based on start-up characteristics. First, we estimate the model presented in our methodology section with a growth outcome equal to 1 if a firm achieves an IPO or acquisition within 6 years, and include all early stage observables as described above. While multiple definitions of growth are possible, we use this outcome to correctly characterize the venture capital process: entrepreneurial quality indicates the potential to achieve a successful exit. Accordingly, we construct a dummy variable equal to 1 if the startup achieves an initial public offering (IPO) or is acquired within 6 years of registration. Both outcomes are drawn from Thomson Reuters SDC Platinum. We observe 1099 positive growth outcomes for the 1995-2005 startup cohorts (used in all our regressions). The median acquisition price is $77 million (ranging from a minimum of $11.9 million at the 5th percentile to $1.92 billion at the 95th percentile). Finally, we use this model to predict the probability of a growth outcome for a firm given its start-up characteristics. This probability is a measure of entrepreneurial quality at the time of firm founding.

Our initial model excludes gender to allow for the estimated quality distribution to be independent of whether a start-up is female- or male-run. For any given firm, our measure estimates its likelihood of achieving a growth outcome, given early stage observables without considering the effect of gender on growth. This allows us to further estimate the effect of gender while controlling for entrepreneurial quality driven by other characteristics.

Before we turn to analyses of gender, we first discuss our computation of the entrepreneurial quality metric, as can be seen in Table A1. We begin by estimating a logit regression specification with all quality observables estimated for all firms registered in CA and MA between 1995 and 2005. We use the observables shown to have good fit as well as a strong predictive power in out of sample tests (Guzman and Stern, 2015a; 2015b). We find corporations are 5.8 times more likely to grow relative to the baseline, firms with a short name 2.5 times more likely to grow, and eponymous firms 70 percent *less* likely to grow. Firms with a trademark are almost 4 times more likely to grow, firms with a patent 35 times more likely, and firms registered in Delaware are 52 times more likely to grow. Firms with both a patent and registered in Delaware are 269 times more likely to grow. Finally, firms associated with high technology industries are 53 percent more likely to grow while firms associated with local industries are 33 percent less likely. Interestingly, this small number of observables is able to account for 34 percent of all statistical variation (*pseudo*-R-squared).

Summary Statistics

Table 1 provides descriptive statistics for the main covariates. The dependent variables are firm-level outcomes, including access to VC funding (in 2 years and 6 years) and growth outcomes, such as IPO and M&A. We also provide summary statistics for controls, including firm observables, intellectual property observables, industry characteristics, and VC-targeted industry controls. Table 2 summarizes the proportion of ventures with female founders by key quality indicators. Only about 10 percent of new ventures with early patents are female-led. Female management is associated with about 22 percent of ventures with early trademark. About 16 percent of ventures in high-tech industries and 24 percent of ventures founded non-locally are female-run.

IV. MAIN RESULTS

Table 3 Panel A presents univariate regressions for incidence ratios between gender and a number of quality indicators. To facilitate the interpretation of our results, we present the results in terms of the odds-ratio coefficients. All specifications include sector dummies as well as VC targeted sector dummies. These analyses are consistent with the descriptives mentioned earlier: there are significant differences between male- and female-led ventures across a number of key indicators of entrepreneurial quality. In Panel B, we re-estimate the baseline specifications but include control variables. As shown in both panels, the coefficients on the *Female-led start-up* dummy are remarkably stable. Women-led start-ups have higher odds than men-led start-ups to be incorporated (Column 7), to have a trademark (Column 10), and to be local (Column 11), but they have lower odds to be incorporated in Delaware (Column 8), to have early patents (Column 9), and to operate in the high-tech sector (column 12). Overall, these tables mirror the patterns presented in Table 2 and Table 3. More generally, we find evidence that female-led start-ups are less likely to exhibit the markers of entrepreneurial quality at the time of founding.

In Table 4, we assess the baseline hypothesis by considering the probability of obtaining VC funding by women-led ventures relative to their counterparts run by men. As shown in column 1, women-led ventures are 63% less likely than men-led ventures to obtain VC funding. In column 2, we re-estimate this baseline specification but match female- and male-run ventures on entrepreneurial quality. As can be seen, gender gap decreases to 24%. Based on our findings, as much as 65% of the observed gender gap in access to entrepreneurial funding can be attributed to systematic sorting of females and males into ventures of differential growth potential. Finally, in Column (3), we re-estimate the baseline specification using a Monte Carlo procedure[7]; these findings mirror previous estimates, with the estimated gender gap equal to 22%.

Overall, our results lead to two important conclusions. First, while women are less likely to access venture capital than men, the majority of this gap arises because of the underlying differences in entrepreneurial quality. At the same time, female-run ventures continue to be less likely to access VC funding – even when compared to male-run ventures of equivalent quality. In the following section, we perform additional analyses to probe deeper the mechanisms that may potentially explain the residual gender-based differences we observed.

V. AUXILIARY RESULTS: STATISTICAL VERSUS TASTE-BASED DISCRIMINATION

Although our results indicate that the majority of gender-based gap in entrepreneurship might be driven by initial differences in resources across female and male-founded ventures, as much as 35% of the gap persists even when such initial differences are taken into consideration. Hence, in additional analyses, we probe the mechanisms likely to explain the residual gender differences in access to VC funding. Two distinct causal mechanisms may explain the persistent gender gap in access to VC funding: statistical discrimination and taste-based preferences. The former suggests that stereotypes are activated when information is limited or ambiguous (Arrow, 1977; Phelps, 1972); therefore the reliance on gender decreases as additional cues on individual merit become more easily available to resource holders or evaluators (Petersen and Saporta, 2004; Ridgeway and Correll, 2000). The latter suggests that evaluators rely on stereotypes even when it is possible to assess entrepreneurial quality directly (e.g., Correll and Ridgeway, 2003; Ridgeway, 1991) because ascriptive characteristics, such as gender, become imbued with beliefs and status (Ridgeway and Correll 2000), putting women at a significant disadvantage. Though the two theories yield similar predictions about the existence of gender-based differences in access to venture capital funding, they yield conflicting predictions regarding the conditions under which such biases might be prevalent. To tease out these two explanations, we therefore

[7] Given the small number of female founded firms that have at the high end, we prefer to match it with multiple different male-led firms. To do so, we find 100 random matches (with replacement) for each female firm, then estimate the coefficient 100 times and report the coefficient's empirical distribution.

exploit cross-sectional heterogeneity in: (a) the position within the quality distribution; and (b) evaluators' sophistication.

Position within Quality Distribution. We first turn to an exploration of a start-up position within the estimated quality distribution. Statistical discrimination theories imply that venture capitalists will be less likely to rely on negative gender stereotypes as a cue when uncertainty is lower. This theory predicts that the reliance on gender in the evaluation process will decrease, as new venture exhibits a stronger signal of quality associated with lower uncertainty about growth potential. Because the quality signal is stronger for ventures in a higher position of the estimated quality distribution, we expect that, conditional on equivalent quality, gender differences in access to VC funding will decrease as a venture's position in the quality distribution increases. By contrast, taste-based theories of discrimination imply that evaluators will discriminate against women regardless of the entrepreneurial quality signal. Hence, conditional on equivalent quality, the size of gender differences should be uniform at different levels of entrepreneurial quality.

The results are presented in Table 5 Panel A. Panel A in this table estimates gender-based differences in VC funding at different levels of entrepreneurial quality distribution. We classify entrepreneurial ventures based on differences in the entrepreneurial quality in the following way: (1) 0-95th percentile; (2) 95th-99 percentile; (3) 99th percentile; (4) 99.5th percentile; and (5) 99.9th percentile. This partition is particularly suitable to the VC context in which the modal investment outcome is a complete failure. Hall and Woodward (2010) report that about 50% of the venture capital-backed startups in their sample had zero-value exits. Similarly, Sahlman (2010) finds that 85% of returns come from just 10% of investments. Because successful exists are rare for venture capitalists, the latter tend to focus on investments with highest potential (a fact that we also validate within our quality distribution).

As can be seen in columns (1)-(5), conditional on matching on quality, gender gap is wider at lower levels of entrepreneurial quality – or for new ventures associated with greater uncertainty. In Column (1), within the 0-95 percentile, a start-up run by a female is 33% less likely to receive VC funding than a comparable-quality start-up that is run by a male. As can be seen in column (2), the gap decreases for ventures that fall between 95-99 percentile of quality: within this category, female-led ventures are 23% less likely to obtain VC funding, relative to male-led ventures of comparable entrepreneurial quality. Column (3) estimates the probability for new ventures that fall within the top 1% of quality. As can be seen, gender-based differences in access to VC capital continues to decrease: female-led ventures are 16% less likely to receive VC funding, relative to comparable male-led ventures. Column (4) and Column (5) further estimate the probability of VC funding for ventures at the top 0.5% and 0.1% of the estimated quality distribution, respectively. As can be seen, gender-based differences disappear entirely within those subsamples, indicating that female-run start-ups at the top of the estimated quality distribution are as likely as male-run start-ups in equivalent position of the estimated quality distribution to secure important entrepreneurial resources. In additional analyses, we tested whether the coefficients in these different models were different statistically. Our results are significant across a large number of sub-groups in the distribution. As can be seen, the coefficients are statistically different between firms in the bottom of our distribution (bottom 95%) and the 95-99 group; among the bottom 95% and the top 1%, and among the top 0.1% and the top 1%. Together, these findings are consistent with the theories of statistical discrimination, indicating that, gender-based differences in access to VC funding are not uniform across different levels of entrepreneurial quality; instead, as venture growth potential increases (and uncertainty decreases), evaluators are less likely to rely on gender to assess the potential quality of a new venture.

Non-Sophisticated Evaluators. As a second test, we examine whether the hypothesized effect of gender varies with the evaluators' sophistication. If, as hypothesized, statistical discrimination accounts for weaker gender gap, then we should also expect the gender gap to be higher amongst non-sophisticated investors. Less experienced evaluators tend to rely on signals more because they are less capable of discerning quality in a more direct manner (e.g., Jensen 2006; Podgorny, 1995). More generally, studies in this vein conclude that the significance of signals for evaluations declines as evaluators are able to better assess the candidate's quality more precisely and more directly. Hence, statistical-discrimination theories posit that gender gap in access to VC funding should be stronger for non-sophisticated investors. By contrast, taste-based theories of discrimination imply that gender biases will be uniform across non-sophisticated investors.

To account for the degree of VC sophistication, we follow Krishnan and Masoulis (2012) who calculate a reputation score for the top 1000 VCs between 1996 and 2002 based on past IPO performance. We use their score, expanded to 1995-2005. At the firm level, VC quality is the maximum of the Series A investors. We consider VCs as being non-sophisticated, when they fall into the bottom quartile of the VC-backed firms, based on the reputation score presented in Krishnan and Masoulis (2012).[8]

In Panel B, we examine the heterogeneous effect of gender on the probability of accessing VC funding for sophisticated and non-sophisticated VCs. To do so, we re-estimate the baseline specification for the two kinds of VCs, separately. Columns (6)-(8) report the estimates for the association between female-run ventures and VC funding for sophisticated VCs. Columns (9)-(11) re-estimate the same baseline specifications for non-sophisticated VCs. A can be seen in columns (6) and (9), the overall gender gap is greater for non-sophisticated VCs (an increase from 14% to 22%), consistent with the notion that less capable evaluators are less likely to rely on gender stereotypes to infer the potential value of a new venture.

It further follows that the quality of an entrepreneurial venture and the quality of an evaluator might both serve to mitigate gender gap in access to external funding and may thus work as substitutes. Hence, we expect that gender gap will be uniform across sophisticated and non-sophisticated evaluators at the top of the quality distribution. In subsequent columns, we decompose the heterogeneous effect according to venture-quality. The effects observed in columns (6) and (9) are driven by ventures outside the top 1% of the estimated quality distribution. As shown in columns (7) and (10), gender gap increases from 25% (for sophisticated VCs) to 56% (for non-sophisticated VCs). Interestingly, as shown in column (8) and (11), no significant differences exist for ventures at the top of the quality distribution. This finding implies that, when the signal of quality is less ambiguous - as is the case for high-quality ventures- the evaluators are less likely to differ in their assessment of female-run ventures. Taken together, these results provide consistent evidence that female-led ventures might face an unequal gap in that, for a comparable venture quality, they are less likely to have a constrained access to VC funding at the top of the quality distribution.

VI .AUXILIARY ANALYSES: ALTERNATIVE EXPLANATIONS

Gender Differences in Benefits of VC funding. A potential alternative explanation for our findings is that the gender gap in access to VC funding arises because women are less likely to benefit from VC funding than men. Fewer complementarities might exist between male-

---

[8] Notably, the period of 1995-2002 is a period where there was a considerable number of non-sophisticated investors in the market. Our list is mostly composed of short-lived funds such as the Boston University Community Technology Fund, and corporate venture capital funds such as the Compaq Computer Corporation.

dominated VCs and female-run ventures ex-post, thus reducing the motivation of venture capitalists to invest in female-run ventures. Alternatively, female-led ventures that receive VC funding may be less motivated than male-led ventures to pursue successful exit strategies ex-post, which would again discourage VCs from making ex-ante investments in female-run ventures. If this were the case, female-run ventures would underperform male-run ventures even conditional on access to VC funding.

To investigate this possibility, we examine whether the benefits of getting VC funding (i.e., their positive impact on growth outcomes) accrue differently for female- and male-led ventures. We consider two key growth outcomes: IPOs and acquisitions because the two are considered as important liquidity events. We then assess the heterogeneous effect of VC funding on those outcomes. As can be seen in Table 6 Column (1) and Column (3), VC funding has a positive and statistically significant impact on the probability of filing for an IPO as well as being acquired. Column (2) and Column (4) further show that the positive impact of VC funding on liquidity events is homogenous across male- and female-led ventures. This suggests that female-led ventures are equally likely as male-led ventures to benefit from access to venture capital.

VC Risk Taking. Another alternative explanation is that higher-quality VCs might not be better evaluators; instead, they might be more able and willing to take risk than lower-quality VCs. If so, then investment in female-led ventures may simply reflect greater propensity to invest in risky ventures. We investigate this possibility by examining whether VCs are more likely to invest in female-led ventures during hot markets. In doing so, we follow Nanda and Rhodes-Kropf (2013), who find that venture capitalists invest in riskier startups in hot markets.

In Table 7 Panel (A) and Panel (B), we re-estimate the baseline specification from Table 4 separately for boom years (Panel A) and bust years (Panel B). The results in Column (1) and Column (4) are comparable and similar to those shown in Table 4 and are not statistically different across boom and bust periods. That is, the *Female-led start-up* coefficient is lower than 1 and statistically significant (column 1) in both Panel (A) and Panel (B). Columns (2)-(3) and (5)-(6) further show that gender gap decreases as growth potential of a new venture increases and is 0 for the top 1% of new ventures in our sample. Hence, venture capitalists appear to be equally likely to make investments in female-led ventures in boom and bust years. In Panel C, we compare the results from Panel (A) and Panel (B) directly by estimating an interaction term between the female dummy and the boom-years dummy. As shown in column (7), we find that access to VC funding is indeed less constrained during boom periods, consistent with prior findings (Nanda and Rhodes-Kropf, 2013). However, the impact of hot markets on investments in entrepreneurial ventures does not vary across male and female-led ventures. Hence, given these findings, it is unlikely that VCs selectively invest in female-led ventures as a form of experimentation and risk taking.

Top Female Founders. Another concern is that women who pursue high-growth ventures may differ along observable and unobservable characteristics that may affect VC evaluation processes and willingness to invest. First, it might be that women outside the "top" quality distribution have relatively weaker networks than women at the top of the estimated quality distribution. Because many studies have related lower performance of women-run ventures to gender differences in network structure and composition (Aldrich, 1989; Cromie and Birley, 1992), this explanation is particularly credible. Nevertheless, this concern is unlikely to explain our results, for the following reasons. First, if the quality of female-led ventures is systematically correlated with women's access to networks, then we would expect that sophisticated VCs would better evaluate such differences. This implies that gender gap in obtaining VC funding should be greater for sophisticated VCs – which are presumably better at evaluating observed differences between female entrepreneurs. However, in Table 5 Panel (B) and (C), we find the opposite: as shown in Column (7) and Column (10), gender gap is greater for non-sophisticated than sophisticated VCs.

A related concern may be that women who pursue opportunities with lower growth potential lack confidence to seek VC funding, relative to women at the top of the quality distribution. However, this would imply that a female will be (intrinsically) more capable than a male counterpart at a similar level in quality distribution. Hence, female-led ventures should be more likely to raise financing relative to comparable male-led ventures. However, in Table 5 column (1) and column (2) we find the opposite: amongst lower-growth-potential ventures, those run by females are less likely to obtain VC funding those run by males.

A related and more general concern is that females who found or run start-ups may systematically differ along some unobserved dimensions correlated with the propensity to obtain VC funding from males in those same positions. Although we expect such unobserved differences to be accounted for with the index of entrepreneurial quality, since these differences would incline women to found new ventures of different quality, we nonetheless conduct additional analyses to probe this effect deeper. To rule out this explanation, we estimate a model with individual-fixed effects. This conditional logit specification removes time-invariant individual heterogeneity, but it cannot be estimated for individuals who never received VC funding. Although we lose many observations, we can estimate this model for serial entrepreneurs who founded more than one venture and who obtained VC funding for some (but not all) founded ventures.

In Table 8, we interact *Female-led start-up* dummy with a dummy indicating that an individual has been a serial entrepreneur. As can be seen, serial entrepreneurs are less likely to obtain VC funding than are non-serial entrepreneurs. At the same time, when the results are estimated "within the individual," the odds of getting VC funding are even lower if an entrepreneur is female, as indicated by the coefficient on the interaction term (*Serial Entrepreneur * Female*). This result is consistent with the notion that women-led ventures are less likely to receive VC funding than male-led ventures. Thus our results are unlikely to be attributed to unobserved qualities and skills.

VII. ROBUSTNESS CHECKS

Falsification Test. If gender, at least in part, drives differences in access to VC funding, we would expect gender differences to be *at least* mitigated when new ventures are composed of gender-mixed management teams. Presumably, a male representation on the management team is likely to compensate for the negative stereotypes, if such were associated with female founders. To assess this possibility, we re-estimate our specification for new ventures with gender-mixed management teams — in this case, we consider the gender of individuals registered as President, Treasurer, or Secretary. These data on top management teams are only available for the subsample of business registrants in Massachusetts. As shown in Table 9 column (1), there is gender gap in getting VC founding, when a firm registers in Massachusetts. Column (2) additionally shows that the effect becomes zero for ventures with gender-mixed management teams. Our inability to replicate the results for these ventures reinforces the notion that our results might reflect the effect of gender.

Generalizability. While our analyses focus on businesses registered in California and Massachusetts, one concern may be that these results only capture the effect of California and are not generalizable to other locations either because (a) female-led ventures outside California are consistently lower quality; or (b) VCs are less likely to statistically discriminate based on gender in other states. While plausible, the possibility that bias toward female-run ventures is systematically higher in California is unlikely – indeed, because California is home to Silicon Valley and vibrant entrepreneurial culture, we expect (a) and (b) to be unlikely in this state, making our empirical context conservative. However, to address this possibility formally, we re-estimate our baseline specification for the universe of business registrants in Massachusetts (for the same study period) and Texas. Because new ventures in Massachusetts are more commonly founded within the bio-tech sector, the gender gap in obtaining VC funding may be different. Similarly, because Texas is a southern state, it is worthwhile to investigate whether gender gap can also be replicated in this setting to alleviate the concern that our results are driven by ventures in northern states alone.

In Table 10, we replicate similar findings for businesses registered in Texas (Columns 1-2) and Massachusetts (Columns 3-4): the coefficient on the *Female-led start-up* dummy continues being lower than 1 with a high statistical significance across all model specifications. Indeed, the gender gap appears to be even wider in Texas, with women-led start-ups having 70% lower odds of getting VC funding than men-led start-ups, and in Massachusetts, with women-led start-ups having 33% lower odds of getting VC funding than men-led start-ups.

VC Funding Time Window. Another potential concern is that gender differences in access to VC funding may be an artifact of the two-year window we chose. Although the majority of ventures tend to obtain VC funding within the first two years, it is possible that women take longer than men to access venture capital. This raises the possibility that female-led ventures might be as likely to obtain VC funding as male-led ventures, when a longer time-window is considered. To address this concern, in Table 11 Panel A, we re-estimate the baseline specification to examine: (a) getting VC funding within a 6-year window; and (b) getting funding "ever" – or within the entire period under study. As can be seen in Table 11 Panel A, our results are unchanged if we focus on a longer time window. Hence, our results are not merely an artifact of different time horizons that female and male-run start-ups might adopt.

Alternative Time Periods. Another concern with our identification strategy might be the period under study. Perhaps gender gap in entrepreneurship has disappeared, if female entry into entrepreneurship has become more prevalent over time. If so, then our results are simply an artifact of the time period chosen. To see whether this possibility affects our results, in Table 12 Panel B, we re-estimate our baseline specification for different time windows: 1999-2000; 2001-2007; 2008-2011. As shown in columns (7) and (8) of Table 10, doing so is immaterial for our results: we find that there are significant gender differences in access to VC funding in each of the windows considered.

VIII. DISCUSSION

Past studies have shown that stark gender differences exist between men and women in entrepreneurship rates and new-venture performance (e.g., Aldrich, 2005; Reynolds, Carter, Gartner, and Greene, 2004; Ruef, Aldrich, and Carter, 2003; Yang and Aldrich, 2015). In this study, we complement past research by shedding more light on the drivers of this well-established empirical pattern. In particular, we revisit growing empirical evidence suggesting that the source of gender gap lies primarily in access to venture capital (Diana Project, 2014; Canning, Haque, and Wang, 2012), because investors tend to hold gendered preferences and biases against female-run ventures. Two theoretical traditions, one rooted in the sociological theories of homophily, similarity-attraction, in-group preference, and stereotypes, the other rooted in resource-based view, offer contrasting predictions regarding the primary source of the differences commonly observed in past accounts. In this study, we apply a novel empirical approach to adjudicate between these two mechanisms and assess the drivers of gender gap in entrepreneurship with greater empirical precision. Building on recent studies of entrepreneurial quality distribution (Guzman and Stern, 2015a; 2015b), we estimate the relative importance of each theoretical mechanism in driving the gap in access to funding for female- and male-led ventures.

Our findings confirm the well-established pattern that female-led ventures are significantly less likely to obtain venture capital funding. But contrary to the common assumption in the existing accounts, the initial differences in entrepreneurial quality are the primary source of this gap. In this regard, we show that women are significantly less likely than men to be involved as executives or founders in start-ups with the kinds of characteristics that correlate with successful venture growth in the future. Once female and male-led ventures are matched on quality, the residual gap decreases for ventures in higher positions within the quality distribution as well as more professional venture capitalists. These findings are consistent with the notion that the residual gender differences might be driven by statistical discrimination: gender is used as a cue to infer information when uncertainty is high or when evaluators are less capable and less experienced.

Collectively, our findings make several contributions. First, we contribute to the growing line of research on female entrepreneurship. Our analyses enrich a growing line of work on gender and entrepreneurship (e.g., Kim, Aldrich and Keister 2006; Loscocco et al. 1991; Kalleberg and Leicht 1991), as well as work that relates to the processes of discrimination on the part of investors (e.g., Jennings and Brush 2013). We show that, while these processes play a role in generating gender imbalance in entrepreneurship, they are not the key source of the gap. Instead, our study suggests that initial resource disparities across female and male-led ventures play a primary role in generating the observed empirical pattern. Second, our work also finds important heterogeneity in the gender gap across the firm quality distribution. We hypothesize that other findings on the gender gap might also vary by the quality and type of entrepreneurial firm, such that evidence found within one type of entrepreneurs need to carry to others. Going forward, we encourage researchers to establish clearly the type of entrepreneurship they are studying before analyzing the role of gender or other minority groups within them.

More generally, our study makes a contribution to work on gender in the strategic context. A vast number of strategy scholars have recognized the role of gender in strategy, but these studies have mostly focused on female participation rates in corporate boards (e.g., Helfat et al., 2006; Hillman, Shropshire, and Cannella, 2007), CEO and executive positions (Cook and Glass, 2014; Hill et al., 2014; Heilman *et al.*, 1989), or managerial roles (Blum, Fields and Goodman, 1994; Petersen and Morgan, 1995) – and examined its influence on important firm outcomes, ranging from firm performance (e.g., Dezso and Ross, 2012; Hillman *et al.*, 2007; Matsa and Miller, 2011), to investors' reaction (e.g., James and Lee, 2007). Yet, the role of gender in driving strategic outcomes in the entrepreneurial context has been less well explored. Hence, our study contributes to this line of inquiry, shedding light on how gender might shape strategic outcomes, such as access to VC funding, in the context of entrepreneurial firms.

Our results have important policy implications. Findings presented in the study lead to a natural focus on interventions that improve the net-new creation of high quality entrepreneurship rather than the performance of existing ones. Policies aimed at allowing women to create more and higher potential firms —such as improving technological education, mentoring and career aspirations, and developing support mechanisms within the family— are likely to have a higher impact in reducing the gender gap in entrepreneurship we documented in this study.

Our findings open up attractive opportunities for future research. First, while our study provides evidence that initial differences in entrepreneurial quality drive the well-established gender gap in access to venture capital, it does not shed any light on the drivers of such differences. Future research could therefore profitably explore the reasons why women tend to found or lead ventures of lower entrepreneurial quality. While these reasons are theorized in our study, further empirical inquiry could investigate in greater depth the drivers of the quality differences we document. Moreover, our study shows that gender differences are likely to be weaker and even non-existent for top performing firms, opening up attractive opportunities for further inquiry. Future studies may, for example, want to further assess whether top-performing start-ups led by females may, under some conditions, gain advantage over start-ups led by males, and reach critical entrepreneurial milestones. For example, future studies may want to assess when venture capitalists are more likely to invest in top-performing female entrepreneurs. Finally, while our study focuses on female entrepreneurs, the mechanisms that we find are more general, and could possibly apply to other types of minority groups or under other contexts for female professionals, many of which would be interesting areas of future research.

REFERENCES

Aldrich, H. E. 2005. "Entrepreneurship." Pp. 451-477 in *The Handbook of Economic Sociology, Second Edition,* edited by Neil J. Smelser and Richard Swedberg. Princeton, NJ, Princeton University Press.

Alvarez, S. & Busenitz, L. (2001). The entrepreneurship of resource-based theory. Journal of Management, 6, 755–775.

Barney JB. 1991. Firm resources and sustained competitive advantage. *Journal of Management* 17(1): 99–120.

Blanchflower, David G., Phillip Levine and David Zimmerman. 2003. "Discrimination in the Small-Business Credit Market." The Review of Economics and Statistics 85(4): 930–943.

Birley, S., 1989. "Female Entrepreneurs: Are They Really any Different?" Journal of Small Business Management, 27(1): 32-37.

Brush, C.G., N.M. Carter, E. Gatewood, P.G. Greene, and M. Hart (2003). Venture Capital Access: Is Gender an Issue? pp. 141-154 in D. Hart, ed., The Emergence of Entrepreneurship Policy: Governance, Start-ups and Growth in the U.S. Knowledge Economy London: Cambridge University Press.

Brush, C., Carter, N., Gatewood, E., Greene, P., and Hart, M., 2006. Women's Entrepreneurship in the United States. In Brush, C., Carter, N., Gatewood, E., Greene, P., and Hart, M. (Eds.), Growth Oriented Women Entrepreneurs and Their Businesses, Cheltenham, UK: Edward Elgar.

Buttner, E., and Moore, D., 1997. "Women's Organizational Exodus to Entrepreneurship: Self-Reported Motivations and Correlates with Success." Journal of Small Business Management, 26: 31-35.

Canning, J., Haque, M. & Wang, Y. (2012). Women at the Wheel: Do Female Executives Drive Start-Up Success? Dow Jones and Company.

Carnahan, S., Agarwal, R., Campbell, B.A., 2012. Heterogeneity in turnover: The effect of relative compensation dispersion of firms on the mobility and entrepreneurship of extreme performers. Strategic Management Journal 33, 1411-1430.

Carter, N., and Allen, K., 1997. "Size-determinants of Women-Owned Businesses: Choice or Barriers to Resources." Entrepreneurship and Regional Development, 9(3): 211-220.

Castilla, E. J. 2008. Gender, Race, and Meritocracy in Organizational Careers. The American Journal of Sociology, 113 (6): 1479–1526.

Cavalluzzo, Ken and Linda Cavalluzzo and John D. Wolken. 2002. "Competition, Small Business Financing, and Discrimination: Evidence From a New Survey." The Journal of Business. 75(4): 641-679.

Carter, Sara, E. Shaw, W. Lam, and F. Wilson. 2007. "Gender, entrepreneurship, and bank lending: The criteria and processes used by bank loan officers in assessing applications." Entrepreneurship Theory and Practice 31:427–444.

Coleman, Susan and Alicia Robb. 2009. "A Comparison of New Firm Financing by Gender: Evidence from the Kauffman Firm Survey Data." Small Business Economics 33:397-411.

Carter, N., Gartner, W., Shaver, K., and Gatewood, E., 2003. "The Career Reasons of Nascent Entrepreneurs," Journal of Business Venturing, 18: 13-39.

Cavalluzzo, Ken and Linda Cavalluzzo and John D. Wolken. 2002. Competition, Small Business Financing, and Discrimination: Evidence From a New Survey. *The Journal of Business.* 75(4): 641-679.

Gatewood, E.J., Carter, N.M., Brush, C.G., Greene, P.G., & Hart, M.M. (2003). Women entrepreneurs, their ventures, and the venture capital industry: An annotated bibliography. Stockholm: Entrepreneurship and Small Business Research Institute.

Greene, P.G., Hart, M.M., Gatewood, E.J., Brush, C.G., & Carter, N.M. (2003). Women entrepreneurs: Moving front and center: An overview of research and theory. Coleman White Paper Series. Available at http://www.usasbe.org, accessed 17 July 2005.

Guzman, Jorge, and Scott Stern. 2015a. "Where is Silicon Valley?" *Science*. Vol. 347. Issue #6222.

Guzman, Jorge, and Scott Stern. 2015b. "Nowcasting and Placecasting Entrepreneurial Quality and Performance" *NBER Working Paper #20954*

Guzman, Jorge, and Scott Stern. 2016. "Zipf's Law in Entrepreneurial Quality" *mimeo*

Haines, G.H., Jr., B.J. Orser and A.L. Riding. 1999. "Myths and realities: An empirical study of banks and the gender of small business clients." *Canadian Journal of Administrative Sciences* 16: 291–308.

Jennings, J.E., Brush, C.G., 2013. Research on women entrepreneurs: challenges to (and from) the broader entrepreneurship literature? Acad. Manag. Ann. 7 (1), 663–715

Kalleberg, A., and Leicht, K., 1991. "Gender and Organizational Performance: Determinants of Small Business Survival and Success," Academy of Management Journal, 34(1): 136-161.

Kim, P. H., H. E. Aldrich, L. A. Keister. 2006. Access (Not) Denied: The Impact of Financial, Human, and Cultural Capital on Entrepreneurial Entry in the United States. *Small Business Economics* 27(1) 5–22.

Kortum, Samuel S. and Josh. Lerner. 2000. Assessing the contribution of venture capital to innovation. Rand Journal of Economics 31(4):674-692.

Krishnan, CVN, and R. Masulis, 2012, Venture Capital Reputation: A Survey (Handbook of Entrepreneurial Finance, Venture Capital and Private Equity, D. Cummings, Editor, Oxford University Press.)

Lippman SA, Rumelt RP. 2003a. A bargaining perspective on resource advantage. Strategic Management Journal 24(11): 1069-1086.

Loscocco, Karyn A., Joyce Robinson, Richard H. Hall, and John K. Allen. 1991. "Gender and small business success: An inquiry into women's relative disadvantage." Social Forces 70(1):65-85.

McPherson, M., Smith-Lovin, L., & Cook, J. M. 2001. Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology, 27: 415-444.

Nanda, R., Rhodes-Kropf, M., 2013. Innovation and the financial guillotine. Unpublished working paper. Harvard University, Cambridge, MA.

Orser, B.J., A.L. Riding and K. Manley. 2006. Women entrepreneurs and financial capital. Entrepreneurship Theory & Practice 30: 643–665.

Petersen, T., & I. Saporta. 2004. The Opportunity Structure for Discrimination. The American Journal of Sociology, 109: 852-901.

Schoar, Antoinette. 2010. "The Divide between Subsistence and Transformational Entrepreneurship" Chapter in NBER book Innovation Policy and the Economy, Volume 10. Edited by Josh Lerner and Scott Stern. p. 57 - 81

Reynolds, P., 2000. "National Panel Study Of U.S. Business Start-Ups: Background And Methodology,"

Databases for the Study of Entrepreneurship, 4: 153-227.

Robb, A., Coleman, S., & Stangler, D. (2014). Sources of Economic Hope: Women's Entrepreneurship. Retrieved from http://www.kauffman.org/~/media/kauffman_org/ percent20covers/2014/11/sources_of_economic_hope_womens_entrepreneurship.pdf

Ridgeway, C. L., & S. J. Correll. 2006. Consensus and the Creation of Status Beliefs. Social Forces, 85: 431–453.

Ruef, M., H. E. Aldrich, N. Carter. 2003. The Structure of Organizational Founding Teams: Homophily, Strong Ties, and Isolation Among U.S. Entrepreneurs. *American Sociological Review* 68(2) 195–222.

Shane, S, Sharon D, Joseph Jankowski, Isabelle Reymen, and Georges Romme. 2012."Which Inventors do Technology Licensing Officers Favor for Start-Ups?," Frontiers of Entrepreneurship Research 32 (18), Article 1.

Sørensen, Jesper B., and Amanda J. Sharkey. 2014. "Entrepreneurship as a MobilitY Process." *American Sociological Review* 79 (2): 328–49.

Tajfel, H., & Turner, J. C. 1979. An integrative theory of intergroup conflict. In W. G. Austin and S. Worchel (eds.), The social psychology of intergroup relations, Monterey, CA: Brooks/Cole

Thebaud, Sarah. 2015. "Business as Plan B Institutional Foundations of Gender Inequality in Entrepreneurship across 24 Industrialized Countries." Administrative Science Quarterly 60:671–711.

Tsui, A. S., & O'Reilly, C. A. (1989). Beyond Simple Demographic Effects: The Importance of Relational Demography in Superior-Subordinate Dyads. Academy of Management

Journal, 32: 402-423.

Yang, Tiantian and Howard E. Aldrich. 2014. "Who's the boss? Explaining gender inequality in entrepreneurial teams." *American Sociological Review* 79(2): 303-327.

Table 1: Summary Statistics

	N	Mean	St. Dev.	Sum
Year	1875087	2004.1653	4.4917	3.76E+09
Gender Measures				
Female-led Start-Up	1875087	0.2212	0.415	414682
Firm Outcome Measures				
Growth (IPO or M&A in 6 years)	1875087	0.0008	0.0278	1455
VC Series A in 2 Years	1875087	0.0021	0.0454	3871
VC Series A in 6 Years	1875087	0.0026	0.0506	4815
Low Quality VC	2064	0.2485	0.4323	513
Firm Observables				
Corporation	1875087	6.30E-01	0.4828	1.18E+06
Short Name	1875087	0.5064	0.5	949510
Eponymous	1875087	1.67E-01	0.3727	312607
Delaware	1875087	0.0466	0.2108	87366
Intellectual Property Observables				
Patent	1875087	0.0045	0.0666	8349
Trademark	1875087	0.003	0.0548	5643
Broad Industry Controls				
Local	1875087	0.1556	0.3625	291774
Traded High Technology	1875087	0.0534	0.2249	100159
Traded Resource Intensive	1875087	0.1091	0.3118	204610
Traded	1875087	0.5365	0.4987	1006077
VC Targeted Industry Controls				
IT Sector	1875087	0.0281	0.1652	52633
Biotech Sector	1875087	0.0028	0.0529	5255
Ecommerce Sector	1875087	0.0458	0.2091	85915
Semiconductor Sector	1875087	0.001	0.0309	1788

Table 2: Quality Indicators and Female-led Start-ups (California and Massachusetts, N=1,856,375)

	Share of Female-led Start-ups
All Firms	21.29%
Early Patenting Firms	9.94%
Early Trademark Firms	21.63%
Firms in High Tech Industries	15.90%
Firms in Local Industries	23.75%

Table 3: Probability of Covariates given Female-led Start-up

	Panel A: Univariate Regressions					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Corporation	**Delaware**	**Patent**	**Trademark**	**Local Firm**	**High Tech Firm**
Female-led Start-up	1.311***	0.478***	0.386***	0.969	1.187***	0.669***
	(0.00490)	(0.00501)	(0.0142)	(0.0314)	(0.00559)	(0.00586)
Broad Sector Dummies	No	No	No	No	No	No
VC Targeted Sector Dummies	No	No	No	No	No	No
Observations	1875087	1875087	1875087	1875087	1875087	1875087
R-squared	0.002	0.008	0.008	0.000	0.001	0.003

	Panel B: Controls Included					
	(7)	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
	Corporation	**Delaware**	**Patent**	**Trademark**	**Local Firm**	**High Tech Firm**
Female-led Start-up	1.309***	0.525***	0.540***	1.186***	1.129***	0.784***
	(0.00493)	(0.00559)	(0.0203)	(0.0391)	(0.00550)	(0.00815)
Corporation	1.239***	0.976***	2.943***	1.671***	0.267***	0.892***
	(0.00408)	(0.00713)	(0.0852)	(0.0519)	(0.00124)	(0.00743)
Short Name	1.195***	0.842***	0.384***	0.444***	1.117***	0.554***
	(0.00503)	(0.00881)	(0.0217)	(0.0245)	(0.00593)	(0.00746)
Eponymous	0.448***		17.94***	6.430***	0.423***	0.964*
	(0.00320)		(0.449)	(0.228)	(0.00587)	(0.0168)
Delaware	2.666***	20.53***		6.533***	0.327***	1.832***
	(0.0778)	(0.532)		(0.380)	(0.0241)	(0.0725)
Patent	1.201***	6.662***	6.297***		0.505***	1.279***
	(0.0363)	(0.240)	(0.377)		(0.0300)	(0.0800)
Trademark		0.448***	1.943***	1.074*	1.157***	1.227***
		(0.00320)	(0.0548)	(0.0321)	(0.00509)	(0.0106)
Broad Sector Dummies	Yes	Yes	Yes	Yes	Yes	Yes
VC Targeted Sector Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1875087	1875087	1875087	1875087	1875087	1875087
Pseudo R-squared	0.014	0.064	0.231	0.081	0.073	0.336

Incidence Ratios. Robust standard errors in parenthesis. * p<0.05 ** p<0.01 *** p<0.001. US CMP Cluster Dummies do not included Local and Traded in Model B-5 and do not include Local and High Tech in Model B-6

Table 4: The Probability of Female-led Ventures on Getting VC Funding in 2 Years

	Estimate Effect of Female-led start-up		
	(1)	**(2)**	**(3)**
	Logit Regression		**Bootstrapped Estimate**
Female-led Start-up	0.368***	0.756***	0.781***
	(0.0202)	(0.0439)	(.0285)
Corporation		19.04***	
		(1.783)	
Short Name		4.010***	
		(0.236)	
Eponymous		0.117***	
		(0.0277)	
Delaware Only		143.7***	
		(8.416)	
Patent Only		47.82***	
		(6.442)	
Patent and Delaware		554.3***	
		(38.02)	
Trademark		1.373***	
		(0.124)	
Broad Sector Dummies	No	Yes	
VC Targeted Sector Dummies	No	Yes	
Observations	1875087	1875087	
Pseudo R-squared	0.008	0.483	

Table 5: Logit Regression Female-led Ventures at Different Levels of Firm and VC Sophistication. Matched Estimates on Entrepreneurial Quality.

Panel A: P(VC Financing in 2 Years)					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	0-95 percentile	**95-99 percentile**	**top 1%**	**top 0.5%**	**top 0.1%**
Regression Coefficient					
Female-led Start-up	0.674***	0.771***	0.842**	.892	1.02
	(0.066)	(0.060)	(0.053)	(.071)	(0.175)
Summary Stats					
Observations	1,763,556	74,255	18,564	9282	1,857
Total Funded Firms	302	736	2,121	1,324	480
# Female-led VC Funded Firms	48	72	192	115	40
Share of All Observations	0.0027%	0.10%	1.03%	1.24%	2.15%
# Male-led Growth Firms	254	664	1929	1209	440
Share of All Observations	0.0144%	0.89%	10.39%	13.03%	23.69%
t-tests	**Cols (1) & (2)**	**Cols (1) & (3)**	**Cols (2) & (3)**	**Cols (3) & (4)**	**Cols (3) & (5)**
T-Statistic of Difference in Means	1.54	2.87	1.25	0.10	1.38
T-Statistic p-value (df=100-1)	0.06	0.00	0.11	0.46	0.09

Panel B: P(VC Financing in 2 Years by Sophisticated VC) (1995-2005 Only)(1)

	(6) All Firms	(7) 0-99 percentile	(8) top 1%	*t*-test of Difference in Means for Cols (7) & (8)	
Regression Coefficient					
Female-led start-up	0.857*	0.749**	0.947	*t-statistic*	1.92
	(0.074)	(0.097)	(0.109)	*p-value*	0.03

Panel C: P(VC Financing in 2 Years by Non-Sophisticated VC) (1995-2005 Only)(1)

	(9) All Firms	(10) 0-99 percentile	(11) top 1%	*t*-test of Difference in Means for Cols (10) & (11)	
Regression Coefficient					
Female-led start-up	0.775+	.440***	1.033	*t-statistic*	3.30
	(0.137)	(.104)	(0.232)	*p-value*	0.00

Ratios Reported; Bootstrapped standard errors in parentheses * p<0.05 ** p<0.01 *** p<0.001.

Table 6: Impact of VC on Outcomes (Only 1995-2005)

	DV: IPO		*DV: Acquisition*	
	(1) IPO Firms	(2) IPO Firms	(3) Acq All Firms	(4) Acq All Firms
VC Series A in 2 Years	1.822***	1.835***	4.274***	4.375***
	(0.213)	(0.219)	(0.358)	(0.375)
VC Series A in 2 Years X Female-led start-up		0.914		0.731
		(0.328)		(0.191)
Corporation	21.87***	21.87***	4.453***	4.455***
	(5.884)	(5.885)	(0.404)	(0.404)
Short Name	1.798***	1.798***	2.134***	2.135***
	(0.165)	(0.165)	(0.122)	(0.122)
Eponymous	0.474***	0.474***	0.344***	0.344***
	(0.106)	(0.106)	(0.0469)	(0.0469)
Delaware Only	75.87***	75.88***	31.28***	31.29***
	(8.098)	(8.099)	(1.833)	(1.833)
Patent Only	66.38***	66.40***	31.44***	31.48***
	(13.06)	(13.06)	(3.847)	(3.850)
Patent and Delaware	517.0***	517.1***	144.7***	144.7***
	(67.93)	(67.94)	(12.31)	(12.32)

Trademark	4.784***	4.782***	5.322***	5.314***
	(0.592)	(0.592)	(0.527)	(0.527)
Broad Sector Dummies	Yes	Yes	Yes	Yes
VC Targeted Sector Dummies	Yes	Yes	Yes	Yes
Observations	1442015	1442015	1442015	1442015
Pseudo R-squared	0.415	0.415	0.333	0.333

Incidence Ratios Reported; Standard errors in parentheses * p<0.05 ** p<0.01 *** p<0.001. Only firms up to 2005 used to allow enough time for growth events to occur in our sample.

Table 7: Logit Regression Impact of Female-led Start-Ups at Different Levels of Firm and VC Sophistication During Boom and Bust Cycles

Panel A: P(VC Financing in 2 Years) during Boom Years (1996-2001) - Split

	(1) All Firms	(2) 0-99 percentile	(3) top 1%	*t*-test of Difference in Means for Cols (1) & (4)	
Regression Coefficient					
Female-led start-up	0.811**	0.690**	0.946	*T-Statistic*	0.67
	(0.063)	(0.080)	(0.114)	*p-value*	0.25

Panel B: P(VC Financing in 2 Years) during Bust Years (2002-2005) Split

	(4) All Firms	(5) 0-99 percentile	(6) top 1%
Regression Coefficient			
Female-led start-up	0.857*	0.704**	0.877
	(0.074)	(0.112)	(0.144)

Panel C: Interaction

	(7) All Firms
Regression Coefficient	
Female-led start-up	0.826+
	(0.093)
Boom	2.37**
	(0.357)
Female-led start-up * Boom	0.993
	(0.150)

Incidence Ratios Reported. + p< 0.1, * p<.05, ** p < .01

Table 8: Serial Entrepreneurship Using Individual Fixed-Effects

	(1) VC Series A in 2 Years
Female-led start-up	0.200
	(1.502)
Serial Entrepreneur	0.385**
	(0.118)
Serial Entrepreneur * Female-led start-up	0.0288*

	(0.0474)
Observations	489
Pseudo R-squared	0.535

Only relevant covariates included. Exponentiated coefficients; Standard errors in parentheses. * p<0.05 ** p<0.01 *** p<0.001. Column (1) defines serial entrepreneurs as individuals with the same name, state, and city. Regressions are run at the individual level - this means that in the cases of more than one firm manager (e.g. LLCs), the firm appears once per each manager.

Table 9: The Probability of Female-led Ventures on Getting VC Funding in 2 Years: Massachusetts Firms. Includes mixed gender teams.

	(1)	(2)
Female-led start-up	0.584**	0.523***
	(0.108)	(0.101)
Mixed-Gender-led start-up		1.031
		(0.138)
Cluster Dummies	Yes	Yes
N	289278	367267
Pseudo R2	0.449	0.445

All models include all controls for venture quality, as in Table 4 Model 2. Model 1 only includes firms with all female or all male founders, model 2 includes mixed gender teams. Mixed gender teams compose 42% of the sample. Female composes 11.23% of the sample. We consider part of the team only those registered as President, Treasurer or Secretary.

Table 10: The Probability of Female-led Ventures on Getting VC Funding in 2 Years: Texas and Massachusetts

	(1) Texas	(2) Texas	(3) Massachusetts	(4) Massachuse
Female-led start-up	0.0995***	0.295**	0.301***	0.667**
	(0.0412)	(0.125)	(0.0439)	(0.101)
Broad Sector Dummies	No	Yes	No	Yes
VC Targeted Sector Dummies	No	Yes	No	Yes
Observations	702455	702455	396635	396635
Pseudo R-squared	0.017	0.392	0.010	0.428

All models include all controls for venture quality, as in Table 4 Model 2.
Exponentiated coefficients; Standard errors in parentheses. * $p<0.05$ ** $p<0.01$ *** $p<0.001$.

Table 11: Robustness Tests

Panel A: Distribution of effect by time to VC financing

	(1) Gets VC in 2 Years	(2) Gets VC in 6 Years	(3) Gets VC Ever
Female-led start-up	0.781***	0.771***	0.798***

	(.028)	(.025)	(.027)

Panel B: Distribution of effect through time periods

	(4) **1995-2000**	**(5)** **2001-2007**	**(6)** **2008-2011**
Female-led start-up	0.805***	0.744***	0.766***
	(.054)	(.044)	(.056)

Only firms between 1995 and 2005 used in analysis. Incidence Ratios. Robust standard errors in parenthesis. Matched sample. Quality Ranges Very High top1%; High between 95% and 99%; Medium between 75% and 95%; Low less than 75%. Panels C and D use as a dependent variable a dummy equal to 1 of a firm patents (trademarks) between years 2 to 6, thus excluding the first year which is included in the quality calculation of the company. + p<.10 * p<0.05 ** p<0.01

APPENDIX TABLES

Table 1A: Quality Estimation Model

	(1)
Corporation	5.751***
	(0.681)
Eponymous	0.301***
	(0.067)
Short Name	2.458***
	(0.202)
Trademark	3.874***
	(0.470)
interactions	
Patent Only	34.70***
	[6.858]
Delaware Only	51.67***
	(4.374)
Patent and Delaware	268.93***
	(26.672)
Industry Dummies	
Local	0.668*
	(0.120)
Traded High Technology	1.525***
	(0.146)
Traded Resource Intensive	0.766*
	(0.093)
Traded	1.107
	(0.079)

Observations	1,064,914
Pseudo R-squared	0.34

Standard errors in brackets. Incidence Ratios Reported.

CHAPTER 4

ENTREPRENEURIAL MIGRATION

ABSTRACT

I study the migration of high growth entrepreneurs to out-of-state MSAs in the United States. Migration is found to be meaningful economically, and most migration is entrepreneurial. Higher quality firms are more likely to migrate; migrant quality is as high, but not higher, than local firms in the destination MSA. I then investigate the role of MSA characteristics in migration. Using count models with MSA and year fixed-effects, I find improvements in the quality-adjusted quantity of entrepreneurship predict migration into the MSA, but changes in simply the quantity of firms do not. Measures of MSA innovation (patenting) and cost of living (median home value) do not predict migration once fixed effects are included; venture capital does, but this relationship loses statistical significance once the entrepreneurial ecosystem is accounted for. The effect holds across the firm quality distribution, even for firms at the top 0.5%. This chapter provides novel evidence on the migration of high growth entrepreneurs and their patterns in the U.S. economy.

I. INTRODUCTION

A good location provides many benefits to entrepreneurs, such as proximity to investors, customers, and suppliers, a capable labor force, and access to novel ideas and institutions[1]. Since some new firms are likely to be born far away from their ideal location, the process of sorting into locations should be a key interest of research in entrepreneurship geography, strategy, and regional economics. However, little research exists on the topic; the main assumption across a range of literatures appears to be that, while employees do move across regions, firms simply grow (or fail to do so) in the place they are born[2]. This assumption stands in sharp contrast to the real-world dynamics of high growth entrepreneurship: many startups, including Amazon, Microsoft, Netscape, Facebook, and Dropbox, have moved early in their lifetimes. The purpose of this paper is to bridge this disconnect by performing a systematic exploration of migration in high growth entrepreneurship across metropolitan statistical areas (MSA) in the United States.

[1] Marshall (1890), Glaeser and Goettlieb, 2009, Ellison et al (2010), and Delgado et al (2010).

[2] This is the assumption, for example, in the main theoretical models of economic geography such as Roback (1982), and Krugman (1991).

Consider the case of Microsoft. Bill Gates and Paul Allen started the company in Cambridge, Massachusetts. They quickly moved to Albuquerque, New Mexico, to follow their first customer. However, a few years later, they realized they were having trouble hiring and moved to Redmond, Washington, which offered a better software engineering labor force and where they had a personal network. Microsoft's location choices were driven by Marshallian agglomeration forces—such as the distance to customers or workers—but they were updated through time, changing the optimal location choice and causing the firm to migrate. Migration costs might have been meaningful, but they were not insurmountable.

How common is migration? What makes firms migrate, even after being founded? What can we learn from the choice to migrate by these entrepreneurs?

To study these questions, I develop a new dataset using business registration records and entrepreneurial quality estimates (Guzman and Stern, 2015) with four useful features. First, I am able to identify a subset of entrepreneurs with a signal of growth intention, independent of actual performance: those firms registered under Delaware jurisdiction. Second, I am able to track the migration of high growth entrepreneurs to out-of-state MSAs, as business registration records provide digital "breadcrumbs" of firm location. Third, I include entrepreneurial quality estimates (the firm potential at birth) at the firm level, and measures of the quality, quantity, and potential (quality-adjusted quantity) at the regional level, which allow me to characterize the heterogeneity of entrepreneurship across locations and firms and their role in influencing migration. And fourth, I include other measures of entrepreneurial ecosystems at the MSA level—such as the level of "bohemia" (as in Florida, 2002), the MSA GDP, cost of living, venture capital, and patenting—which allow me to study the relationship of migration to some common measures of ecosystems. The dataset covers 25 US states, accounting for 60% of the U.S. by 2013 GDP, and the set of firms at risk of migration represent all Delaware corporations, partnership, and limited liability companies founded between 1988 and 2012 in those states.

Using this data, this paper delivers three interrelated sets of results.

The first set is theoretical, I outline the five main reasons that might lead a firm to migrate systematically after founding: migration costs, arbitrage between locations, uncertainty in location, dynamic firm strategies, and macroeconomic changes.

The second set is the main facts of migration (within my data). Migration is relatively common amongst Delaware firms, about 10% of firms migrate. Because I am only able to observe migration between my subset of states, this is likely to be an underestimate. Migration is also entrepreneurial, the risk of migration is highest in the first year of life for the firm and decreases monotonically as the firm ages. The quality of movers is lower than non-movers, but they also disproportionately come from lower quality ecosystems. Controlling for the entrepreneurial quality of the birth state, firms of higher quality are more likely to migrate. The quality of these firms appears as good, but not better, than the average firm in the destination.

The third set of results is the relationship between migration rates and the characteristics of the MSAs to which they move through count data regressions on a panel of 162 MSAs. I begin by looking at the relationship of migration counts to five *common* entrepreneurial ecosystem measures—bohemia, MSA GDP, median housing price (cost of living), venture capital, and patenting—, and two *new* ecosystem measures—entrepreneurial quantity (the quantity of new firms) and entrepreneurial potential (the quality-adjusted quantity of new firms). In individual pooled regressions (one for each measure), with year fixed effects and independent variables lagged by one year, all coefficients are positive and significant. However, on (MSA) fixed-effects regressions, the results change. Bohemia and MSA GDP are excluded as there is little variation left after MSA fixed-effects on which to run the regression. Of the remaining measures, only two—venture capital and the MSA entrepreneurial potential—continue to relate to migration rates into a region after including MSA fixed effects. Cost of living, number of new patents, and the MSA entrepreneurial quantity do not. The coefficient of entrepreneurial potential is an order of magnitude higher than venture capital and, when the two measures are included together, the significance of venture capital disappears.

In the last section, this paper investigates the differences of these effects across the entrepreneurial quality distribution. The effects are surprisingly similar all the way to the highest categorization that is possible (due to power considerations) within my data, firms at the top 0.5% of firm quality.

The results in this paper contribute to three distinct strands of literature. First, its main contribution is to the literature at the nexus of geography and entrepreneurship (Guzman and Stern,2017; Delgado et al 2010; Glaeser, 2007; Gleaser et al, 2014; Kerr and Nanda, 2010; Samila and Sorenson, 2011), in three ways: by highlighting entrepreneurial migration as an important area of study, by showing how migration can be used to observe the revealed preference of entrepreneurial firms for different locations, and by using this choice to shed light on attributes of regions that firms are valuing in their migration choices. Second, it contributes to the nascent literature that incorporates location choice within entrepreneurial strategy. Since different locations offer different benefits to firms, the choice of location is part of the key choices entrepreneurs undertake as they seek to develop a competitive position. While some prior work has hinted at the importance of this choice (Gans and Stern, 2016; Gans, Stern, and Wu, 2016; Arzhagi and Henderson, 2008), my paper focuses on this topic more intently, and offers initial results on this area of study[3]. Third, it contributes to the literature on regional dynamics, economic clusters, and their evolution. In particular, recent studies have highlighted large geographic divergence within the United States in wellbeing, productivity, wages, and innovation (Hsieh and Moretti, 2015; Diamond, 2016; Forman et al, 2016). This paper contributes by highlighting how dynamics of location changes in firms, not only workers, might also contributes to this divergence. Furthermore, by documenting the migration of firms in the United States, this paper provides evidence on one of the mechanisms through which economic clusters (Porter, 2003; Delgado et al, 2014) might gather strength, an area of recent debate in the literature (see Duranton, 2011).

Finally, at a policy level, this paper offers a refinement of regional entrepreneurship policy by including migration in policy design. When one takes migration seriously, the focus of local policies shifts beyond simply creating high growth firms, to policies that also consider the likelihood that these firms stay. And policies that focus on attracting entrepreneurs, not simply creating them locally, become feasible[4].

The results of this paper proceed in seven subsequent sections.

In Section II, I outline the economic intuition of why would high growth entrepreneurs migrate at all.

In Section III, I introduce the dataset, and explain the approach to measure regional entrepreneurial quality, quantity, and potential; the approach to measure migration; and describe the five common measures of entrepreneurial ecosystems at the MSA level. The dataset is presented in two formats, at the firm level—with each firm being an observation—and as a panel

[3] Dahl and Sorenson (2011) and Michelacci and Silva (2007) are also related papers, but they study differences in performance of firms given the time the founder has lived in the region, rather than the act of firm migration.

[4] Gonzalez-Uribe and Leatherbee (2016) develop initial estimates on the ambitious program of Startup Chile, which provided ample incentives for international entrepreneurs to move to Chile temporarily, finding significant regional spillovers from the program.

of MSAs, with MSA observables and migration counts into each MSA. Section IV presents the summary statistics of this dataset as well as some basic distributions of the sample data.

The main results are in Sections V, VI, and VII.

In Section V I present an analysis of the basic facts of migration using the firm-level dataset of all migrants. In Section VI I present the results of count data regressions on the panel of 162 MSAs, including the relationship of migration counts to common entrepreneurial ecosystem measures, and to two new entrepreneurship ecosystem measures—entrepreneurial quantity, and entrepreneurial potential. Finally, in Section VII, I investigate the effect of these observables on the migration rates of firms at different levels of the entrepreneurial quality distribution.

Section VIII concludes.

II. THE ECONOMIC INTUITION OF ENTREPRENEURIAL MIGRATION

There is overwhelming evidence that differences in the attributes of locations affect firm performance (see Glaeser and Goettlieb, 2009, for a review). However, the extent to which firms take advantage of these locational differences, and change their location, is not known. Why would a startup change location? In a static world where the strategies and assets of startups are perfectly matched with their birth location, or where moving a firm after it is founded is very costly, locational changes would be few and idiosyncratic. Yet, there are at least five non-idiosyncratic mechanisms through which founders might choose to change location even after founding the firm—in turn making entrepreneurial migration more common.

The first is migration costs. Even if migration costs are surmountable, they can be meaningful for the firm, the founders, and their families. If a project has initial uncertainty on whether it will be successful, or of there are capital constraints, the firm might prefer to achieve a small level of success before moving to a different location and incurring these costs. The firm will then be born wherever the idea occurred, and migration would follow later.

The second reason is the possibility of arbitrage between different locations. For example, due to differences in cost of inputs, entrepreneurs might optimally choose different regions to develop their initial product and to grow their company. Senor and Singer (2009) document this as a common strategy for Israeli startups: they develop their initial technology in

Israel, and then move to the U.S. to commercialize it, but continue to perform substantial product development in Israel afterwards.

The third reason is the role of uncertainty on the location of a complement—such investors or customers. Early in the process of seeking financing, entrepreneurs might want to locate close to their investors to allow them to add more value (Bernstein et al, 2016), but cannot know where these will be yet. Their optimal strategy could then be to begin in some location to develop a minimum product, and then move close to their investors once financing is secured and the location of their lead investor is known. Similarly, entrepreneurs with new technology might choose to postpone their location choices until they understand better which market their technology will serve and where those customers will be located.

The fourth reason is the possibility of dynamic firm strategies. Dynamic strategic choices could make firms have a different ideal location at different points of their lifetime. Entrepreneurs might choose to locate in one region to maximize their first project, knowing they will later move. This could have been the case with Microsoft. Albuquerque could have been the right choice for the first few years when they developed their first product, while Redmond was a better location later on; Bill Gates and Paul Allen could have moved to Albuquerque knowing this would move again later on[5].

Finally, fifth, dynamically changing economic conditions can change the appeal of locations through time. As some regions and markets grow and others shrink, locations once considered ideal might not be anymore, causing entrepreneurs to update their location choice.

Together, these mechanisms create patterns of entrepreneurial migration. Their distribution, causes, and empirical determinants, are important facts in the process of understanding the geography of high growth entrepreneurship.

III. DATA OVERVIEW

The underlying data for this paper is business registration records and entrepreneurial quality measures across 25 U.S. states, representing 60% of U.S. GDP (the list of states is available in Appendix Table A1). Business registration records are public records created endogenously when a firm is created as a corporation, partnership, or limited liability company

[5] Whether this was actually the reason for Microsoft's location choices is not relevant in this case, but instead the goal is that it is a possibility.

in the Secretary of State (or Commonwealth) office of any U.S. state[6]. The underlying data contains over 20 million firms, representing all business registrants from 1988 to 2012. Other work using this dataset includes Guzman and Stern (2015, 2016a, 2016b, 21017), Fazio, Guzman, Murray and Stern (2016), and Guzman and Kacperczyk (2016). In each of these instances, these papers have used business registration datasets and predictive analytics to create entrepreneurial quality estimates at the firm level, and aggregate measures of quality, to analyze different elements of entrepreneurship. The most complete entrepreneurial quality approach is documented in Guzman and Stern (2016b), I follow this version exactly.

The approach is based on three interrelated insights. First, a practical requirement for any growth-oriented entrepreneur is business registration. These public documents allow us to observe a "population" sample of entrepreneurs observed at a similar (and foundational) stage of the entrepreneurial process. Second, we are able to measure characteristics related to entrepreneurial quality *at or close to the time of registration.* These characteristics include how the firm is organized (e.g., as a corporation, partnership, or LLC, and whether the company is registered in Delaware), how it is named (e.g., whether the owners name the firm eponymously after themselves), and how the idea behind the business is protected (e.g., through an early patent or trademark application). These start-up characteristics may reflect choices by founders who perceive their venture to have high potential. As a result, though observed start-up characteristics are not causal drivers of start-up performance, they may nonetheless represent early-stage "digital signatures" of high-quality ventures. Third, we leverage the fact that, though rare, we observe meaningful growth outcomes for some firms (e.g., those that achieve an IPO or high-value acquisition within six years of founding), and are therefore able to estimate the relationship between these growth outcomes and start-up characteristics.

This mapping allows us to form an estimate of entrepreneurial quality for any business registrant within our sample—the predicted probability of achieving growth given a firm's at-birth characteristics.

Estimating Entrepreneurial Quality and Dataset Overview. Using all business registrations for 25 U.S. states, I split the sample into two random groups. A first group, containing 25% of the data, is used to estimate the entrepreneurial quality predictive model. To do so, I run a logit regression on all available observables using an equity growth outcome (IPO

[6] In fact, it is the act of *registering* a firm itself that legally creates the firm.

or acquisition in 6 years. The result of this regression is provided in Appendix Table A2. Then, I predict a quality score in the complete (100%) dataset. The quality estimate is performed at the moment of birth in its birth state—in the case of migrants, before migration.

The entrepreneurial quality estimate turns out to be highly predictive of realized performance: through a 10-fold out-of-sample testing procedure, we find in Guzman and Stern (2016) find that 71% grow (out of sample) are in the top 5% of the entrepreneurial quality distribution.

Once quality is estimated, I use the 25% sample to also estimate MSA-level statistics of entrepreneurship following the US Census 2013 MSA definitions. The other 75% is used to keep the sample of all Delaware firms, and estimate migration patterns across the United States.

Measuring the Regional Entrepreneurial Ecosystem. To measure the regional entrepreneurial ecosystem, I build on Guzman and Stern (2016a, 2016b), and propose two statistics to characterize the entrepreneurial ecosystem of an MSA. The first measure is *MSA Entrepreneurial Quantity*, which is simply the number of new *local* firms registered in an MSA and year. This represents a quantity measure of entrepreneurial production, and does not take into account the differences in the quality of firms in different regions—which, as document in Guzman and Stern (2016b, 2017), can be substantial.

To also incorporate variations in quality at the regional level, I propose a second measure, *MSA Entrepreneurship Potential* which is the quality-adjusted quantity of firms in a region[7]. Empirically it is measured as product of average entrepreneurial quality for a region and *MSA Entrepreneurial Quantity*. Since quality is the expected probability of growth given at-birth startup characteristics, it also represents the expected number of growth events for a startup cohort given its underlying quality.

Estimating Migration. I estimate migration of all Delaware firms. Delaware firms represents about 5% of all business registrants and are an overarchingly high growth subset[8]. All firms (Delaware or other) need to register with the local Secretary of State in each state in which they open an office, rent real-estate, hire people, or transact with local banks. These registrations include at least the name of the firm, registration date, and address of principal office. Because

[7] This measure is also proposed in Guzman and Stern (2016a, 2016b) as RECPI.

[8] Guzman and Stern document the incidence of achieving a growth outcome to be 42 times higher for Delaware registered firms than local firms.

Delaware corporate law specifically requires companies to name sufficiently different from one another[9], and because companies must register with their true name in each state[10], the matching across registries is very simple and allows high confidence that two Delaware firms under the same name are the same firm.

I operationalize migration through three conditions: (1) The first state in which the firm registers is assumed to be the birth state, and its registration year the date of founding; (2) if a firm then registers in a second state with its principal office in this new state, this is considered a migration as long as the firm lived in the birth state for at least 3 months; (3) the date of registration in the destination state is the migration date.

This limits the analysis to migration of registered firms across states, but with an ability to see the specific destination address (and hence also the destination MSA). In a way, this is useful, as it generally abstracts away from migrations across contiguous MSAs in the same economic area (e.g. from the San Francisco-Oakland-Freemont, CA MSA to the San Jose-Sunnyvale-Santa Clara, CA MSA).

The resulting dataset of migrants is composed of 206,776 Delaware registered firms, registered between 1988 and 2012, of which 7.7% of them migrates (obviously, this migration rate is right-censored for later years). Each observation also includes the firm's quality at birth in its origin region, the birth state, the final MSA location (destination MSA for migrants, birth MSA for non-migrants), it also includes the age at which it migrates (for migrants), and the entrepreneurial quality and quantity.

The Firm-Level and MSA Datasets. From this process, I create two datasets. The first is a firm-level dataset, contains all Delaware firms registered between 1988 and 2012[11], their quality, year of birth, whether they migrate, their migration date, their birth state, and their final MSA, as well as MSA and state measures of the quality and quantity of entrepreneurship.

The second dataset is a panel of the 162 MSAs in my sample, from 1988 to 2011. It includes the ecosystem measures for each MSA, the number of migrant firms into each MSA

---

[9] In fact, this appears to be true for all states, not just Delaware. The logic for it is that (1) similar names would confuse people engaging in transactions with firms and (2) a new firm should not be able to impersonate another firm by virtue of having a similar name.
[10] The only exception to this are cases where this firm name is already taken by a local firm. In this case, I would simply not document this migration as there will not be a Delaware firm of the name of the migrant. This is very rare.
[11] In the 75% sub-sample

(lagged forward by one year), and a series of other ecosystem variables—bohemia, MSA GDP, venture capital, cost of living, and patenting. To focus directly on 'entrepreneurial migrants', only firms that migrate within the first five years of age are included in the migration counts.

Counting Migrants of Different Quality. To account for differences in migrant quality in the MSA panel, I also include variation across the quality of migrants by counting the migrants that are above specific quality thresholds of the entrepreneurial quality distribution of *all* registered firms. While Delaware firms are overarchingly of high quality, they do vary across the quality distribution. By mapping of these firms to the overall firm quality distribution I am able to make statements of the changes in migration rates across the quality levels of all firms.

Therefore, in the MSA panel, I include five different count variables of migration, all estimated in the subsequent year. *Number of Migrants (t+1)* represents the count of migrants that moved to the MSA in the year $t + 1$. The rest of the variables counts firms only if the are above different entrepreneurial quality thresholds. *Number of Migrants in the top 10% (t+1)* is those firms in the top 10% of the entrepreneurial quality distribution. *Number of Migrants in the top 5% (t+1), Number of Migrants in the top 1% (t+1),* and *Number of Migrants in the top 0.5% (t+1)* are the number in the top 5%, 1%, and 0.5%, respectively. Data limitations in the incidence of migration do not allow to go beyond the top 0.5%.

Other Regional Measures. Finally, I complement this dataset with five measures which the economic literature uses as common measures of the vibrancy of a regional entrepreneurial ecosystem. *Bohemian Index* is a measure of "bohemia" developed following precisely the strategy outlined by Florida (2002). Florida proposes that bohemia is an urban element that attracts knowledge workers and idea creation, resulting in economic growth. He uses a location quotient of bohemian occupations using the 5-percent sample of the 1990 U.S. Decennial Census[12], and then normalizes the location quotient to a mean of 0 and standard deviation of 1. I replicate this using the same occupational codes of Florida with the U.S. Census American Community Survey which allows annual estimation of "bohemia" from 2003 to 2011 for most MSAs. *Median Home Value* for each MSA is downloaded directly from the website Zillow (http://www.zillow.com). Zillow is a online real estate database that produces home market-

[12] These occupations are—following 1990 definitions—authors (code 183); designers (182); musicians and composers (186); craft-artists, painters, sculptors, and artist printmakers (188); photographers (189); dancers (193); and artists, performers, and related workers (194).

value estimates for about 100 million homes nationwide (Zillow, 2016). The Zillow Home Value Index (ZHVI) is simply the median home value estimate for a given MSA and year. Housing costs have long been used in regional economics as a proxy of cost of living, and this is the interpretation used in this paper (e.g. Roback, 1982; Hsieh and Moretti, 2015). *Venture Capital $* is a measure of the available venture capital in a region for new-firm financing. It is the amount of money fundraised in that year and MSA by venture capital firms, as reported by Thompson Reuters VentureXpert. *Number of Patents* is a proxy of the flow of ideas (and idea vibrancy) occurring in an MSA, and it is the number of patents granted in the year and MSA, reported by the U.S. Patent and Trademark Office (USPTO) for the years 2000 to 2012[13]. Finally, *MSA GDP* is the economic product of the region, indicating the strength of its economy in dollar terms. This measure is developed by the Bureau of Economic Analysis (BEA) and is used in constant 2009 dollars[14].

IV. DATA DESCRIPTION AND SUMMARY STATISTICS

Table 1 presents the key summary statistics for the for each of the datasets, the firm-level data and the MSA panel.

Panel A of Table 1 is the firm-level data of all Delaware registrants in the sample of states registered in the years of 1988 to 2012. The total number is 206,776 firms. The average entrepreneurial quality of these firms (*Firm Entrepreneurial Quality*) is 0.011, a very high value compared to the population of firms. Figure 1 shows the quality distribution of all Delaware registrants and all firms registered in these states. Given that this is a log-distribution, the differences are substantial[15]. The quality of the destination MSA (*Final MSA Entrepreneurial Quality*) and the birth state (*Birth State Entrepreneurial Quality*) are naturally much lower as they are measured for all business registrants, not just Delaware. Finally, *Months to Migration* has a mean of 45, but the high standard deviation shows that this data is substantially skewed and, given the structure of the data, it is also right-censored.

---

[13] This list is available at https://www.uspto.gov/web/offices/ac/ido/oeip/taf/cls_cbsa/allcbsa_gd.htm .

[14] The BEA calls this measure Gross Metropolitan Product (GMP). In this case, I instead choose to call it MSA GDP, as it is more easily understood.

[15] In Guzman and Stern (2016b) Delaware registrants are found to be 45 times more likely to achieve an equity growth outcome than other firms.

Figure 2 compares the entrepreneurial quality of movers and non-movers in the sample of Delaware firms (once again, in logs). Movers have lower entrepreneurial quality on average that non-movers, with particular differences in the incidence of each group at the low-end of the quality distribution.

Panel B of Table 1 reports summary statistics on the panel of 162 MSAs from 1988 to 2012. To focus on entrepreneurial migration, migrations are only included if they occur for firms within the first five years of firm life. The average number of high growth migrants to an MSA is 3.24, but the data is, once again, very skewed, with a standard deviation of 12.1. *MSA Entrepreneurial Quality* has a mean of .00038[16]. The log-distribution of this measure is shown in Figure 3, highlighting, once again, its skewness. *MSA Entrepreneurial Quantity*, the number of *local* firms in that MSA and year has a mean value of 971, suggesting the level of entrepreneurship is meaningful in these locations. Finally, *MSA Entrepreneurial Potential,* the quality-adjusted quantity of firms, has a mean of 0.4 and is also substantially skewed.

The table then presents the external observables of entrepreneurial ecosystems. *Bohemia* is—by the definition of Florida (2002)—a standardized index with mean zero and standard deviation of 1. The *Median Home Price* for an average MSA and year is $167,488. The average *Number of Patents* is 363, suggesting some level of innovation on average for the destination MSAs. The amount of *Venture Capital $* is, on average 105 million USD, though many of the regions have a value of zero for this measure. Finally, the MSA GDP is, on average, $46 billion USD. Since many of the datasets used for this had limitations in coverage, and I was unable to obtain complete coverage of all the MSAs and years. When all of these values are used together, the final number of observations drops further to 863.

V. THE INCIDENCE, AGE, GEOGRAPHY, AND QUALITY OF MIGRANTS

I now move on to the empirical analysis using the dataset of individual firms to document a few basic facts on migration, including the incidence, the age of migrants, the geographic distribution, and the quality of migrants compared to source and destination regions.

---

[16] Note that the value of *MSA Entrepreneurial Quality* is different in the individual file and the panel file, but this is due to the design of the sample. In first case, it is weighted by the number of firms in each MSA, but this is not so in the second case.

The Incidence of Migration and Age at Migration. I begin by presenting the share of firms that are migrants in my dataset. Because my sample is right-censored in later years, I use the sub-sample of firms registered between the years 1988 and 2000 and look at migration for all firms over their first 15 years.

10.1% of all firms migrate in their first 15 years, with 6.7% of the all firms migrating within the first 5 years, and 4.2% over the first two years. This estimate is economically high, representing a meaningful number of the high growth entrepreneurs in a region. Furthermore, this estimate is most likely a lower bound: I am only able to see migrations from and to the 60% of the US in my dataset, and including the rest would likely raise the incidence of migration.

In Figure 4 I analyze in more detail the age at migration by presenting the distribution of the share of migrations that occurred by quarter of age, for the first 10 years of these firms. Panel A shows all migrant firms within the Delaware sample of study. The highest rate of migration occurs at the firm's second quarter of age—its first quarter at risk of migration. 13% of all migrants move within the second quarter. The share then monotonically decreases over the lifetime of the firm. 7.1% of firms move in the third quarter and 5.3% in the fourth quarter. The first year of the firm accounts for 25% of all migration.

Panels B and C show the distribution for different quality of migrants. Panel B constraints the set of Delaware migrants to be in the top 5% of the entrepreneurial quality distribution of all firms in the economy, and Panel C constraints the ones in the top 1%. For all purposes, there are no apparent differences between each of the distributions except for the fact that Panel C—which has much less firms—is simply a little noisier.

Migration is common, with the bulk of it being entrepreneurial, and the risk of migration monotonically decreasing with age at any level of firm quality.

The Regional Distribution of Migration. Where do these migrants go and how does their quality vary by destination? I provide an overview of this for the largest MSAs in my sample in Figure 5. This figure includes the total number of migrants in the X axis and the average quality of these in the Y axis. It is possible to see differences in the size of the MSA in attracting migrants, with the highest migration rates occurring to the New York-New Jersey-Long Beach, NY-NJ-PA MSA, followed by the Dallas-Forth Worth-Arlington, TX MSA. There is also an overall positive relationship between the average quality of migrants and the number of migrants to a city, suggesting some scale effects with quality.

The average quality of migrants, however, is still heterogeneous, and the MSAs in the Northwest region of the U.S. are important outliers. At the top is the Portland-Vancouver-Beaverton, OR-WA MSA, followed by both the San Francisco-Oakland-Fremont CA, MSA and the San Jose-Sunnyvale-Santa Clara, CA MSA, and the Seattle-Tacoma-Bellevue, CA MSA. Other MSAs continue down the list.

Firm Quality and Likelihood of Migration. In Tables 3 and 4 I report the results of regressions studying the role of the firm's own quality in its likelihood of migration. I run a linear probability model on all firms, with the dependent variable being a binary measure that is equal to 1 if the firm moves (and 0 otherwise) and the independent variable being the log of firm quality. To focus on entrepreneurial migration, I limit the dependent variable to only migrations in the first 2 years of firm life[17]. The coefficient can be interpreted as the change in probability of migration from a change in one log-point of firm quality. Robust standard errors clustered at the final MSA level.

Table 3 compares the firm quality to the quality of other firms in its birth region. Column 1 is a baseline estimate that uses only *Firm Entrepreneurial Quality* as the independent variable. The point estimate is negative at -.003, though noisy and not statistically significant. Column 2 uses only *Birth State Entrepreneurial Quality* as an independent variable. The coefficient is again a negative point estimate (-.0097). While not statistically significant, the point estimate is counterintuitive: we would expect firms in regions of lower quality to be *more* likely to leave their region.

There is, of course, a problem with the correlation between these two observables: firms of higher quality are more likely to be from higher quality regions. Therefore, what we would hope to do is study the likelihood of migration with firm quality conditional on differences in the quality of regions. I get closer to this relationship in Columns 3 and 4.

In Column 3, I introduce both measures at the same time in the regression. The coefficients are similar. Both are still negative in magnitude and not statistically significant.

In Column 4 I use a different approach: I seek to absorb all possible regional differences by including a fixed-effect for each region-year pair in the sample. The result of this is that differences in the *measured* average firm quality by birth state and year are removed completely, leaving only differences in firm quality within state in the regression. The coefficient of *Firm*

[17] No results change if I focus on the first 5 years instead.

Entrepreneurial Quality is now positive and statistically significant. The magnitude is 0.0024. This variable has a 10-90 range (difference between the 10th and the 90th percentile) of 4.6 and a 25-75 range of 2.1. Which would imply that *within high growth entrepreneurs,* moving from the 10th percentile of quality to the 90th percentile would increase the likelihood of migration by 1.1 percentage points, or 26% of the rate of migration (in the first two years).

This difference helps explain the puzzling results from Column 2: firms of higher quality are more likely to migrate, but they are also more likely to be born in states of higher quality.

In Table 4 I perform a similar exercise but compare the firm to the quality of firms in destination MSAs, rather than firms in state of birth. The interpretation of the coefficients is different than Table 3. Because these are destinations, the coefficients do not tell us whether a firm is more or less likely to move (as in Table 3), but how do migrant firms compare to local firms in the destination.

Column 1 is again the baseline relationship of firm quality and migration rates, with the same coefficient as in Table 3.

Column 2 compares the likelihood of being a migrant given *Final MSA Entrepreneurial Quality*. Surprisingly, this coefficient is negative—with a magnitude of –.0059 —and statistically significant: out of all high growth firms, the share of firms who are migrants is *lower* in higher quality ecosystems. This is striking when put together with Figure 4 and Table 3, which show that firms appear to be consistently moving to high quality ecosystems. The result suggests that the *local* production of entrepreneurs in those regions increases even faster than migration, causing that the overall share of migrants drops in higher quality ecosystems.

Column 3 includes both measures at the same time. Finding no significance in either coefficient. In Column 4, I repeat the exercise of Table 3 and include a fixed-effect for each MSA-year pair in the data. The coefficient is basically zero (–.0002), even if the confidence interval is very large. This suggests that there are very little systematic differences in the quality of migrants compared to the quality of locals in an MSA, though large idiosyncratic differences.

VI. THE RELATIONSHIP OF MSA CHARACTERISTICS TO MIGRATION

In this section, I use the panel of MSAs to study the relationship between MSA characteristics and migration. I study two types of relationships: the relationship between some common measures of entrepreneurial ecosystems—*Bohemia, MSA GDP, Median Home Value,* *Venture Capital $*, and *Number of Patents*—and migration rates; and the relationship between *MSA Entrepreneurial Quantity* and *MSA Entrepreneurial Potential* and migration rates. Because migration indicates the revealed preference of firms for regions at different points in time, the determinants of migration rates might shed some light on the importance (or attractiveness) of these characteristics to entrepreneurs.

To study this relationship, I run count data regressions through a Poisson Quasi-Maximum Likelihood Estimator (QMLE). The dependent variable is number of entrepreneurial migrants[18] who arrive to the MSA in year $t + 1$, and the independent variables are each of the MSA characteristics in year t. I include all of the independent variables in their natural log to account for their skewness, except for *Bohemia* which is already standardized. In the case of *Venture Capital $*, which contains a value of zero for many regions, I use the log plus 1 instead.

Common Measures of Entrepreneurial Ecosystems and Migration Rates. Table 5 reports the results of pooled regression of migration rates on common measures of entrepreneurial ecosystems. All regressions include year fixed-effects, and standard errors are clustered at the MSA level.

In Columns 1 through 5, I include each measure independently. Each of the coefficients is positive and significant, suggesting a positive empirical relationship between migration rates and the level of bohemia in a region, its GDP, its cost of living, its supply of venture capital, and its level of innovation. Each of these positive relationships would be theoretically expected, except for cost of living, where we would have expected the coefficient to be *negative* instead (lower cost of living cities should be more attractive to entrepreneurs, not less).

Of course, these measures are also highly correlated amongst themselves, and it would be important to try to differentiate them from each other. In Column 6 I make some progress by including all the observables at the same time. Most coefficients decrease substantially in magnitude, and the sign and significance of coefficient make more sense. *Bohemian Index*, *Venture Capital* and *Number of Patents* are all positive and significant, in agreement with general agglomeration theories that would suggest each of these would matter for regional

[18] I consider 'entrepreneurial migrants' as those firms migrating within the first five years after founding. In unreported regressions, I run all analysis using only firms migrating in the first two years. All results are qualitatively the same.

attractiveness. *MSA GDP* is not significant. The coefficient for *Median Home Value* (cost of living) is now negative and significant.

Including MSA Fixed-Effects. The analysis in Table 5 centered on the "between variation" of MSAs (the variation that drives the positive relationships found is between the regions themselves). It might be of more interest to study the "within" variation: Do increases in some MSA characteristic correlate to increases in migration to that MSA? To do so, I add MSA fixed-effects to the regressions, which absorb all permanent differences between the MSAs in the time period, leaving only the within MSA variation to drive the results.

However—because MSA characteristics are serially correlated—adding fixed-effects could take away a large portion of the variation (for example, the most bohemian regions are likely to be the same ones, such as New York, or San Francisco, over many years). Therefore, to run fixed-effects regressions, it is first necessary to establish whether there is enough remaining variation after including fixed effects for the regression.

I use a simple test: I estimate the OLS residual of each measure after MSA and year fixed effects (this will be the variation left for the fixed-effect regressions) and compare this to the original variation, to see if there is still significant variation left.

In Table 6, I report the mean, standard deviation, and range (max – min) for the original values, in Panel A, and the residuals, in Panel B. The extent of 'remaining' variation is measured in Panel C, as two statistics: the first column of Panel C estimates the ratio between the residual standard deviation, and the original standard deviation; the second column estimates the ratio between the range of the residuals and the range of the original values.

There appears to be very little variation left for *Bohemia* and *MSA GDP* after including fixed-effects. The ratio of the standard deviations is less than 0.1, and the ratio of the range is less than 0.15, for each. Therefore, I remove these two measures from subsequent analysis.

There is more—but still not too much—variation left for *Median Home Value* and *Number of Patents.* The ratios of their standard deviations is .25 and .14, respectively, and the ratio of their range is .31 and .28. Though the choice of the exact threshold of variation is certainly ad-hoc, I decide to keep these variables in the subsequent models and assume that this variation is enough to be included.

The remaining variables, *Venture Capital*, *MSA Entrepreneurial Quantity*, and *MSA Entrepreneurial Potential*, all have higher variation and are included.

Fixed-Effects Results. With this analysis in hand, I present, in Table 7, the same regressions of Table 5 after including MSA fixed-effects for the variables with enough variation.

In Columns 1 through 3 I include each of these measures independently. Interestingly, all coefficients are now not significant and substantially decreased in magnitude. Once fixed-effects are included, *Median Home Value,* and *Number of Patents,* at the MSA level all stop holding a relationship to migration rates. *Venture Capital $* is still positive, and significant, though the effect is relatively small (0.021). The interpretation of this coefficient is as follows: an increase of 1 log-point in venture capital fundraised in an MSA correlates to an increase in migration counts by 2.1% in the subsequent year to that MSA.

In Column 4 I include these three measures together. *Venture Capital $* is again positive and significant, with a higher coefficient of (0.039), while the other two continue to be non-significant. Notably, the coefficients should not be compared to each other, as there are some sample changes between each regression due to differences in coverage of each observable. In an unreported regression that uses the same sample as Column 4 but includes only *Venture Capital* as an independent variable, the coefficient is 0.037, with a standard error of .008.

Migration and MSA Entrepreneurial Potential. I continue this analysis in Table 8 by including two new measures on the entrepreneurial ecosystem of a region, based on estimates of entrepreneurial quality and business registration records. Specifically, I include *MSA Entrepreneurial Quantity*, the number of firms born in a MSA and region, and *MSA Entrepreneurial Potential*, the quality-adjusted quantity of firms born in a MSA and region, as independent variables (in log form), and run Poisson QMLE regressions with year fixed effects and number of migrants in $t+1$ as the dependent variable.

Columns 1 and 2 report pooled regressions that study the 'between variation' of these measures and migration rates. Column 1 shows a positive and statistically significant relationship between entrepreneurial quantity and rates of migration to that region. Column 2 shows a positive and significant relationship between entrepreneurial potential and rates of migration to that region.

In Columns 3 through 7 I move to the within variation by including MSA fixed-effects.

Column 3 reports the relationship between *MSA Entrepreneurial Quantity* and number of migrants. Though the magnitude is positive, the result is not significant anymore, and the confidence interval is quite large.

In Column 4 I show the relationship between *MSA Entrepreneurial Potential* and migration rates with MSA fixed-effects. The relationship is positive, significant above the 5% level, and economically meaningful with a magnitude of 0.40. The economic interpretation is that an increase of 1 log-point in aggregate entrepreneurial potential of *local* firms in a region relates to an increase of 49% of the number of migrants to this region in the subsequent year ($e^{0.40} = 1.49$).

In Columns 5 through 7 I perform some robustness tests on these relationships by including other controls, particularly focusing on *Venture Capital*, which had a positive relationship in Table 7.

In Column 5 I report a regression that includes *MSA Entrepreneurial Quantity* and *Venture Capital* together, as well as MSA fixed-effects. The coefficient of *MSA Entrepreneurial Quantity* is basically unchanged and still not significant, while that of *Venture Capital* is positive, significant, and similar in magnitude—once the quantity of new firms is controlled for, the amount of venture capital continues relate to higher rates of migration to that MSA.

In Column 6 I report a similar regression but include *MSA Entrepreneurial Potential* rather than *MSA Entrepreneurial Quantity*, and include again venture capital and MSA fixed effects. The effects here differ. The coefficient for entrepreneurial potential is (basically) unchanged from the one in Column 4, but the coefficient of *Venture Capital $* drops (slightly) and is now not statistically significant (p-value of 0.15). In Column 7, I repeat this regression but also include all other observables (bohemia, GDP, cost of living and patenting) in the regression. Once again, the coefficient for entrepreneurial potential is basically the same, while the venture capital coefficient drops substantially and is basically zero.

VII. THE EFFECT OF ENTREPRENEURIAL POTENTIAL ACROSS THE QUALITY OF MIGRANTS

The relationships in the prior section are compelling. This section pushes further into understanding the role of *MSA Entrepreneurial Potential* on entrepreneurial migration by looking at counts of migrants above different thresholds of the entrepreneurial quality distribution of migrants. While the sample used to measure migration—Delaware registrants—is overarchingly of high quality, those firms do vary across the overall entrepreneurial quality distribution. In the data section, I introduced four measures that count firms only in the top 10%, 5%, 1%, and 0.5% of the distribution of *all* firms. Table 9 reports regressions of *MSA Entrepreneurial Potential* on migration counts at different thresholds of quality. The regressions include MSA and year fixed effects, with robust standard errors clustered at the MSA level.

Panel A reports the results of Poisson QMLE regressions. Column 1 is the main coefficient—the same regression as Table 8, Column 4. Columns 2, 3, 4, and 5, change the outcome variable to be the count of only those firms in the top 10%, 5%, 1%, or 0.5% of entrepreneurial quality of all firms, respectively. All coefficients are overarchingly similar, between 0.350 and 0.385, and are not statistically different from each other.

While the coefficients do not change much in Panel A, the number of observations drops as I move up in the quality threshold—some regions are excluded due to the MSA fixed effects. To look at differences in the complete sample, I present in Panel B OLS regressions with the log-number of migrants plus 1 as the dependent variable. All coefficients continue to be positive and significant, though the magnitude of the coefficient does drop as we move up the threshold.

VIII. CONCLUSION

As we conclude this analysis, it is worth pausing and asking *what exactly* is learned from these results. This paper does not seek to establish definitive causality in the regressions; however, the relationships found are more than simply correlations of an interesting phenomenon, and suggest underlying economic dynamics. Besides meaningful summary statistics documenting the phenomena of entrepreneurial migration, the approach of this paper offers three unique empirical benefits relative to prior work: it includes all firms born in a region while carefully controlling for the quality of these firms; it includes unique measures of the entrepreneurial ecosystem of regions that directly measures the local supply of new firms in those regions; and it is able to use migrants to show the revealed preference of firms for certain regions.

The summary statistics are interesting, and highlight the importance of entrepreneurial migration as a phenomenon in regional dynamics. It is interesting that migrants of higher quality are more likely to migrate, causing higher quality regions to be more likely to have migrants leave the region even though the risk of leaving drops with birth region quality. The reasons for this could be multiple and it appears an interesting fact on which future research could be developed.

At the MSA-level. This paper shows the significance of regional fixed effects in determining the location choice of firms. This result casts doubt on policies that seek to simply imitate the observable characteristics of entrepreneurial clusters (such as the high level of patenting or venture capital financing in Silicon Valley) since a substantial portion of migration is driven by the unobservable fixed effects.

This paper also shows little relationship between the migration rates of high growth entrepreneurs and either the average housing costs or the local level of patenting. The housing costs result should give pause to researchers and policymakers who have focused on affordability as an approach to develop entrepreneurial clusters, often focusing directly on the role of low cost of living in attracting migrant entrepreneurs. The patenting result suggests differences between the attributes migrants seek in a region and the determinants of local entrepreneurship production. It is reasonable to expect local patenting to relate to local startups, while also being potentially true that startups do not move to regions seeking knowledge externalities at the level of new ideas.

This paper shows a positive relationship between venture capital and migration rates, leaving open the possibility that available venture capital financing *causally* attracts firms to regions. The coefficient, however, is rather small. It is possible that better measures of venture capital increase this coefficient. Much is left to be understood on the relationship between regional VC availability and firm migration, or the value it plays for firms directly.

As its main result, this paper shows a very strong and positive relationship between the entrepreneurial potential of a region and migration. A higher local production of entrepreneurs correlates to attracting more migrant entrepreneurs, conditional on these local entrepreneurs being of high quality. This result is novel in two ways. First, it is the first to document the role of the local entrepreneurial ecosystem through a systematic measure of this ecosystem directly. In agreement with a much discussed but seldom documented principle in regional entrepreneurship: simply creating new firms does not appear to make the region more attractive, it is necessary that these firms are of high quality. And second, it shows some dynamic of increasing returns in regional entrepreneurship, where the production of a strong ecosystem also attracts startups from other locations. This is in line with accounts, for example, of many high potential founders who continue to move to Silicon Valley. The impact of these increasing returns should make policymakers and researchers re-consider the ways in which entrepreneurship can lead to regional development—it increases the importance of the role of 'kickstarting' an ecosystem, and highlights the risk of a 'spiraling descent' in regional entrepreneurship, where the few good firms in low-quality ecosystems are more likely to leave, in turn making it even lower quality.

The ultimate question is, of course, what are the elements of a regional ecosystem that influence firm productivity and drive economic growth. This paper offers only slight progress on this question. Nonetheless, the evidence is informative and, while much is left, some progress has been made. Future research—at both the empirical and theoretical levels—can build on these contributions as it seeks to advance our understanding of *what is it* that makes entrepreneurship create economic growth, and how to influence those parameters for the benefit of regions, firms, and welfare at large.

BIBLIOGRAPHY

Arzaghi, Mohammad, and J. Vernon Henderson. 2008. "Networking off Madison Avenue." *Review of Economic Studies* 75(4): 1011–38.

Bernstein, Shai, Xavier Giroud, Richard R. Townsend. 2016. *"The Impact of Venture Capital Monitoring"* Journal of Finance. *71(4): 1591-1622*

Dahl, Michael S, and Olav Sorenson. 2012. "Home Sweet Home : Entrepreneurs ' Location Choices and the Performance of Their Ventures." *Management Science* Articles i(6): 1–13.

Delgado, M., M. E. Porter, and S. Stern. 2010. "Clusters and Entrepreneurship." *Journal of Economic Geography* 10(4): 495–518.

Delgado, Mercedes, Michael E. Porter, and Scott Stern. 2014. "Clusters, Convergence, and Economic Performance." *Research Policy* 43(10): 1785–99.

Diamond, By Rebecca. 2016. "The Determinants and Welfare Implications of US Workers ' Diverging Location Choices by Skill : 1980 – 2000 †." *The American Economic Review* 106(3): 479–524.

Ellison, Glenn, Edward L Glaeser, and William R Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review* 100(June): 1195–1213. http://www.nber.org/papers/w13068.

Fazio, Catherine, Jorge Guzman, Fiona Murray, and Scott Stern. "A New View of the Skew: Quantitative Estimates of the Quality and Quantity of Entrepreneurship." *MIT Innovation Initiative. Policy Brief Series.*

Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2016. "Agglomeration of Invention in the Bay Area: Not Just ICT." *American Economic Review* 106(5): 146–51.

Gans, Joshua, and Scott Stern. 2016. *Entrepreneurial Strategy*. ed. Mimeo.

Gans, Joshua S, Scott Stern, and Jane Wu. 2016. *Foundations of Entrepreneurial Strategy*. SSRN Working Paper Series. # 2844843

Glaeser, Edward L., and Joshua D. Gottlieb. 2009. "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." *Journal of Economic Literature* 47(4): 983–1028.

Glaeser, EL. 2007. "Entrepreneurship and the City." http://www.nber.org/papers/w13551 (July 9, 2013).

Glaeser, Edward L, Sari Pekkala Kerr, and William R Kerr. 2014. "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines." *Review of Economics and Statistics* 97(2): 498–520.

Gonzalez-Uribe, Juanita, and Michael Leatherbee. 2016. "The Effects of Business Accelerators on Venture Performance: Evidence from Start-up Chile". Mimeo.

Guzman, Jorge, and Aleksandra Kacperczyk. 2016. "Gender Gap in Entrepreneurship." *MIT Working Paper.*

Guzman, Jorge, and Scott Stern. 2015. "Where Is Silicon Valley ?" *Science* 347(6222): 606–9.

Guzman, Jorge, and Scott Stern. 2016. "Nowcasting and Placecasting Entrepreneurial Quality and Performance." *NBER/CRIW Measuring Entrepreneurial Businesses: Current Knowledge and Challenges conference* (December): 67.

Guzman, Jorge, and Scott Stern. 2016. "The State of American Entrepreneurship: New Estimates of The Quantity and Quality of Entrepreneurship for 15 US States, 1988-2014." *NBER Working Paper # 22095*

Guzman, Jorge, and Scott Stern. 2017. "The Entrepreneurial Quality and Quantity of US Regions". Mimeo.

Hsieh, Chang-Tai, and Enrico Moretti. 2015. "Why Do Cities Matter? Local Growth and Aggregate Growth." *National Bureau of Economic Research Working Paper* 21154: 40. http://www.nber.org/papers/w21154.

Florida, Richard. 2002. "Bohemia and Economic Geography." *Journal of Economic Geography* 2(1): 55–71.

Kerr, William R., and Ramana Nanda. 2009. "Democratizing Entry: Banking Deregulations, Financing Constraints, and Entrepreneurship." *Journal of Financial Economics* 94(1): 124–49.

Krugman, P. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99: 483–99.

Marshall, Alfred. 1890. *Principles of Economics.* Macmillan. London.

Michelacci, Claudio, and Olmo Silva. 2007. "Why So Many Local Entrepreneurs?" *The Review of Economics and Statistics* 89(November): 615–33.

Porter, Michael. 2003. "The Economic Performance of Regions." *Regional Studies* 37(6–7): 545–46.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(6): 1257–78.

Samila, Sampsa, and Olav Sorenson. 2011. "Venture Capital, Entrepreneurship, and Economic Growth." *The Review of Economics and Statistics* 93(February): 338–49. http://www.mitpressjournals.org/doi/abs/10.1162/REST_a_00066 (July 7, 2014).

Duranton, Gilles. 2011. "California Dreamin': the Feeble Case for Cluster Policies." *Review of Economic Analysis* 3: 3–45.

Senor, Dan, and Saul Singer. 2011. *Start-up Nation: The Story of Israel's Economic Miracle*. Twelve.

Zillow. 2016. What's the Zillow Home Value Index? Retrieved from http://www.zillow.com/wikipages/What's-the-Zillow-Home-Value-Index/ Accessed on December 29, 2016.

TABLE 1

Summary Statistics

A. File of all Delaware registrants in 28 US states representing 70% of US GDP. 1988-2012

	Description	**Obs**	**Mean**	**Std. Dev.**
Registration Year	Year of birth	206776	2003.066	6.225
Firm Entrepreneurial Quality	The firm's quality at birth in its original state	206776	0.011	0.028
Final MSA Entrepreneurial Quality	Final MSA is destination for migrants, place of birth for locals	206776	0.00073	0.00080
Birth State Entrepreneurial Quality	Entrepreneurial quality of the state of birth.	206776	0.00058	0.00047
Months to move	Months from birth to migration date.	16024	45.97	49.06
Moves in 2 Years	Only for firms that move in 2 years or never move	198197	0.04	0.19

B. Panel of 162 MSAs with migration counts, 1988-2012.

	Description	**Obs**	**Mean**	**Std. Dev.**
Year	Year of observation	4025	2000.00	7.21
Number of Migrants (t+1)	Total number of Delaware movers the following year	4025	2.91	10.99
Number of Migrants in top 10% (t+1)	Movers in top 10% of the quality of *all* firms	4025	2.37	9.31
Number of Migrants in top 5% (t+1)	Movers in top 5% of the quality of *all* firms	4025	2.17	8.43
Number of Migrants in top 1% (t+1)	Movers in top 1% of the quality of *all* firms	4025	1.04	4.00
Number of Migrants in top 0.5% (t+1)	Movers in top 0.5% of the quality of *all* firms	4025	0.56	2.25
MSA Entrepreneurial Quality	Average quality of firms *born* in this MSA and year	4025	0.00038	0.00057
MSA Entrepreneurial Quantity	Total count of firms *born* in this MSA and year	4025	977.07	3111.69
MSA Entrepreneurial Potential	Quality-adjusted quantity of firms *born* in this MSA and year. Empirically estimated as quality times quantity.	4025	0.41	1.45
Bohemia	Replica of Florida's (2003) "Bohemia" index using the American Community Survey. Only available for the 122 largest MSAs from 2005-2012.	963	-0.04	0.20
Median Home Price	Source: Zillow Home Index Value. Available from 1995-2012 for 130 MSA, with a few exceptions / gaps.	2177	167720.70	101384.10
Number of Patents	Granted in that year and MSA. Reported by the USPTO.	2067	365.39	1010.25
Venture Capital $	Dollars fundraised in that year. Source: ThompsonReuters.	4025	106.61	882.44
MSA GDP	Source: Bureau of Labor Statistics. BLS begins reporting MSA GDP after 2001.	1871	47107.23	118310.60

TABLE 2

Migration Rates for All Firms in Sample born Before 2001*

Does not migrate	89.9%
Entrepreneurial Migration	
Migrates before 2 years of age	4.2%
Migrates before 5 years of age	6.7%
All Migration	
Migrates from age 6 to 15	10.1%

*Migration rates estimated only for firms born before 2001 to allow at least 15 years for firms to migrate.

TABLE 3

Migrant Firm Quality Controlling for Source Region Differences
Linear Probability Model with Binary Outcome
Dependent Variable is Binary Measure Equal to 1 if Firm Migrates in First Two Year

	(1)	(2)	(3)	(4)
Ln(Firm Entrepreneurial Quality)	-0.00310		-0.000477	0.00237*
	(0.00439)		(0.00468)	(0.00127)
Ln(Birth State Entrepreneurial Quality)		-0.00970	-0.00922	
		(0.00742)	(0.00868)	
Birth State -Year Pair Fixed Effects	No	No	No	Yes
N	197723	197723	197723	197723

Sample is all Delaware registrants founded between 1988 and 2011 in 25 US states representing 60% of US by GDP. Outcome variable is equal to 1 if firm migrates in first two years, which account for about two-thirds of all entrepreneurial migration (see Table 2). Dependent variables are in log form to control for substantial skewness in their distribution (documented in Figures 1 and 2). Firm Entrepreneurial Quality is the entrepreneurial quality estimate when a firm is born in its birth location. Birth State Entrepreneurial Quality is the average quality of firms born in that state and year. Column 4 absorbs all difference in average quality if firms in a year and state by including fixed effects for each state-year pair. Robust standard errors are clustered at the level of the firm location MSA. * $p < .1$ ** $p < .05$

TABLE 4

Migrant Firm Quality Controlling for Destination Region Differences
Linear Probability Model with Binary Outcome
Dependent Variable is Binary Measure Equal to 1 if Firm Migrates in First Two Year

	(1)	(2)	(3)	(4)
Ln(Firm Entrepreneurial Quality)	-0.00310		-0.00184	-0.000203
	(0.00439)		(0.00538)	(0.00771)
Ln(Final MSA Entrepreneurial Quality)		-0.00593**	-0.00443	
		(0.00269)	(0.00514)	
Final MSA - Year Pair Fixed Effects	No	No	No	Yes
N	197723	197723	197723	197723

Sample is all Delaware registrants founded between 1988 and 2011 in 25 US states representing 60% of US by GDP. Outcome variable is equal to 1 if firm migrates in first two years, which account for about two-thirds of all entrepreneurial migration (see Table 2). Dependent variables are in log form to control for substantial skewness in their distribution (documented in Figures 1 and 2). Firm Quality is the entrepreneurial quality estimate when a firm is born in its birth location. Final MSA Entrepreneurial Quality is the average quality of firms in the final MSA (destination for migrants) and year. Column 4 absorbs all difference in average quality of firms in a cohort year and MSA by including fixed effects for each MSA-year pair. Robust standard errors are clustered at the level of the firm location MSA. * $p < .1$ ** $p < .05$

TABLE 5

Common Measures of Strong Entrepreneurial Clusters in Literature
Quasi-Maximum Likelihood Poisson
Dependent Variable: Number of Migrant Firms to MSA in t+1
Strongly balanced panel of 162 MSAs 1988-2012

	(1)	(2)	(3)	(4)	(5)	(6)
Bohemian Index	2.014**					0.690**
	(0.124)					(0.252)
Ln(MSA GDP)		1.080**				0.408
		(0.0800)				(0.258)
Ln(Median Home Value)			1.023**			-0.906**
			(0.499)			(0.366)
Ln(Venture Capital $+1)				0.413**		0.0935**
				(0.0569)		(0.0370)
Ln(Number of Patents)					0.806**	0.379**
					(0.122)	(0.141)
N	963	1871	2177	4025	2067	855
Log-Likelihood	-5013.3	-5660.7	-17191.8	-15640.8	-7762.4	-3054.6

This regression represents the number of firms founded in the same year that will migrate to the region within the next 5 years. Standard errors clustered at the MSA level and year fixed-effects included in all regressions. MSA is a metropolitan statistical area (MSA) using 2013 US Census MSA Definitions. Bohemian Index is estimated by replicating Florida's (2003) location quotient of occupations with the yearly U.S. Census American Community Survey. MSA GDP as reported by the Bureau of Economic Analysis. Real Estate costs is the median home value provided in the Zillow ZHVI data series. Venture Capital $ represents the amount of money fundraised that year in that city, provided by Thompson Reuters VentureXpert. Patenting is the number of patents granted by MSA as reported in the Patents by MSA Table by the US Patent and Trademark Office. * p < .1 ** p < .05

TABLE 6

Variation in Variables with and without Including MSA and Year Fixed Effects

		A. Summary Statistics of Original Value			B. Summary Statistics of Residual After Fixed Effects		C. Ratio of values from A over B	
	N	μ_o	σ_o	Range	σ_r	Range	Ratio σ_r/σ_o	Ratio Range
Bohemia	971	-0.04	0.20	2.08	0.01	0.21	0.06	0.10
Ln(MSA GDP)	1884	9.73	1.27	6.50	0.06	0.90	0.05	0.14
Ln(Median Home Value)	2194	11.90	0.48	2.68	0.12	0.83	0.25	0.31
Ln(Venture Capital $ +1)	4050	0.62	1.75	10.12	0.90	10.57	0.51	1.04
Ln(Number of Patents)	2080	4.15	1.81	9.35	0.26	2.60	0.14	0.28
Ln(MSA Entrepreneurial Potential)	4050	-1.67	1.95	14.17	0.61	11.07	0.31	0.78
Ln(MSA Entrepreneurial Quantity)	4050	6.72	1.73	12.01	0.38	7.89	0.22	0.66

This table measures the residual variation with and without fixed effects in a panel of 162 MSAs from 1988 to 2012. Panel A is the unconditional variation on the data. Panel B is the variation conditional on MSA and year fixed effects. Panel C is the ratio of these two values, a measure of the 'residual' variation. *Bohemia* and *Ln(MSA GDP)* are found to have to low variation left after including fixed effects to do analysis, and hence are excluded from the fixed-effects regressions. μ represents the mean, σ represents the standard deviation.

TABLE 7

Common Measures of Strong Entrepreneurial Clusters in Literature
Panel of 162 MSAs, 1988-2012
Poisson QMLE with Year and MSA Fixed Effects.
Dependent Variable: Number of Migrants to MSA in t+1

	(1)	(2)	(3)	(4)
Ln(Median Home Value)	0.0883			-0.175
	(0.287)			(0.354)
Ln(Venture Capital $+1)		0.0214*		0.0392**
		(0.0122)		(0.0162)
Ln(Number of Patents)			-0.00439	-0.0549
			(0.226)	(0.225)
N	2017	3850	1846	1526
Log-Likelihood	-2777.2	-4426.3	-2328.1	-2046.6

This regression represents the number of firms founded in the same year that will migrate to the region within the next 5 years. Standard errors clustered at the MSA level and year fixed-effects included in all regressions. MSA is a metropolitan statistical area (MSA) using 2013 US Census MSA Definitions. Bohemian Index is estimated by replicating Florida's (2003) location quotient of occupations with the yearly U.S. Census American Community Survey. MSA GDP as reported by the Bureau of Economic Analysis. Real Estate costs is the median home value provided in the Zillow ZHVI data series. Venture Capital $ represents the amount of money fundraised that year in that city, provided by Thompson Reuters VentureXpert. Patenting is the number of patents granted by MSA as reported in the Patents by MSA Table by the US Patent and Trademark Office. * p < .1 ** p < .05

TABLE 8

The Effect of Quantity and Potential on Migration Rates

QMLE Poisson Model.
Dependent Variable: Number of Migrant Firms to MSA in t+1
Strongly Balanced Panel of 162 MSAs 1988-2012

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln(MSA Entrepreneurial Quantity)	0.950**		0.373		0.379		
	(0.0789)		(0.271)		(0.267)		
Ln(MSA Entrepreneurial Potential)		0.750**		0.403**		0.398**	0.353**
		(0.0877)		(0.107)		(0.106)	(0.162)
Ln(Venture Capital $ +1)					0.0231*	0.0177	-0.00379
					(0.0124)	(0.0116)	(0.0232)
MSA Fixed Effects	No	No	Yes	Yes	Yes	Yes	Yes
Common Controls	No	No	No	No	No	No	Yes
N	4025	4025	3850	3850	3850	3850	730
Log-Likelihood	-9800.6	-10725.2	-4400.1	-4356.7	-4393.1	-4352.6	-961.0

This regression represents differences in the migration counts of entrepreneurial firms to MSAs. All measures are included in logs due to substantial skewness. *MSA Entrepreneurial Quantity* represents the number of firms founded in the MSA and year. *MSA Entrepreneurial Potential* represents the quality-adjusted quantity of firms founded in the MSA. *Venture Capital $* is the amount of VC fundraised in that location, from Thompson Reuters. Common controls are all covariates of Table 5. Year fixed effects included in all regressions. Robust standard errors clustered at the MSA level are reported in parenthesis. * p < .1 , ** p < .05.

TABLE 9

Elasticity on Migration Rates by Quality of Migrants.
Fixed Effects Model on Strongly Balanced Panel of 162 MSAs, 1988-2012.
OLS and Poisson QMLE Regressions including MSA and year fixed-effects.

A. Poisson QMLE Fixed Effect Regression. Dependent Variable Number of Migrants

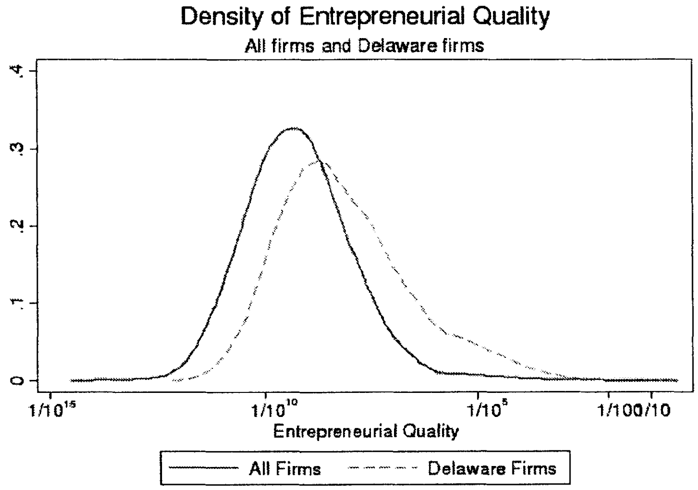
	All Migrants	Migrants in top 10% of quality	Migrants in top 5% of quality	Migrants in top 1% of quality	Migrants in top 0.5% of quality
	(1)	(2)	(3)	(4)	(5)
Ln(MSA Entrepreneurial Potential)	0.403**	0.383**	0.385**	0.350**	0.382**
	(0.107)	(0.0991)	(0.0935)	(0.0807)	(0.102)
N	3850	3600	3550	3100	2625
Log Likelihood	-4356.7	-3795.0	-3622.3	-2385.2	-1637.7

B. OLS Fixed Effect Regression. Dependent Variable Ln(Number of Migrants + 1)

	All Migrants	Migrants in top 10% of quality	Migrants in top 5% of quality	Migrants in top 1% of quality	Migrants in top 0.5% of quality
	(1)	(2)	(3)	(4)	(5)
Ln(MSA Entrepreneurial Potential)	0.0408**	0.0360**	0.0357**	0.0258**	0.0180**
	(0.0179)	(0.0159)	(0.0154)	(0.0101)	(0.00798)
N	4025	4025	4025	4025	4025

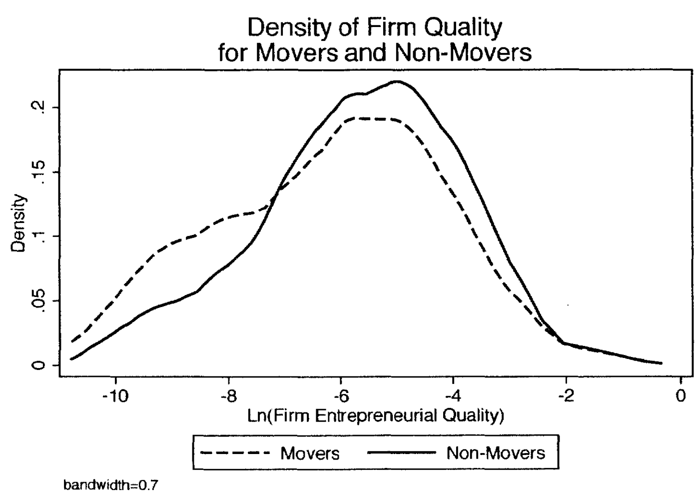
This regression replicates model 4 of Table 8 estimating the role of local entrepreneurial potential on rates of migration to an MSA, after including MSA and year fixed effects. Robust standard errors are clustered at the MSA level. * p < .1 , ** p < .05

FIGURE 1



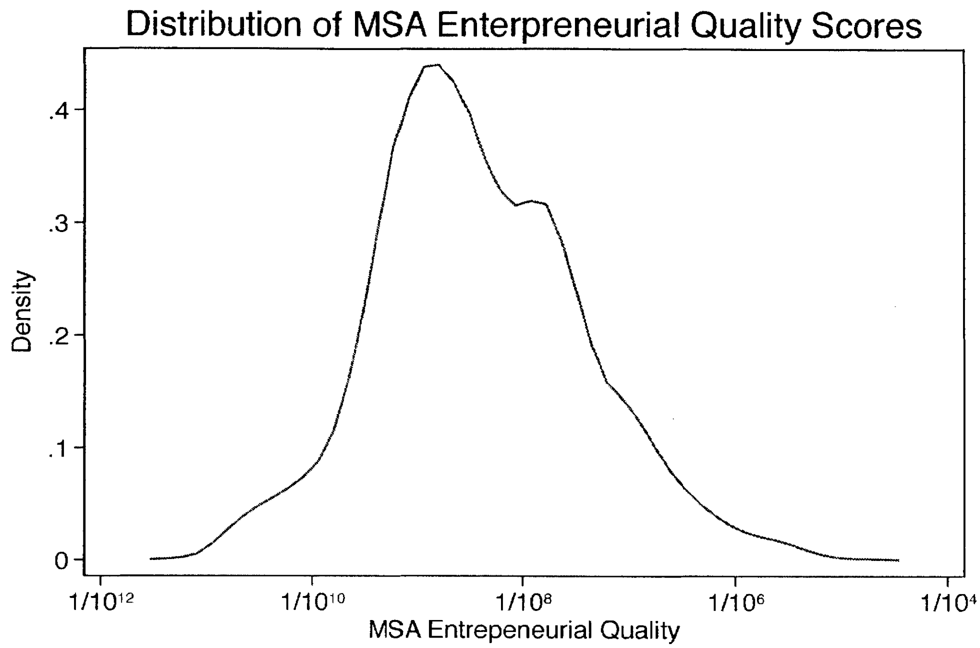
Notes: This figure represents the Entrepreneurial Quality distribution (estimated from the model in Guzman and Stern, 2016b) of all firms in the sample a Delaware firms in the sample. It is easy to see the quality of Delaware firms is higher than all firms and the incidence is much higher at a higher level of quality. It is also possible to see the large skewness in the measure, suggesting the use of the natural log as appropriate.

FIGURE 2



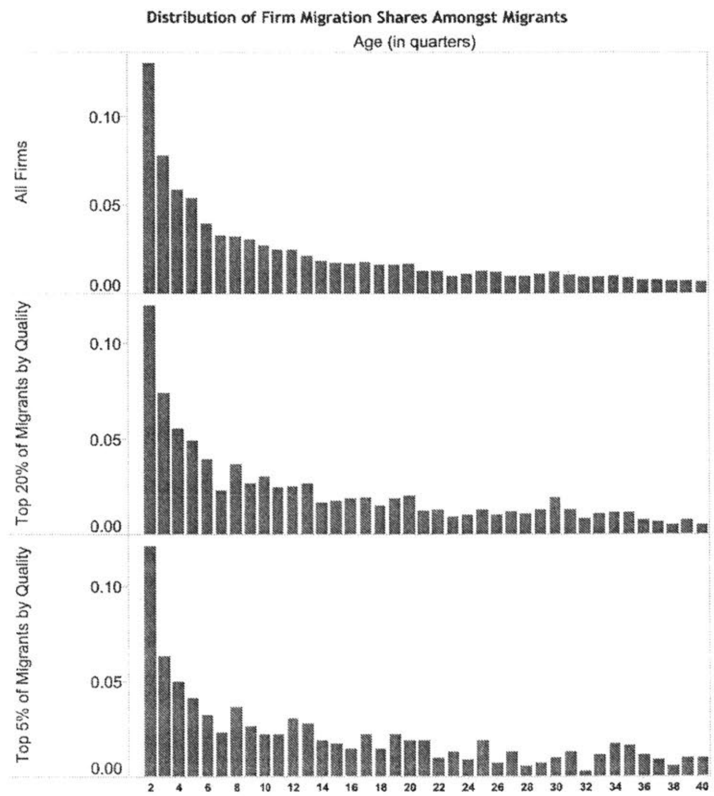
Notes: This figure compares the quality distribution of movers and non-movers, the distribution of non-movers is lower quality than movers.

FIGURE 3



Note: This graph plots the Entrepreneurial Quality of each of the 162 MSAs in my sample. MSA Entrepreneurial Quality is simply the average quality of all local firms to an MSA during the time period.

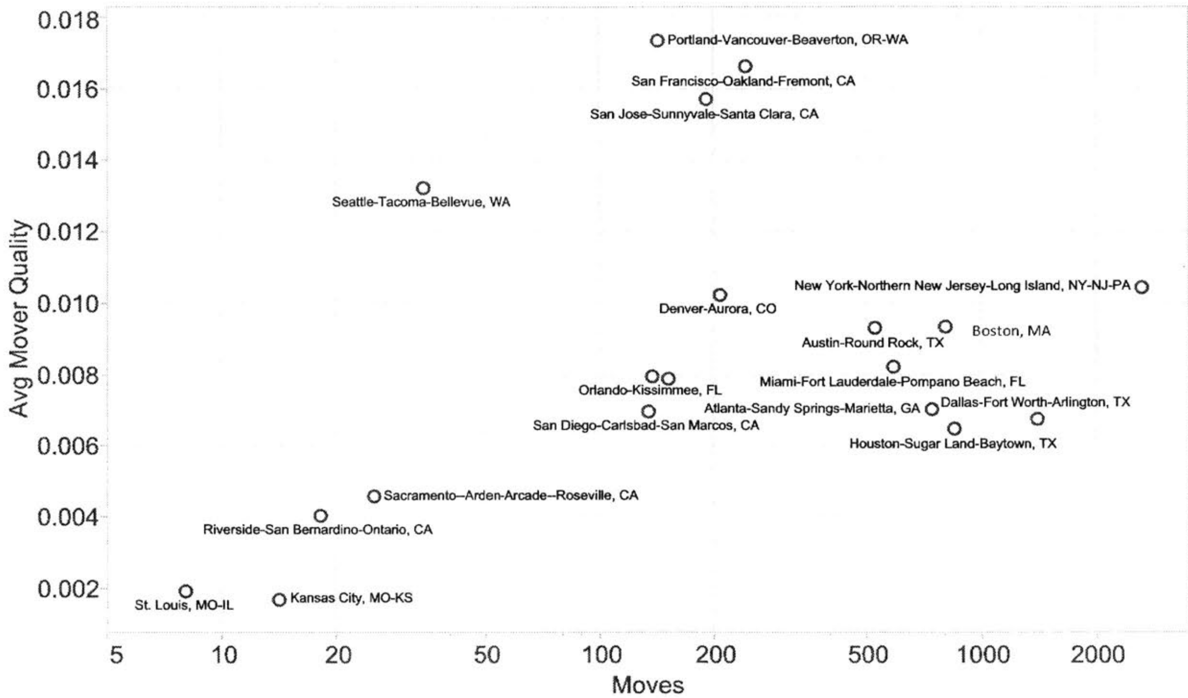
FIGURE 4



Note: this figure shows the share of migrant firms that migrate within each of the quarters of firm life at different levels of firm quality. Firms are required to live at least a quarter in their location of birth to be considered migrants. It is easy to see a monotonic decline in migration rates, suggesting migration is mostly entrepreneurial.

FIGURE 5

MIGRATION RATES TO LARGEST CITIES



APPENDIX

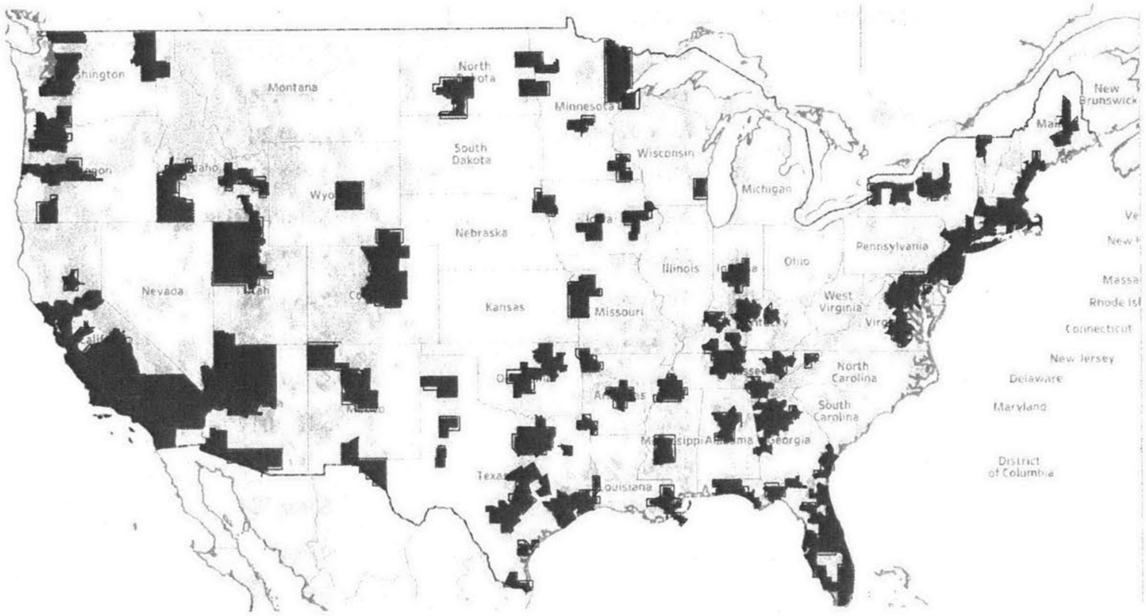
TABLE A1

List of States in Dataset

Rank in US GDP	*State*	*GDP*
1	*California*	$2,287,021
2	*Texas*	$1,602,584
3	*New York*	$1,350,286
4	*Florida*	$833,511
8	*New Jersey*	$560,667
10	*Georgia*	$472,423
11	*Virginia*	$464,606
12	*Massachusetts*	$462,748
13	*Michigan*	$449,218
14	*Washington*	$425,017
17	*Minnesota*	$326,125
18	*Colorado*	$309,721
19	*Tennessee*	$296,602
21	*Arizona*	$288,924
22	*Missouri*	$285,135
25	*Oregon*	$229,241
27	*Oklahoma*	$192,176
28	*South Carolina*	$190,176
29	*Kentucky*	$189,667
30	*Iowa*	$174,512
32	*Utah*	$148,017
34	*Arkansas*	$129,745
39	*New Mexico*	$95,310
43	*Idaho*	$66,548
45	*North Dakota*	$62,772
46	*Alaska*	$60,542
47	*Maine*	$56,163
49	*Wyoming*	$48,538
50	*Rhode Island*	$45,962
52	*Vermont*	$30,723
	GDP in Sample	$12,134,680
	US GDP	$17,411,875
	Share of US GDP in Sample	70%

FIGURE A1

MSAs in Sample



Note: Figure indicates all MSAs in sample, a total of 164, except for Anchorage, AK which is not in the figure. The set of MSAs includes all MSAs with at least one migrant for all states in the business registration dataset used, except for three cities where institutional details on how registrations are recorded oversampled migrations to those (Phoenix, AZ, Minneapolis, MN, and Albany, NY). MSAs that never had a migrant that was registered under Delaware jurisdiction, such as the Laredo, TX MSA, are not included either

CHAPTER 5

CONCLUSION

This dissertation has sought to contribute to scholar research on the economics of entrepreneurship in two ways: methodologically, by developing a novel approach to measure the quality of firms, and, substantively, by using this approach (and the underlying data infrastructure) to study three questions of theoretical interest.

My hope is that entrepreneurial quality as a tool provides benefits in many avenues of research and policy. One possibility, for example, is policy targeting. As recently as 2011, Alan Blinder noted that "one thing economists still agreed on, he thought, was that the government was incapable of picking those winners (i.e. 'growth firms') ex-ante" (see discussion notes for Hurst and Pugsley, 2011). It is easy to see how an entrepreneurial quality approach could help solving this issue, and it is my expectation that it will soon be improved to allow for careful policy targeting and efficient policy–making. Many other applications are in principle possible.

I also hope each of the research insights offered in this dissertation are useful in the field's quest to understand the phenomenon of entrepreneurship. Each chapter is unique in its ability to shed light on a long-standing question of interest, and opens new and exciting questions. Some of them, I am already working—with co-authors—to find an answer; others, will be best answered by other people at other times.

The last chapter, "Entrepreneurial Migration", offers perspective on a new area of entrepreneurship research—the migration of entrepreneurs—, that has had little, if any, exploration. It's the beginning of ongoing work, and I expect to provide more perspective on this topic in future papers, including analyzing the performance, and social consequences of location changes for entrepreneurial firms.

Most importantly, I hope this dissertation benefits societies and economies around the world. The goal of the social scientist is to understand how the world works, and how to make it better. Entrepreneurship, made of private inventiveness and commercialization of new ideas and services, is often seen as the lynchpin of a process of social and economic improvement. It is my sincere hope that the work put in creating dissertation, and the insights that derive from it, provide meaningful progress towards such goal for the benefit of all human society.

Thank you

Jorge Guzman
Massachusetts Institute of Technology
Cambridge, Massachusetts
2017