The State of American Entrepreneurship:

New Estimates of the Quantity and Quality of Entrepreneurship for 34 US States, 1988-2014[1]

Jorge Guzman, MIT

Scott Stern, MIT and NBER[2]

May 2017

[1]Acknowledgments: We are thankful for comments and suggestions by Marianne Bertrand, Erik Brynjolffson, Ankur Chavda, Catherine Fazio, Joshua Gans, John Haltiwanger, Bill Kerr, Fiona Murray, Abhishek Nagaraj, Roberto Rigobon, David Robinson, and Hal Varian, as well as seminar and conference participants at Duke University, Harvard Business School, University of Toronto, the University of Virginia, the Kauffman Foundation New Entrepreneurial Growth Conference, and the NBER Pre-Conference on Entrepreneurship and Economic Growth. We also thank Open Corporates for providing data for New York and Michigan, and to RJ Andrews for his development of the visualization approach. Sarah Andries, Jintao Chen, and Ji Seok Kim provided excellent research assistance. Finally, we acknowledge and thank the Jean Hammond (1986) and Michael Krasner (1974) Entrepreneurship Fund and the Edward B. Roberts (1957) Entrepreneurship Fund at MIT, and the Kauffman Foundation for financial support. All errors and omissions are of course our own.

[2] Jorge Guzman: 100 Main Street, E62-484. Cambridge, MA. 02142. jorgeg@mit.edu. Scott Stern: 100 Main Street, E62-476. Cambridge, MA. 02142. sstern@mit.edu.

The State of American Entrepreneurship:
New Estimates of the Quantity and Quality of Entrepreneurship for 34 US States, 1988-2014

ABSTRACT

Assessing the state of American entrepreneurship requires not simply measuring the quantity but also the initial quality of new ventures. We use comprehensive business registries and predictive analytics to build estimates of entrepreneurial quantity and quality from 1988-2014 for 34 states. In contrast to prior work highlighting a secular pattern of declining business dynamism, a quality-adjusted index of entrepreneurship follows a cyclical pattern, exhibiting a sharp rise during the dot-com boom and after the Great Recession. Conditioning on the level of initial entrepreneurial quality, the probability of growth in the form of an IPO or significant acquisition has declined sharply since the mid-1990s. Entrepreneurial quality is highly concentrated geographically, particularly in well-known start-up hubs such as Silicon Valley, Boston and Austin. These estimates can be used to offer new insight into the interplay between entrepreneurship and economic growth. First, consistent with a theory of investment cycles, we find a positive relationship between GDP growth and the subsequent quality-adjusted quantity of entrepreneurship; in contrast, there is no relationship between GDP growth and quantity (without quality adjustment). We separately take advantage of the regional variation in the data to evaluate the relationship between the initial quantity and quality-adjusted quantity of entrepreneurship and subsequent 10-year economic growth; while there is no impact of quantity on growth, there is a strong positive association between quality-adjusted quantity and subsequent economic performance. Our results offer a novel perspective on the distribution and dynamic evolution of entrepreneurship in the United States, and the relationship between entrepreneurship and economic performance.

I. Introduction

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 and Haltiwanger, 1992; 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*. These studies established that a disproportionate share of new job creation has been linked to new firms, and economic growth is grounded in 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 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.

Notwithstanding these advances, there is increasing recognition that the relationship between entrepreneurship and economic growth depends not simply on the quantity but also on the underlying *quality* of new firms (Schoar, 2010; Hurst and Pugsley, 2011). While systematic population-level indices of the quantity of entrepreneurial activity (such as the Business Dynamics Statistics database, hereafter BDS) document a secular decline in the rate of business dynamism and the "aging" of US private sector establishments (Hathaway and Litan (2014a, 2014b, 2014c)), researchers focused on venture capital and high-growth firms have documented a sizeable increase after the Great Recession in the funding of growth-oriented entrepreneurial ventures (Gornall and Strebulaev, 2015).

To put these differences in perspective, consider the gap, for the 34 states which will form the basis of our analysis, between 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). 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.[3]

How can we resolve this puzzle? Building on Guzman and Stern (2015; 2017), this paper focuses not only on the quantity of entrepreneurship (nor on highly selective measures of the rate of successful entrepreneurs) but instead on the measurement and assessment of entrepreneurial "quality." While it has long been known that the growth consequences of start-up activity are concentrated in the outcomes of a very small fraction of the most successful firms (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. The challenge is fundamentally 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).

Our approach to measuring entrepreneurial quality combines three interrelated insights.[4] First, a practical requirement for any growth-oriented entrepreneur is business registration (as a

[3] This divergence is reinforced by comparing BDS firm births and economic growth. While the BDS has little cyclical variation (and is on a downward decline), GDP growth is far more variable with a sharp upward trend through the 1990s and a downward decline over the subsequent period. In recent work, Haltiwanger

[4] In our earlier work, we undertook preliminary explorations of the approach that we develop in this paper. In Guzman and Stern (2015), 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, 2017). This paper builds on these earlier exercises to develop an analysis for 34 "representative" US states (comprising more than 80%

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 34 states comprising ~ 83% of total US economic activity over a 27-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., for equity growth, we can observe firm that achieve an IPO or high-value acquisition within six years of founding; we also measure alternative measures of growth including employment), 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

---

of overall GDP) a 27-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).

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 (or RECPI / GDP) 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 34 states included in our sample for the period from 1988-2014, documenting several key findings.[5] First, in contrast to the secular and steady decline observed in the BDS, RECPI / GDP has followed a cyclical pattern that seems sensitive to the capital market environment and overall economic conditions. Second, while the peak value of RECPI / GDP 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, with an additional 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

[5] 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.

their potential (REAI): REAI declined sharply in the late 1990s, and did not recover through 2008. While we focus on estimates of entrepreneurial quality based on a predictive model of equity growth outcomes (the achievement of an IPO or significant acquisition within six years of founding), these broad patterns of results also hold if one focuses on alternative definitions of equity growth (e.g., only focusing on IPOs) or alternative growth measures such as the realization of more than 500 employees within the first six years after founding.[6]

Beyond these descriptive patterns, we are able to use these new measurements to test specific hypotheses about the drivers and potential consequences of entrepreneurship and economic activity. We highlight two key theoretical tests. First, building on the theoretical predictions of the 'investment cycles and innovation' literature (Nanda and Rhodes-Kropf, 2013), we evaluate how the quantity and quality-adjusted quantity of entrepreneurship depend on shocks to the economic growth rate; while arguments dating back to Schumpeter's perspective on the "cleansing effect" of recessions suggests that economic downturns might actually spur entrepreneurship—specifically high-quality entrepreneurship—Nanda and Rhodes-Kropf argue that the negative financing environment that arises during economic downturns reduces the equilibrium level of entrepreneurial quality (i.e., we should see a reduction in high-quality entrepreneurs during economic downturns). We test this hypothesis using structural vector autoregression (SVAR) on the entire U.S. economy, finding that quality-adjusted quantity is procyclical and is negatively impacted by recessions (while the impact on quantity alone is noisy and if anything countercyclical). Second, beyond this aggregate test, we also conduct a simple exercise evaluating the impact of entrepreneurial quantity and quality-adjusted quantity on regional economic performance. In a simple cross-sectional growth regression, we find suggestive

[6] Our employment growth results are based on a private sector data source, Infogroup USA. While we highlight the robustness of our core findings to this potentially noisy measure of employment growth in this paper, we do not undertake a systematic assessment of employment-oriented growth outcomes which could more naturally be conducted with a comprehensive public dataset such as the LBD.

evidence that while regional quantity has no correlation with subsequent economic growth, the regional RECPI predicts subsequent economic growth.

Though our focus in on the aggregate level of entrepreneurship, we also document striking regional and intra-regional variation I entrepreneurial quality. Consistent with Guzman and Stern (2015), we document an extremely high and persistent level of entrepreneurial quality in regions such as Silicon Valley, but a more detailed view of this region highlights that, even within Silicon Valley, high growth entrepreneurship is concentrated in only a few areas, with a much broader distribution of low-growth entrepreneurship.

Our approach of course comes 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 their 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, though we show some robustness of our findings to the use of employment-oriented growth outcomes, a more complete assessment of the differences between equity growth outcomes and employment-oriented outcomes remains outstanding.

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[7]. 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.

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.

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—an emphasis of this paper—, characterizing entrepreneurial quality can allow, for example, for novel tests of alternative theories of the firm-size distribution (e.g., testing the relative importance of initial conditions versus firm dynamics),

[7] In recent work, Decker et al (2016) use business dynamics estimates to also document important variation in rates of dynamisms across industries.

evaluate alternative potential entrants in the context of industrial organization models, and 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.

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. Section VII compares the relationship of our index to an alternative measure of economic growth using employment outcomes. In section VIII, we empirically test the relationship between GDP growth and changes in national entrepreneurship. Section IX studies the correlation between regional entrepreneurship and regional economic growth. And Section X studies variation of entrepreneurial quality and potential in the micro-geography of Silicon Valley. Section XI 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),[8] 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).[9] Despite

[8] 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).

[9] 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; Klette and Kortum, 2004; Akcigit and Kerr, 2010; and Luttmer, 2007).

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 probabilities of survival (Evans, 1987; Dunne, Roberts, and Samuelson, 1988).[10]

Davis and Haltiwanger (1992) clarified this empirical debate by considering both the role of size and age at the same time, and developed a systematic 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 (Jarmin, Haltiwanger, and Miranda, 2013; Akcigit and Kerr, 2010). Building on these studies, Decker et al (2014) further use 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, with meaningful variation across industry groups (Decker et al, 2016).

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. In other words, for many questions in 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

[10] 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, 1995; Klepper, 1996; Hopenhayn, 1992; Klette and Kortum, 2004).

underlying differences in the firms themselves (Hurst and Pugsley, 2011; Kaplan and Lerner, 2010; Schoar, 2010). 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 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.[11] 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

[11] 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.

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 B, 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 outcome g. 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 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 (2015; 2017).

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 at-birth 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 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).[12]

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:

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

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.

[12] 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 (2017), 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.

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:

$$RECPI_{r,t} = EQI_{r,t} \times N_{r,t} \quad (3)$$

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:

$$REAI_{r,t} = \frac{\sum g_{i,r,t}}{RECPI_{r,t}} \quad (4)$$

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. Appendix C of the Supplementary Materials 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, 2015; 2017).

We focus on 34 US states from 1988-2014[13]. 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 (or Secretary of the Commonwealth) 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.[14]

[13] These are Alaska, Arizona, Arkansas, California, Colorado, Florida, Georgia, Idaho, Illinois, Iowa, Kentucky, Maine, Massachusetts, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Rhode Island, South Carolina, Tennessee, Texas, Utah, Vermont, Virginia, Washington, Wisconsin, and Wyoming.

[14] This section draws on Guzman and Stern (2015, 2017), where we introduce the use of business registration records in the context of entrepreneurial quality estimation.

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 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. The resulting dataset contains 29,961,838 observations.[15] 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[16], as reported in Thomson Reuters SDC database[17]. During the period of 1988 to 2008, we identify 6,708 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

[15] 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.

[16] 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

[17] Although the coverage of IPOs is likely to be nearly comprehensive, the SDC data set excludes some acquisitions. SDC captures their list of acquisitions by using over 200 news sources, SEC filings, trade publications, wires, and proprietary sources of investment banks, law firms, and other advisors (Churchwell, 2016). Barnes, Harp, and Oler (2014) compare the quality of the SDC data to acquisitions by public firms and find a 95% accuracy; Nette, Stegemoller, and Wintoki (2011), perform a similar review. While we know this data not to be perfect, we believe it to have relatively good coverage of 'high value' acquisitions. Further, none of the cited studies found significant false positives, suggesting that the only effect of the acquisitions we do not track will be simply an attenuation of our estimated coefficients.

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 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.[18] We hypothesize that eponymous firms are likely to be associated with lower entrepreneurial quality. Our second 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.[19]

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

[18] Belenzon et al (2014; 2017), perform a more detailed analysis of the interaction between eponymy and firm performance, highlighting name as a signal chosen by entrepreneurs given differences in growth intention.

[19] 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.").

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 (Appendix C).

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.[20] 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 of binary outcomes (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 57% increase in the probability of growth, and having an eponymous name with a 78% *lower* probability of growth. Corporate form measures are also significant. Corporations are 3.7 times more likely to grow and firms registered under Delaware jurisdiction (instead of the local jurisdiction) are 38 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 0.14—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 141 times more likely to grow, while firms with a trademark are 86 times 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 72% 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

[20] 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.

with the biotechnology sector are 16 times more likely to grow, firms associated with ecommerce 1.8 times, associated to IT 5 times, medical devices 3 times, and 19 times for firms with name associated to semiconductor industries. These coefficients 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.[21] 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 to account for idiosyncratic differences across states in corporate registration policies and fees. Though differences across states likely influence the "marginal" registrant (and would be of independent interest), it is unlikely that firms with significant growth potential would be deterred from registration depending on the state in which they were founded.[22]

[21] 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).

[22] In unreported results, we verify all our main results are not sensitive to the exclusion or inclusion of these fixed effects.

Models (3-1) through (3-3) investigate the joint role of different groups of measures after including state fixed effects. 3-1 investigates corporate governance measures, corporations are 5.3 times more likely to grow and Delaware firms are 45 times more likely to grow. Since these are incidence-rate ratios (odds-ratios), the joint coefficients can be interpreted multiplicatively: Delaware corporations are 239 times more likely to grow ($45 \times 5.3 = 238.5$). 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 2 times more likely to grow while eponymous firms are 85% *less* likely to grow. Finally, in column 3, we study the role of intellectual property measures to *Growth*. Firms with a patent are 68 times more likely to grow and firms with a trademark are 14 times more likely to grow.

In (3-4) and (3-5) we develop predictive models by including the measures in prior models plus industry controls. Our first specification (3-4) 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 36 times more likely to grow. Our two industry-agnostic name-based measures are informative as well. Firms with a short name are 2 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 67% less likely to grow, firms associated with traded industries are 9.5% more likely to grow, and firms specifically associated with resource intensive traded industries are 27% more likely to grow. Names associated with specific high-technology sectors are also associated with growth: firms related to biotechnology are 3.5 times more likely to grow, firm associated with e-commerce are 15% more likely to grow, firms associated with IT

2.1 times, firms associated with medical devices 19%, and firms associated with semiconductors 2.9 times more likely to grow.

We extend this specification in (3-5) 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.[23] In particular, receiving a patent is associated with a 46 times increase in the likelihood of growth for non-Delaware firms, and the combination of Delaware registration and patenting is associated with a 199 times increase in the likelihood of growth (simply registering in Delaware without a patent is associated with only a 29X 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.[24]

These two models offer a tradeoff. On the one hand, the "richer" specification of model (3-5) 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 (3-5) is not as timely as model (3-4). Indeed, specifications that rely exclusively on information encoded within the business registration record can be

[23] 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.

[24] 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, 2015) and Massachusetts (Guzman and Stern, 2017). Figure A2 reports the coefficients associated with each state-level fixed effect; overall, our results are not sensitive to the inclusion or exclusion of these fixed effects in our regression analysis, predictive analytic estimates, or mapping of entrepreneurial quality.

calculated on a near real-time basis, and so provide the most timely index for policymakers and other analysts.[25] We will calculate indices based on both specifications; while our main historical analyses will be based off the results from model (3-5), model (3-4) 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 (3-4) and refer to (3-5) as the "full information" model.

Robustness and Predictive Quality. In Figure 2, we evaluate the predictive quality of our estimates by undertaking a tenfold cross-validation test (Witten and Frank, 2005),[26] and report the out-of-sample share of realized growth outcomes at different portions of the entrepreneurial quality distribution. The results are striking. The share of growth firms in the top 5% of our estimated growth probability distribution ranges from 64% to 69%, with an average of 66%. The share of growth firms in the top 1% ranges from 42% to 52%, with 47% on average (interestingly, these results are extremely similar to the findings for California from Guzman and Stern (2015) and Massachusetts from Guzman and Stern (2017)). Growth, however, is still a relatively rare event even among the elite: the average firm within the top 1% of estimated entrepreneurial quality has only a 1.6% chance of realizing a growth outcome.

In Table 4, we repeat our full information model with a series of robustness tests to verify that the magnitudes in our model are not driven by variation across years or states. In (4-1) we report a variation of our model after also including year fixed-effects, (4-2) includes state-specific time trends, and (4-3) incudes both year fixed-effects and state-specific time trends (note that these

[25] 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.

[26] 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.

cannot be included in our predictive model as we would not know the fixed-effect value for future years). 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.

Finally, in Table A1 of our appendix we further study differences in the robustness of our index across states by estimating the out-of-sample share of firms in the top 5% and top 10% of quality by state, and the correlation between entrepreneurial quality estimates performed individually for each state and our national estimate. While there is variation in each of these statistics across states, all of them indicate a relatively strong correlation between quality at the state level and our national measure.

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 34 states.[27] Finally, we also include a confidence interval estimated through a Monte

[27] It is also possible to adjust by population instead of GDP. RECPI / population shows a starker positive increase than RECPI / GDP, as GDP per capita has also increased through the time period represented.

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.[28] 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 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).

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

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. 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. Figure 4 illustrates the geographic variation in entrepreneurial quality for the 34 states in our sample. For the 29 states where we have a zip-code level address for each firm's founding location, we present RECPI by zip code, where the size of each point is equal to the quantity of entrepreneurship, and the color of the point indicates the EQI for that zip code (with darker coloring indicating a higher EQI). For six states, we do not have consistent micro-address information for each firm, and so we simply distribute uniformly the state total firms across all ZIP Codes and assign the average state quality to each ZIP Code.[29]

This map offers insight into the distribution of entrepreneurial quantity and quality across the United States. First, the most intense areas for entrepreneurial potential are in well-known entrepreneurial ecosystems such as Silicon Valley, Boston, and Austin. Second, several large cities, including Los Angeles, Houston, Dallas and even Detroit host not simply a high level of new registrants but a high average level of entrepreneurial quality among their start-ups. Third, a number of smaller cities that host strong research institutions such as Santa Fe, Boulder, and El Paso register a high average EQI. At the same time, there are large areas of the United States that

[29] These six states are Arkansas, Ohio, New York, Oklahoma, Virginia, Washington, also highlighted with a white border in Figure 4.

host a high level of entrepreneurship but where estimates of start-up quality are relatively low. Florida, in particular, seems to have a very high average quantity with low average quality. Many of the Mountain States (e.g., Wyoming, Idaho, and Utah), as well as Northern New England (Vermont and Maine) also seem to have a relatively low average estimated quality even within key cities such as Salt Lake City.

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 from 1988 to 2014, 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 = Number of 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, we 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. [30]

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 0.75 down to 0.51. 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 in the availability of entrepreneurial finance. While this rise is economically important, its realization once all growth outcomes realize is still to be seen.

[30] 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.

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. Equity versus Employment Growth Outcomes

Our analysis so far has used a measure of equity growth – an IPO or significant acquisition – as the measure of entrepreneurial success used to construct our predictive analytic and resulting estimates of entrepreneurial quality. However, while an important measure of success for founders and investors, alternative measures of entrepreneurial success more closely tied to broader economic performance could include a measure of employment (or productivity) (e.g. Krishnan, Nandy, and Puri, 2014; Davis and Haltiwanger, 1992). While a full analysis of the relationship between business registration observables and comprehensive employment outcomes is beyond the scope of this paper (as such an analysis could more naturally be conducted in the context of an integrated longitudinal database such as the LBD), we undertake a preliminary robustness check to evaluate how the use of an employment-based success metric influences our analysis and findings. To do so, we take advantage of a dataset of employment levels for more than 10 million firms available from Infogroup USA (a private sector business database similar to Dun and Bradstreet) to construct two new outcome variables, *500 Employment Growth* and *1000 Employment* Growth, each equal to 1 for all firms recorded as having greater than 500 or 1000

employees or more respectively within 6 years, and 0 otherwise.[31] Though this measure does not capture the employment levels of all firms (and all employment data are themselves categorical estimates rather than the fine-grained measures available through administrative data), this rough cut allows us to identify the vast majority of firms that experience the (rare and usually highly observable) event of becoming a large employer in a relatively short period of time.[32]

We use these data to conduct three interrelated exercises. First, in Table 5, we compare our baseline entrepreneurial quality model using *Growth* versus the *Employment Growth* measures as the dependent variable. The estimates are surprisingly similar not just in sign but also in relative magnitude, with a higher concordance between *Growth* and *Employment Growth 1000*. For example, firms with a trademark are 6.7 times more likely to get 1000 employees (6.8 times for equity growth), firms with a patent 10.1 times (27.6), firms registered in Delaware 46.9 times (42.7), and firms with both a patent and Delaware registration 119.6 times (184.8). This similarity between coefficients suggests that our baseline model not only captures financial outcomes but also captures significant variation across firms in their potential to achieve a rare and outsized employment growth outcome.

As a second exercise, we use the model with the lower level of concordance (*Employment Growth 500*) as an alternative baseline for our predictive analytic to form an entrepreneurial quality estimate for each firm in our sample and compare our initial entrepreneurial quality estimates with this alternative. The Pearson correlation coefficient between a predictive analytic based on *Growth* versus *Employment Growth 500* is 0.90.

Finally, we examine how the incidence of *Employment Growth 500* is predicted by our estimates of entrepreneurial quality using our baseline equity growth regression and report the share of firms that achieve employment growth in the top 5% and 10% of quality. The results are

[31] 500 employees is also the threshold to qualify as a small business under the guidelines of the US Small Business Administration, effectively allowing us to measure the likelihood of a business being large under this definition.

[32] This dataset has also been used by McDevitt (2014, 2011), Waldfogel (2008), Seim (2006) and Ellickson (2007).

striking: more than 50% of all measured employment growth outcomes occur within the top 10% of our entrepreneurial quality distribution, with more than 40% in the top 5%.

While we emphasize that this analysis is incomplete insofar as our measures of employment growth may be incomplete, our analysis nonetheless suggests that there is a meaningful relationship between equity and employment growth, and that both of these highly skewed outcome variables have a predictable relationship with measures of underlying entrepreneurial quality.

VIII. The Impact of Investment Cycles on Entrepreneurial Quantity and Quality

While our analysis so far has been largely descriptive, the quantity and quality metrics we construct can be used to offer new evidence about the economy-wide drivers of entrepreneurship. Consider the relationship between entrepreneurship and business cycles. On the one hand, arguments dating back to Schumpeter highlight the potential "cleansing effect" of recessions; during a downturn, the weakening of existing firms, as well as a higher "threshold" for entry quality, might result in a decrease in quantity but an increase average firm quality of firms (Schumpeter, 1934, Aghion and Howitt, 1992). However, as emphasized by Nanda and Rhodes-Kropf (2013), investors funding a new venture may be concerned about the availability of follow-on financing, particularly for riskier projects with high-potential, and this concern may be exacerbated during a downturn; indeed, they show that venture-backed firms backed during a downturn are actually less likely to go bankrupt, but, conditional on success, exhibit a higher level of innovativeness and scale. In a related vein, recent evidence by Moreira (2016) documents that companies born during a recession start at a smaller scale (in terms of initial employees), grow at a slower rate, and register a lower TFP level, which she interprets as a reduction in the level of underlying entrepreneurial quality. While informative, these prior estimates abstract away from direct population-level estimates of how underlying firm quality changes during a boom and bust

(e.g., estimates based on initial employment levels confound quality with for example the level of early demand).

We are able to use our estimates to offer novel evidence on the overall relationship between economic fluctuations and both the quantity and quality of entrepreneurship. In particular, consider a simple two-equation system in which economic growth affects entrepreneurship through both its current and prior level, while the impact of entrepreneurship on growth is only manifested after a lag (i.e., a higher quality-adjusted quantity of entrepreneurship influences economic growth in the time periods after founding). These relationships can be summarized as follows:

$$Ln(\Delta GDP_t) = a_1 Ln(\frac{RECPI_{t-1}}{GDP_{t-1}}) + \cdots + a_n Ln(\frac{RECPI_{t-n}}{GDP_{t-n}}) + b_1 Ln(\Delta GDP_{t-1}) + \cdots + b_n Ln(\Delta GDP_{t-n}) + u_t \quad (5)$$

$$Ln(\frac{RECPI_t}{GDP_t}) = c_1 Ln(\frac{RECPI_{t-1}}{GDP_{t-1}}) + \cdots + c_n Ln(\frac{RECPI_{t-n}}{GDP_{t-n}}) + d_0 Ln(\Delta GDP_t) + d_1 Ln(\Delta GDP_{t-1}) d_n Ln(\Delta GDP_{t-n}) + v_t \quad (6)$$

where $Ln(\Delta GDP_t)$ represents GDP growth (in logs) in time t and $Ln\left(\frac{RECPI_t}{GDP_t}\right)$ represents the quality-adjusted flow of entrepreneurship in *t* that year, and u_t and v_t are the idiosyncratic disturbances in the growth rate and the entrepreneurship rate respectively. This system can be estimated as a recursive SVAR model where the d_n coefficients represent the percent increase in quality-adjusted entrepreneurship from an increase in the annual GDP growth rate.

Using the (admittedly small) sample of 27 annual observations for the United States from 1988-2014 inclusive, Table 6 reports coefficient estimates from this approach as well as equivalent regressions using only the quantity of firms instead of RECPI. Figure 6 presents the impulse response functions. We begin in (6-1) and (6-2) with a reduced-form single-lag VAR model; while (6-1) reports a positive relationship between GDP growth in *t-1* and quality-adjusted

entrepreneurship in *t*, (6-2) indicates no relationship between changes in GDP growth and start-up quantity. We then turn to a three-lag SVAR model that allows not only to capture dynamics but also allows for contemporaneous impact between GDP growth and entrepreneurship.[33] These structural parameter estimates are similar: a positive relationship between GDP growth and subsequent quality-adjusted entrepreneurship and no relationship between GDP growth and the raw quantity of entrepreneurship. This can be seen directly in Figures 6A and 6B, which map the impulse response function for each. To put these results in context, a doubling of the GDP growth rate leads to a 2% increase in $\frac{RECPI}{GDP}$ in current year *t*, a 4% increase in years *t+1* and *t+2*, and then tapers off. In contrast, as illustrated in Figure 6B, there is only a noisy and insignificant net relationship between GDP growth and $Ln\left(\frac{Obs}{GDP}\right)$. We further this analysis in (6-4) and (6-5) by considering an alternative measure of economic performance: the presence of absence of an economic recession (as determined by the NBER Recession Dating Cycles). According to the estimates as summarized in the impulse response functions in Figures 6C and 6D, the onset of a a recession decreases $\frac{RECPI}{GDP}$ by 5% in t and t+1 (and then tapers off), with no net impact on $Ln(\frac{Obs}{GDP})$.

We emphasize that these results should be viewed with caution. We are basing our inferences on only a relatively short time-series, and it is of course possible that the relationship between economic performance and entrepreneurship changes over time and place. With that important caveat, these results are consistent with the hypothesis that, while economic shocks have an ambiguous and noisy impact on the overall start-up rate, there is a meaningful relationship in which negative economic shocks diminish the propensity for starting ventures with high growth potential at founding.

[33] The three-lag structure minimizes both the Aikake Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (SBIC).

IX. Do Changes in Entrepreneurial Quality Correlate With 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 (MSA number of new firms), and EQI from our data, as well as MSA GDP measures obtained from the Bureau of Economic Analysis.[34] Our data contains 209 MSAs for which we can estimate all measures. Our core specification is a simple "long differences" analysis, in which we examine the relationship between GDP growth from 2003 to 2013 as a function of the initial level of GDP (average between 2001-2003) and the initial quantity and quality of entrepreneurship (both measured as an average between 2001-2003 for number of firms and EQI).[35]

Figure 7 shows the scatterplot and correlation between log GDP growth and our two entrepreneurship measures *Ln(MSA Entrepreneurial Quantity)* (Panel A) and *Ln(MSA Entrepreneurial Quality)* (Panel B). The fitted line is weighted by MSA GDP in 2003. The relationship between EQI and GDP growth is positive, with a slope of .06, and significant at the 1 percent level with robust standard errors. The relationship between quantity and GDP growth, though noisier and lower in magnitude, is also slightly positive, with a slope of .005, and not significant.

In Table 6 we measure this relationship in a regression framework comparing the effect of simply the quantity of firms and of the quantity-adjusted quality of firms at the MSA level (*MSA RECPI*) to GDP Growth. Models (6-1) and (6-2) measure the correlation between each variable and GDP Growth. *Ln(MSA RECPI)* has a positive and statistically significant relationship while the coefficient for *Ln(MSA Entrepreneurial Quantity)* is lower and much noisier. Model (6-3) includes both measures together, and (6-4) also includes the level of GDP—*Ln(MSA GDP)*—as a

[34] Formally, the BEA calls MSA GDP "Gross Metropolitan Product".
[35] In Guzman and Stern (2017) we perform a more in-depth analysis on this relationship.

control. *Ln(MSA RECPI)* continues to have a positive and significant coefficient of 0.024, while quantity has a negative coefficient of –0.032.

Models (6-5) and (6-6) repeat the analyses of (6-3) and (6-4) but weight the observations by the GDP of each MSA. In (6-6)—our main specification—the coefficients both entrepreneurship measures more than double, and continue to be statistically significant. *Ln(MSA RECPI)* increases to 0.0575 while *Ln(MSA Entrepreneurial Quantity)* drops to –0.0638.

These results are striking. While the initial level of GDP and number of firms have no positive (and sometimes negative) 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 5.75% in GDP 10 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 15X increase in quality), moving from the bottom to the top of the distribution of initial entrepreneurial quality is associated with a 86% increase in GDP growth.

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 number of firms 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 entrepreneurship 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 performance, our

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

X. Entrepreneurial Quality In Silicon Valley: A Case Study

While our results so far have focused on the aggregate experience across 34 (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 8A. 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 micro-spatial distribution of entrepreneurial quality (EQI) across Silicon Valley. In Figure 7B, we present all locations in the San Francisco Bay area that have a firm registered in 2012. The size of the point represents the number of firms, and the color of the point the quality of those firms. Three observations are apparent.

First, there appears to be entrepreneurship everywhere. New business formation occurs in almost every populated location of the bay area. However, high quality entrepreneurship is highly clustered, even within this geographic area. It is possible to appreciate a few clusters of high quality entrepreneurship across the traditional area of Silicon Valley at the south of the bay (with a particular strong cluster exactly around the Stanford campus), as well as on the San Mateo area, in Market Street in the northeast corner of San Francisco, and around Berkeley. And third, even in Silicon Valley—which contains the highest entrepreneurial quality ZIP Codes in the United States— the bulk of entrepreneurship is not high quality entrepreneurship. Instead, most of it likely represents businesses more related to of 'local economy', serving local residents and there are very large areas, such as the west side of the bay, with virtually no high growth entrepreneurship.

XI. Conclusion

This paper develops a quality-based approach with business registration records for 34 states to create and evaluate synthetic entrepreneurship indexes at the national level. Not simply a matter of data, the predictive analytics approach 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), the ability of an entrepreneurial ecosystem to realize that potential over time (REAI) and the relationship between entrepreneurship and overall economic performance.

This analysis offers several new findings about the state of American entrepreneurship. First, in contrast to the secular decline observed in aggregate quantity-oriented measures of business dynamism (Decker et al, 2014), 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. As well, a

quality-adjusted predictive analytics approach captures striking regional variation in the growth potential of start-ups across the United States, including the presence of strong ecosystems such as Silicon Valley or Boston and relatively quantity-oriented entrepreneurship regions such as Miami.

By accounting for quality, our estimates offer a different perspective on the role of start-ups in the US economy over the past thirty years. 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). Indeed, in contrast to the secular decline in start-up activity observable in the BDS, our estimates of U.S. RECPI indicate a net *upward* trend across the full time-series of our sample. For example, the rate of expected successful startups fell to its lowest point in 1991, and reached its second highest level in 2014 (the final year of our sample). This finding suggests that the challenges to growth arising from entrepreneurship may be less directly related to the lack of formation of high-growth potential startups, but instead by other dynamics or ecosystem concerns. In particular, while there is high cyclicality in RECPI / GDP, 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. In other words, 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.

Accounting for entrepreneurial quality through a predictive analytics approach is not simply a question of more nuanced measurement of the same phenomena. Instead, a shift towards entrepreneurial quality allows one to more directly connect entrepreneurship and overall economic performance. On the one hand, there is a positive (though not necessarily causal) relationship between the level of quality-adjusted quantity at the regional level and subsequent 10-year regional economic performance, while there is no statistical relationship between entrepreneurial quantity and regional growth. At the same time, economic shocks are associated with a procyclical impact on the quality-adjusted quantity of entrepreneurship, while there is only a noisy relationship with quantity alone. Though we caution against a causal interpretation of these relationships (e.g., a common factor such as a technology shock could both give rise to a higher level of entrepreneurial quality and also economic performance), our approach nonetheless connects the phenomena of entrepreneurship to overall economic performance in a much closer manner.

More generally, our analysis suggests that directly taking a quantitative approach to the measurement of entrepreneurial quality can yield new insight into the precursors and consequences of entrepreneurial ecosystems and the impact of entrepreneurship on economic and social progress. Several follow-on research directions are possible. First, our data reveal striking variation across regions and time in both the quality-adjusted quantity of entrepreneurship as well as the potential for growth condition on initial quality. Examining how regional and temporal determinants of entrepreneurial ecosystems impact both entrepreneurial quality and the growth process seems like promising area for future research[36]. Second, while our current analysis examines the link between entrepreneurial quality at founding and subsequent growth (measured as either equity or employment growth), it is separately possible to examine how particular institutions that impact

[36] In related work one of us (Guzman (2017)) finds that the quality-adjusted quantity of entrepreneurship in a region also meaningfully influences the rate at which high growth entrepreneurs move into the region, suggesting the potential for increasing returns in regional RECPI.

start-ups after founding (such as the receipt of venture capital) impact that growth process. For example, in Catalini, Guzman and Stern (2017), we examine both the selection into and impact of venture capital on start-up firms exploiting this predictive analytics approach, and further work connecting firm founding, capital investment, and growth is likely to allow for a more structured understanding of the role of external capital in start-up growth. Finally, a striking feature of our predictive analytics results is the unusual level of skewness in the entrepreneurial quality distribution (e.g., more than 50% of all equity growth outcomes are contained within the top 1% of the estimated quality distribution). Directly measuring the high level of skewness in the initial distribution of firms likely offers new insight into a number of areas, such as industrial organization, finance, and organizational economics.

REFERENCES

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

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*

Belenzon, Sharon, Chatterji, Aaron and Brendan Daley. 2017. “Choosing Between Growth and Glory” *Working Paper*

C. Churchwell. (2016). “Q. SDC: M&A Database”. *Baker Library – Fast Answers.* Url: http://asklib.library.hbs.edu/faq/47760. Accessed on January 17, 2017.

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”. *European Economic Review.* 86: 4-23.

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.

Ellickson, P. B. (2007), “Does Sutton Apply to Supermarkets?”, *RAND Journal of Economics* 38(1), 43–59

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

Gornall, Will, and Ilya A. Strebulaev. 2015. "The Economic Impact of Venture Capital: Evidence from Public Companies". Working Paper.

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

Guzman, Jorge, and Scott Stern. 2017. "Nowcasting and Placecasting Entrepreneurial Quality and Performance" *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*. Edited by John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar. University of Chicago Press. Chapter 2.

Guzman, Jorge, and Scott Stern. 2017. "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

Krishnan, Karthik Debarshi K. Nandy & Manju Puri, 2015. "Does Financing Spur Small Business Productivity? Evidence from a Natural Experiment," *Review of Financial Studies* 28 (6): 1768-1809

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.

McDevitt, R. (2011), ‘Names and Reputations: An Empirical Analysis’, American Economic Journal: Microeconomics 3(3), 193–209.

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

Moreria, Sara. 2016. “Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles”. *Working Paper.*

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

J. Netter, M. Stegemoller, and M. B. Wintoki. (2011) “Implications of Data Screens on Merger and Acquisition Analysis: A Large Sample Study of Mergers and Acquisitions from 1992 to 2009” *The Review of Financial Studies*. 24 (7): 2316–2357.

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.

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

Seim, K. (2006), “An Empirical Model of Firm Entry with Endogenous Product-Type Choices”, *RAND Journal of Economics* 37(3), 619–640.

Waldfogel, J. (2008), “The Median Voter and the Median Consumer: Local Private Goods and Population Composition”, *Journal of Urban Economics* 63, 567–582.

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

TABLE 1

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

(1) All variables are dummy variables with values of 0 or 1. A description of how each measure is built is available in the data appendix.
(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.

	Definition	Source	Mean	Std Dev
Outcome Variable				
Growth	1 if a firm achieves an equity growth outcome (IPO or acquisition) within 6 years or less, 0 otherwise. Estimated for cohorts 1988-2008	SDC Platinum	0.00033	0.01825
Corporate Form Observables				
Corporation	1 if a firm is registered as corporation, 0 if registered as LLC, or partnership.	Bus. Reg. Records	0.52	0.50
Delaware	1 if registered under Delaware jurisdiction, 0 if registered under local (focal state) jurisdiction	Bus. Reg. Records	0.031	0.172
Name Observables				
Short Name	1 if the firm name is two words or less, 0 otherwise.	Bus. Reg. Records	0.49	0.63
Eponymous	1 if first or last name of top manager (president, CEO, partner) is part of firm name, 0 otherwise.	Bus. Reg. Records	0.0679	0.2516
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. Only cohorts 1988-2012.	USPTO	0.0016	0.0404
Trademark	1 if firm obtains a trademark within a year of founding, 0 otherwise. Only cohorts 1988-2012.	USPTO	0.0012	0.0351
US CMP Cluster Dummies (2)				
Local	1 if firm name is associated to local industries, 0 otherwise.	Estimated from name	0.19	0.40
Traded (3)	1 if firm name is associated to traded industries, 0 otherwise.	Estimated from name	0.135	0.341
Traded Resource Int.	1 if firm name is associated to resource intensive industries, 0 otherwise.	Estimated from name	0.535	0.499
Biotech Sector	1 if firm name is associated to industries in the biotechnology sector, 0 otherwise.	Estimated from name	0.002	0.044
Ecommerce Sector	1 if firm name is associated to industries in the ecommerce sector, 0 otherwise.	Estimated from name	0.052	0.222
IT Sector	1 if firm name is associated to industries in the IT sector, 0 otherwise.	Estimated from name	0.022	0.146
Medical Dev. Sector	1 if firm name is associated to industries in the medical devices sector, 0 otherwise.	Estimated from name	0.028	0.166
Semiconductor Sector	1 if firm name is associated to industries in the semiconductor sector, 0 otherwise.	Estimated from name	0.000	0.020
Observations			29,961,838	

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	*Coefficient*	*Pseudo R2*
Short Name	1.569***	0.008
	(0.00934)	
Eponymous	0.222***	0.004
	(0.0207)	

Corporate Form Measures:

Variable	*Coefficient*	*Pseudo R2*
Corporation	3.693***	0.017
	(0.124)	
Delaware	38.31***	0.140
	(0.943)	

IP Measures:

Variable	*Coefficient*	*Pseudo R2*
Patent	141.0***	0.085
	(4.396)	
Trademark	86.68***	0.032
	(3.861)	

Industry Measures:

Variable	*Coefficient*	*Pseudo R2*
Local	0.273***	0.008
	(0.0139)	
Traded Resource Intensive	1.402***	0.001
	(0.0446)	
Traded	1.474***	0.002
	(0.0374)	
Biotech Sector	15.71***	0.006
	(1.163)	
Ecommerce Sector	1.839***	0.001
	(0.0773)	
IT Sector	5.221***	0.010
	(0.202)	
Medical Dev. Sector	3.036***	0.004
	(0.133)	
Semiconductor Sector	18.85***	0.002
	(2.608)	

Observations	18,2583,53

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. *Growth* is only defined for firms born in the cohorts of 1988 to 2008. 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. * p<0.05 ** p<0.01 *** p<0.001

	Preliminary Models			***Nowcasting Model (estimated up to real-time)***	***Full Information Model (2 year lag)***
	(1)	(2)	(3)	(4)	(5)
Corporate Governance Measures					
Corporation	5.293***			4.626***	3.788***
	(0.181)			(0.169)	(0.139)
Delaware	45.17***			35.92***	
	(1.279)			(1.045)	
Name-Based Measures					
Short Name		2.085***		1.994***	1.830***
		(0.0510)		(0.0430)	(0.0393)
Eponymous		0.165***		0.271***	0.300***
		(0.0155)		(0.0256)	(0.0283)
Intellectual Property Measures					
Patent			67.67***		
			(2.905)		
Trademark			13.70***		6.914***
			(0.999)		(0.430)
Patent - Delaware Interaction					
Patent Only					45.91***
					(3.018)
Delaware Only					28.88***
					(0.952)
Patent and Delaware					198.5***
					(10.14)
US CMP Cluster Dummies					
Local				0.632***	0.655***
				(0.0327)	(0.0343)
Traded Resource Intensive				1.271***	1.263***
				(0.0437)	(0.0445)
Traded				1.095**	1.223***
				(0.0318)	(0.0363)
US CMP High-Tech Clusters					
Biotechnology				3.448***	2.598***
				(0.276)	(0.232)
E-Commerce				1.148**	1.069
				(0.0511)	(0.0493)
IT				2.135***	1.781***
				(0.0996)	(0.0875)
Medical Devices				1.186**	0.965
				(0.0631)	(0.0539)
Semiconductors				2.914***	1.737**
				(0.439)	(0.292)
N	18,258,353	18,258,353	18,258,353	18,258,353	18,258,353
R-squared	0.193	0.050	0.123	0.212	0.242

TABLE 4

Robustness Models. State Fixed Effects and State-Specific Time Trends

Dependent Variable: 1[IPO or Acquisition in six years or less]

We repeat the main regression model of Table 3 but include year fixed effects, and state-specific time-trends, to evaluate the robustness of our findings. We perform other tests on the performance of our predictive model in our appendix. Robust standard errors in parenthesis. * p < .05 , ** p < .01, *** p < .001

	(1)	(2)	(3)
Corporate Governance Measures			
Cororation	2.599***	2.825***	2.676***
	(0.102)	(0.111)	(0.106)
Name-Based Measures			
Short Name	1.852***	1.874***	1.875***
	(0.0405)	(0.0420)	(0.0424)
Eponymous	0.294***	0.295***	0.296***
	(0.0277)	(0.0278)	(0.0279)
Intellectual Property Measures			
Trademark	7.260***	7.098***	7.242***
	(0.466)	(0.451)	(0.467)
Patent - Delaware Interaction			
Delaware Only	28.91***	30.01***	29.31***
	(0.947)	(0.982)	(0.963)
Patent Only	42.44***	43.35***	41.79***
	(2.824)	(2.891)	(2.790)
Patent and Delaware	239.1***	241.4***	245.3***
	(12.33)	(12.45)	(12.72)
US CMP Cluster Dummies			
Local	0.661***	0.662***	0.663***
	(0.0346)	(0.0347)	(0.0347)
Traded Resource Intensive	1.221***	1.218***	1.224***
	(0.0433)	(0.0432)	(0.0435)
Traded	1.194***	1.207***	1.195***
	(0.0355)	(0.0358)	(0.0355)
US CMP High-Tech Clusters			
Biotechnology	2.822***	2.775***	2.828***
	(0.260)	(0.251)	(0.261)
E-Commerce	1.008	1.000	1.001
	(0.0470)	(0.0465)	(0.0469)
IT	1.737***	1.738***	1.723***
	(0.0851)	(0.0852)	(0.0845)
Medical Devices	0.962	0.965	0.960
	(0.0541)	(0.0542)	(0.0542)
Semiconductors	1.709**	1.742***	1.712**
	(0.288)	(0.293)	(0.289)
Year FE	Yes	No	Yes
State FE	Yes	Yes	Yes
State Trends	No	Yes	Yes
N	18258353	18258353	18258353
pseudo R-sq	0.256	0.252	0.258

TABLE 5

Entrepreneurial Quality Models with High Employment Growth Outcomes

We develop models with the same regressor as our full information entrepreneurial quality model (Table 3, Column 5) but substitute high equity growth outcomes for high employment growth outcomes. Our outcome variable is 1 if a firm has high employment six years after founding and zero otherwise, at different thresholds. Employment measures are taken from the Infogroup USA panel data. We have a long-term project with the US Census to develop entrepreneurial quality estimates using continuous employment outcomes. Robust standard errors in parenthesis. * p < .05 , ** p < .01, *** p < .001

Dependent Variable	(1) Equity Growth (IPO or Acquisition)	(2) Employment > 500	(3) Employment > 1000
Corporate Governance Measures			
Corporation	3.473***	1.845***	1.566***
	(0.121)	(0.0763)	(0.107)
Name-Based Measures			
Short Name	2.388***	1.748***	1.591***
	(0.0695)	(0.0696)	(0.108)
Eponymous	0.297***	0.723***	0.848
	(0.0282)	(0.0630)	(0.124)
Intellectual Property Measures			
Trademark	6.766***	7.577***	6.682***
	(0.416)	(0.747)	(1.051)
Delaware Only	42.69***	25.88***	46.86***
	(2.731)	(2.580)	(6.884)
Patent Only	27.58***	9.365***	10.14***
	(0.896)	(0.453)	(0.845)
Patent and Delaware	184.8***	75.34***	119.6***
	(9.282)	(6.688)	(16.61)
US CMP Cluster Dummies			
Local	0.717***	1.057	1.145
	(0.0377)	(0.0588)	(0.108)
Traded Resource Intensive	1.265***	1.455***	1.277**
	(0.0445)	(0.0720)	(0.107)
Traded	1.297***	1.077	1.258**
	(0.0385)	(0.0454)	(0.0908)
Biotechnology	2.466***	0.982	0.175*
	(0.220)	(0.202)	(0.124)
E-Commerce	1.090	1.343***	1.111
	(0.0508)	(0.0878)	(0.130)
IT	1.771***	1.087	0.899
	(0.0869)	(0.0960)	(0.138)
Medical Devices	0.972	1.274**	1.432*
	(0.0543)	(0.108)	(0.205)
Semiconductors	1.761***	2.952***	2.494*
	(0.293)	(0.696)	(1.013)
N	17831208	14232544	14016581
pseudo R-sq	0.243	0.101	0.104

TABLE 6

Vector Autoregression Models (VAR) on the impact of changes in GDP Growth to Entrepreneurship

Notes: All models are run on a 27 observation time series representing each year observed in the data, from 1988 to 2014 (inclusive). VAR models are estimated through simultaneous equations, only the equation with GDP as a dependent variable is presented in the table. Three lag structure chosen as the optimal one using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). US Recession is a dummy variable equal to 1 if the year is 1992, 2001, 2008, or 2009. All regressions also pass VAR stability tests—the eigenvalues of all models lie within the unit circle. Standard errors in parenthesis. * p < .1 ** p < .05

	VAR		SVAR		SVAR: US Recession	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Ln(RECPI/GDP)	Ln(N/GDP)	Ln(RECPI/GDP)	Ln(N/GDP)	Ln(RECPI/GDP)	Ln(N/GDP)
Intersection ΔLn(GDP)(t) (Cholesky Decomposition)			0.022	0.007		
ΔLn(GDP)(t-1)	1.525*	-0.0885	3.95**	-0.835		
	(0.918)	(0.538)	(.98)	(.56)		
ΔLn(GDP)(t-2)			1.217	-0.747		
			(1.13)	(.56)		
ΔLn(GDP)(t-3)			-1.2	0.51		
			(.85)	(0.48)		
Intersection 1[US (Cholesky Decomposition)					-0.0486	-0.007
1[US Recession](t-1)					-0.152**	0.054**
					(.06)	(.023)
1[US Recession](t-2)					-0.033	0.053**
					(.06)	(.024)
1[US Recession](t-3)					-0.012	0.032*
					(.042)	(.019)
Ln(RECPI/GDP)(t-1)	0.739**	0.966**	0.042		0.369	
	(0.106)	(0.0360)	(.22)		(.29)	
Ln(RECPI/GDP)(t-2)			-0.05		0.141	
			(.232)		(.35)	
Ln(RECPI/GDP)(t-3)			0.539**		0.225	
			(.184)		(.26)	
Ln(N/GDP)(t-1)				1.33**		1.41**
				(.19)		(.19)
Ln(N/GDP)(t-2)				-0.125		-0.113
				(0.33)		(0.34)
Ln(N/GDP)(t-3)				-0.292		-0.365*
				(.206)		(.203)

TABLE 7

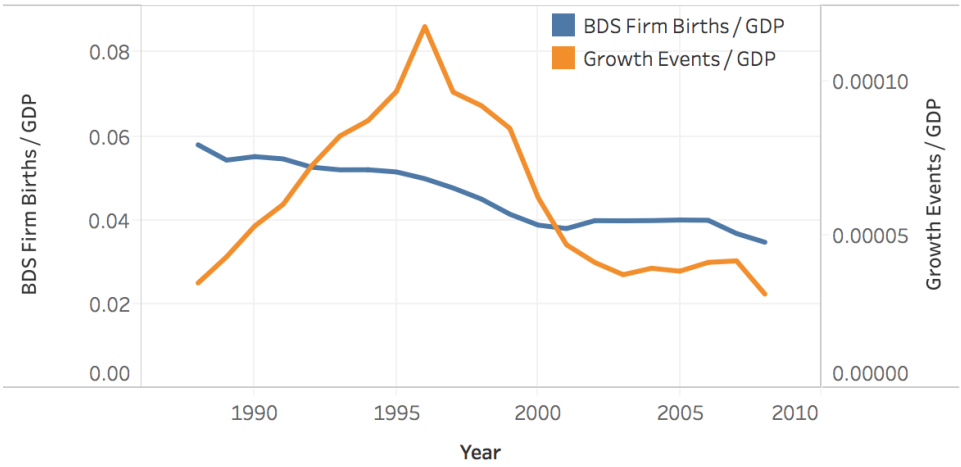
Relationship of MSA Entrepreneurship and MSA GDP Growth
209 Metropolitan Statistical Areas
Dependent Variable: Ln(GDP in 2013/GDP in 2003)

NOTES. MSA GDP is the Gross Metropolitan Product reported by the Bureau of Economic Analysis. MSA entrepreneurship measures are used as aggregates from 2001-2003 to control for year-specific variation. MSA Quantity is the total number of firms registered from 2001 to 2003 in the MSA. MSA RECPI is quality-adjusted quantity of firms. MSA GDP is in 2003. All MSAs in our sample for which we can document all firms confidently included, except for two outliers—Midland, TX, and Odessa, TX. US Census 2013 MSA definitions used. Robust standard errors in parenthesis. * p < .1 ** p < .05

	Unweighted Regressions				Weighted Regressions by MSA Share of Aggregate GDP	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(MSA RECPI)	0.00819*		0.0255**	0.0240*	0.0575**	0.0555**
	(0.00491)		(0.0128)	(0.0128)	(0.0176)	(0.0180)
Ln(MSA Entrepreneurial Quantity)		0.00540	-0.0247	-0.0319*	-0.0578**	-0.0638**
		(0.00601)	(0.0157)	(0.0177)	(0.0202)	(0.0215)
Ln(MSA GDP)				0.0123		0.00964
				(0.0126)		(0.0132)
N	209	209	209	209	209	209
R-sq	0.014	0.004	0.028	0.032	0.198	0.200

FIGURE 1

Panel A. Firm Births in Business Dynamics Statistics vs. Number of Growth Events per Cohort
34 US States (83% of US GDP)



Panel B. Firm Births in Business Dynamics Statistics vs. GDP Growth Over Next 5 Years
34 US States (83% of US GDP)

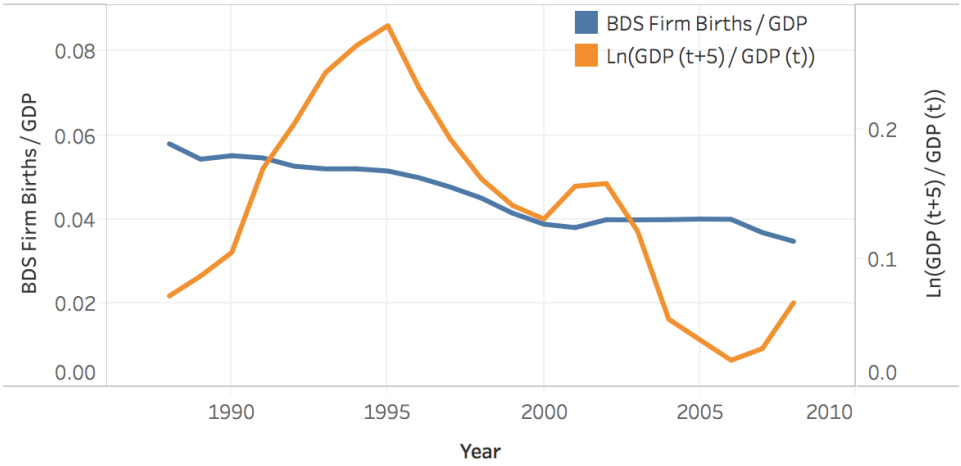
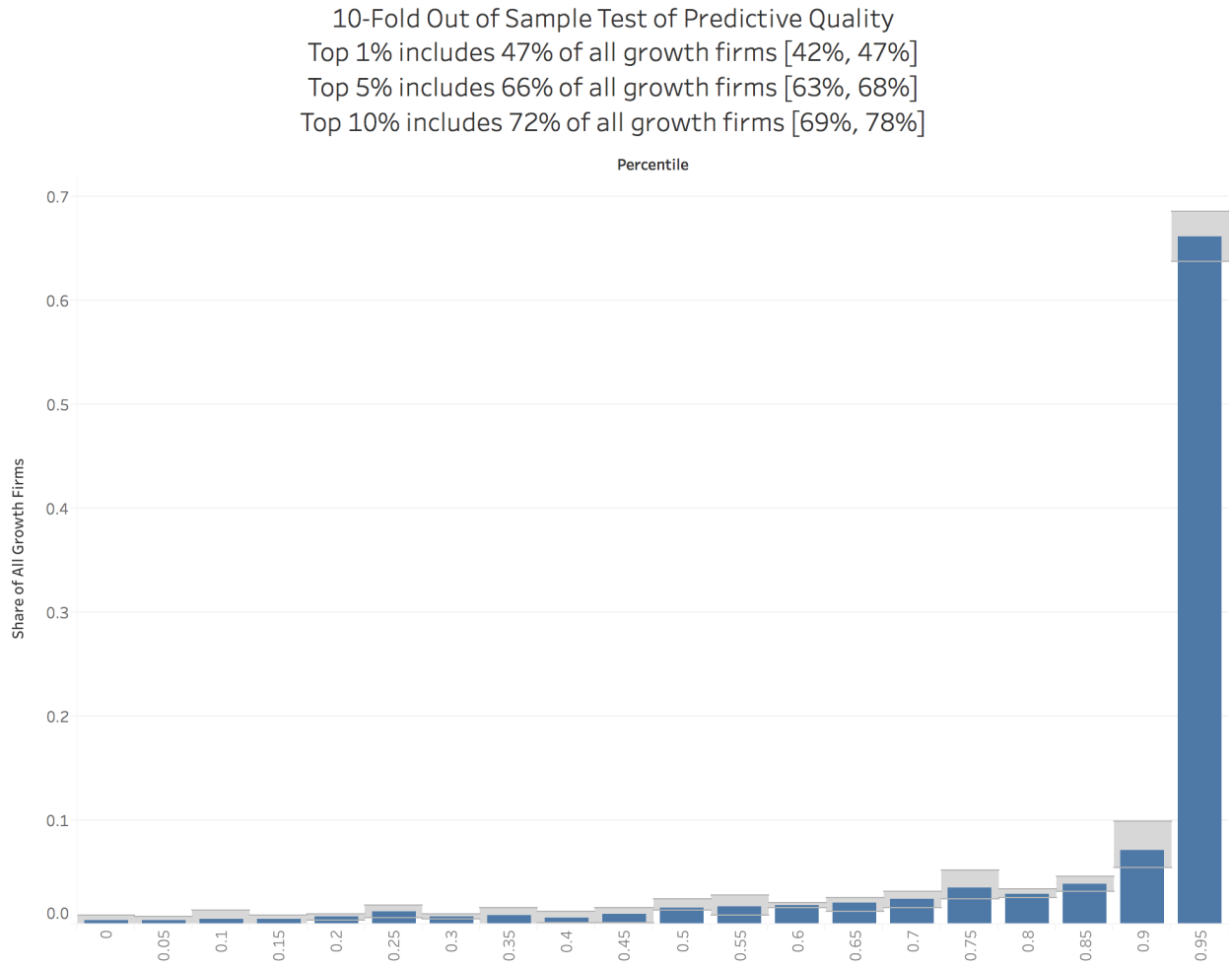
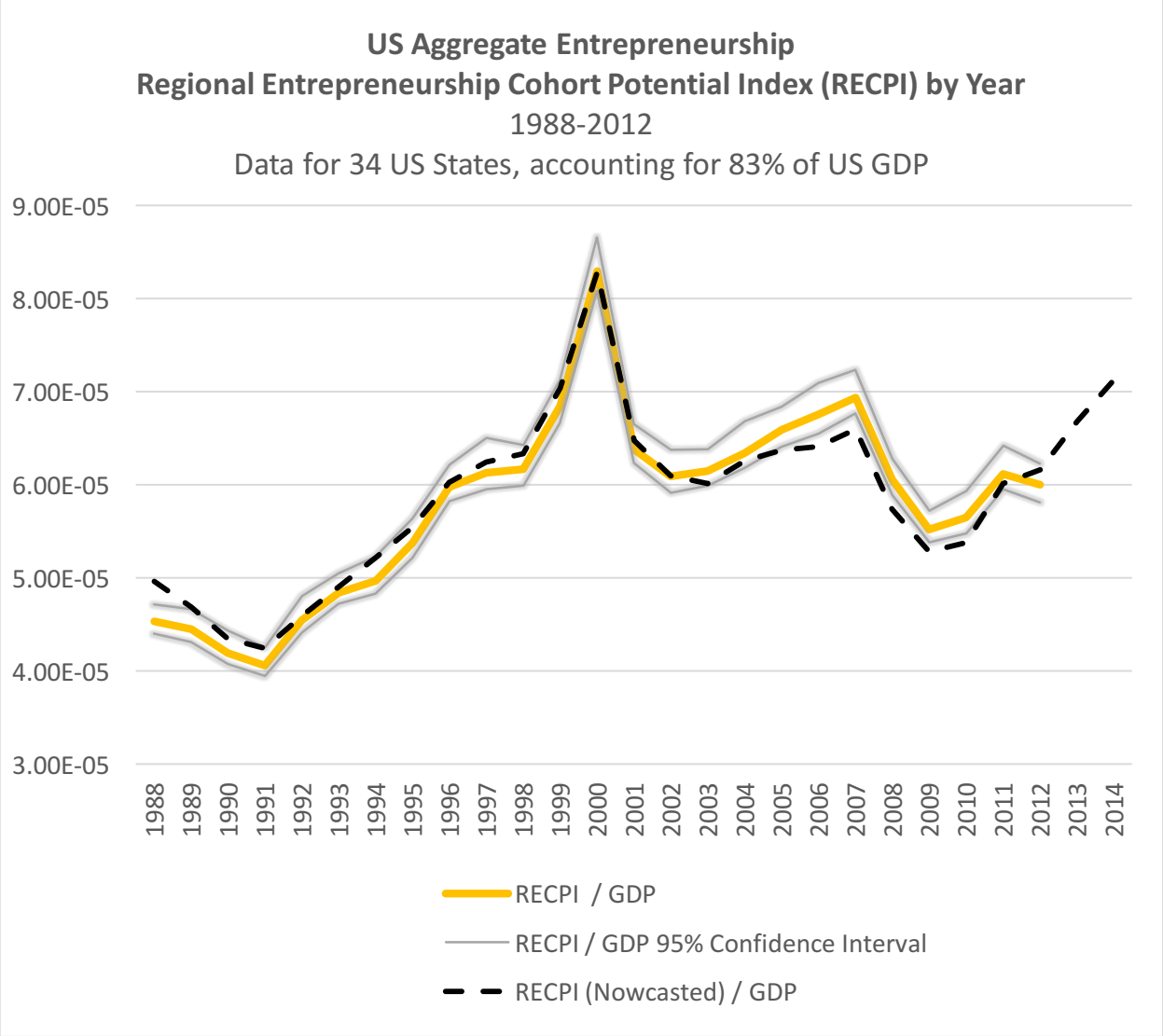


FIGURE 2



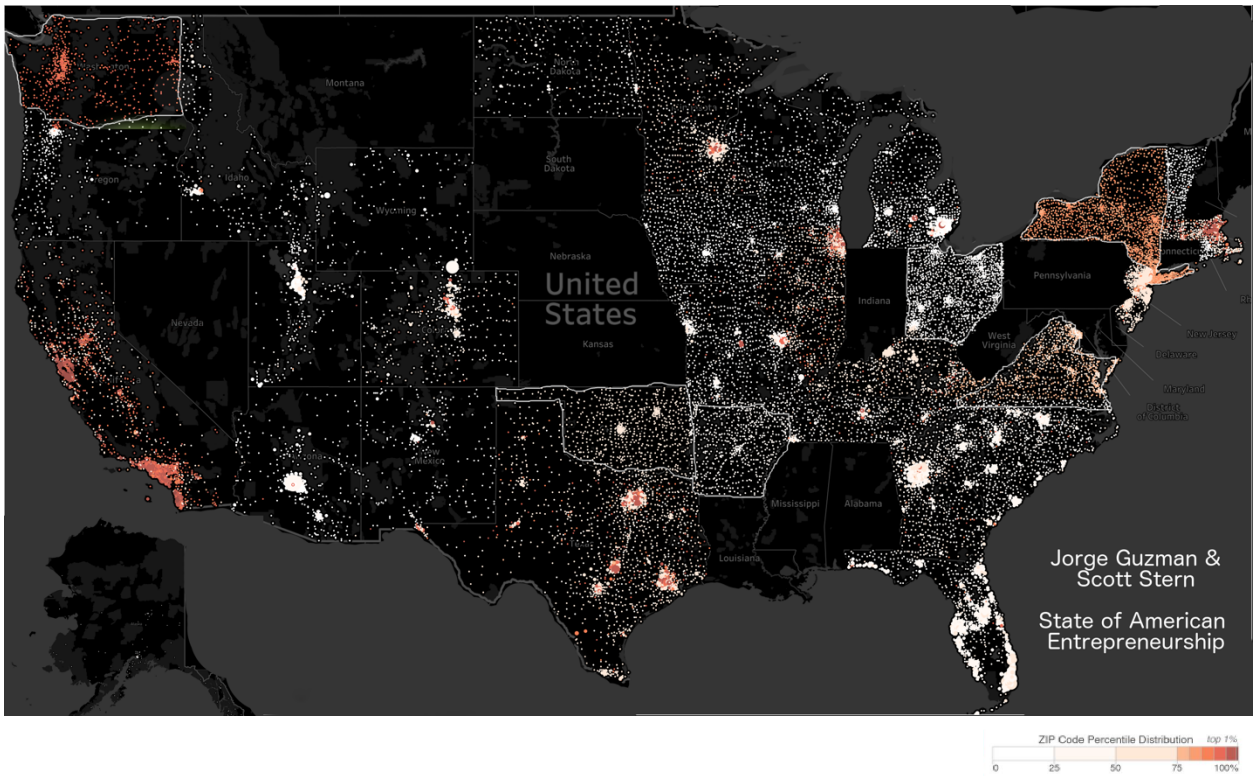
Notes: This Figure presents the results of an out of sample cross validation procedure performed on all firms born between 1988 and 2008 in our database. The goal of this procedure is to compares the predicted quality of firms with their realized outcomes, to assess the quality of these predictions. We use the standard machine learning process, a 10-fold cross validation. This process first separates the sample into 10 random samples, then, for each sample estimate the model with the available 90% and use the remaining 10% as the validation sample. The blue lines are the median value of these 10 tests, grey lines are the maximum and minimum.

FIGURE 3



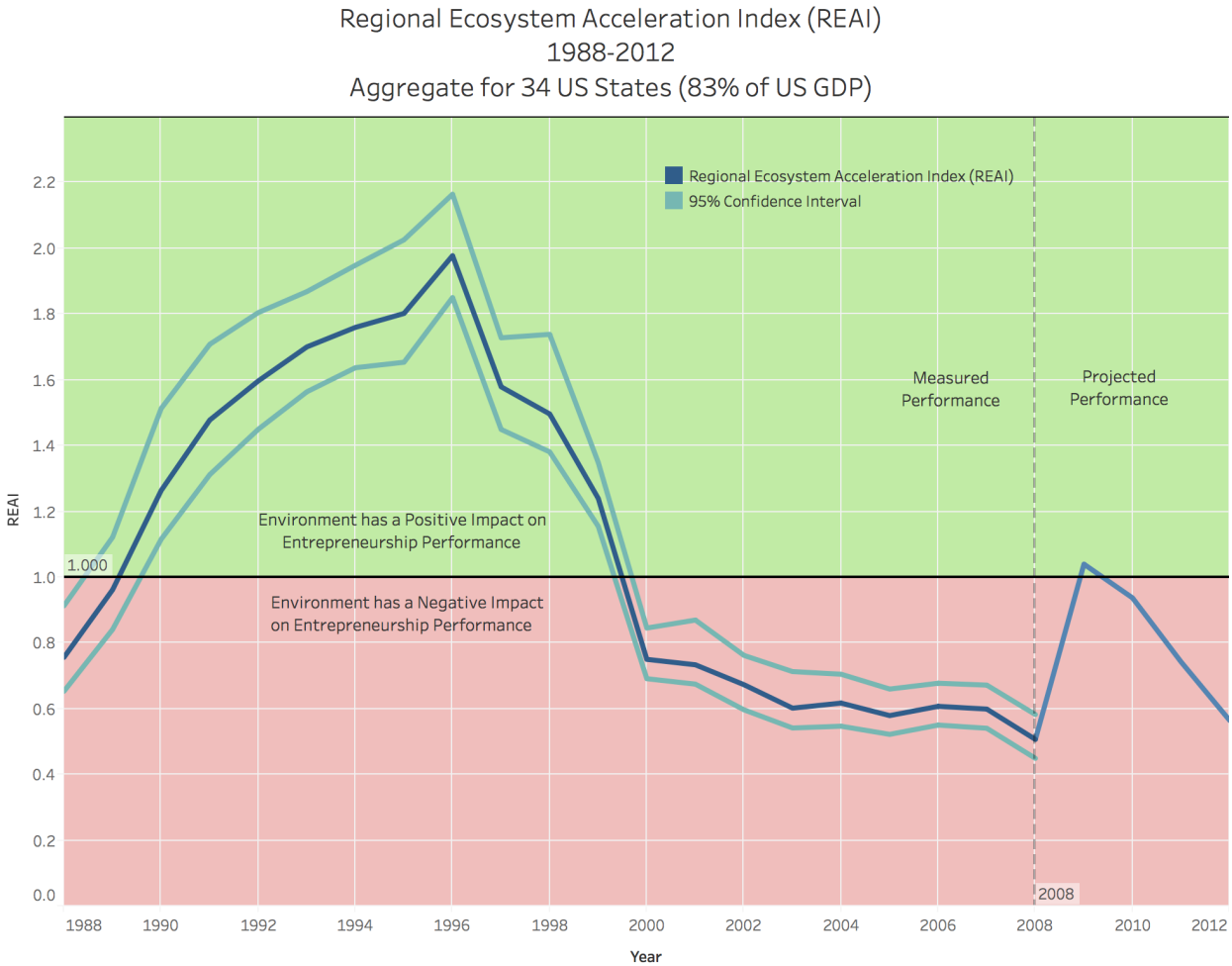
Notes: RECPI / GDP represents the total, quality adjusted, entrepreneurship production in a region after controlling for the size of the economy in that year. GDP is estimated using state GDP measures by the Bureau of Economic Analysis for the 34 states in the sample. 95% confidence interval is built using a Monte Carlo process with 100 random draws of the same size as the full sample of the data. RECPI / population shows a starker positive increase than RECPI / GDP, as GDP per capita has also increased through the time period represented.

FIGURE 4
THE STATE OF AMERICAN ENTREPRENEURSHIP



Notes: This map represents the quality and quantity of entrepreneurship in 2012 all 34 US states for which we have data in our sample. Data is presented by ZIP Code. The size of the point represents the quantity of firms and the color of the point represents the average quality of entrepreneurship in that ZIP Code (white is the lowest average quality and dark red the highest). The entrepreneurial quality model includes state fixed-effects to account for the institutional differences in firm registration. For six states, Washington, Oklahoma, Arkansas, Ohio, Virginia, and New York, our data does not allow us to track the precise ZIP Code in which the firm is located for all firms. In these cases, we simply distribute uniformly the number of new firms across all ZIP Codes and mark each ZIP Code with the state entrepreneurial quality. For ease of analysis, we also create a white border across these states. ZIP Code quality percentiles are defined within the *complete* distribution of ZIP Code - year observations.

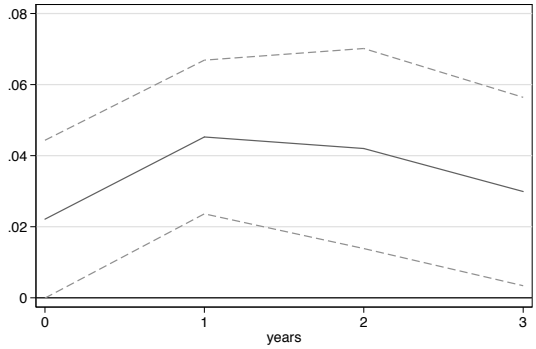
FIGURE 5



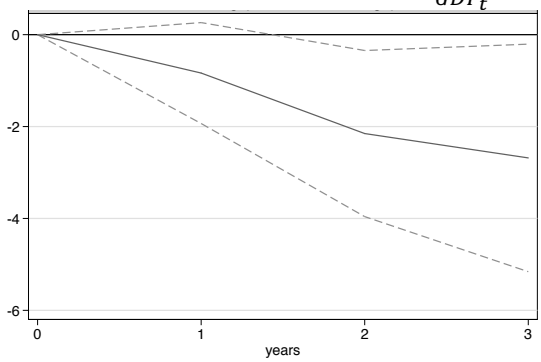
Notes: The Regional Entrepreneurship Acceleration Index (REAI) measures the realized performance of an entrepreneurial ecosystem relative to the expected potential of that ecosystem. It is defined as the number of growth events (IPO or acquisition within six years of founding) to occur from a cohort over the RECPI of that cohort. Confidence intervals are estimated through a Monte Carlo process drawing 100 random samples of size N. Projected performance is a preliminary estimate given the number of growth events that have occurred so far for that cohort.

FIGURE 6
EFFECT OF ECONOMIC CONDITIONS ON ENTREPRENEURSHIP
Impulse Response Functions

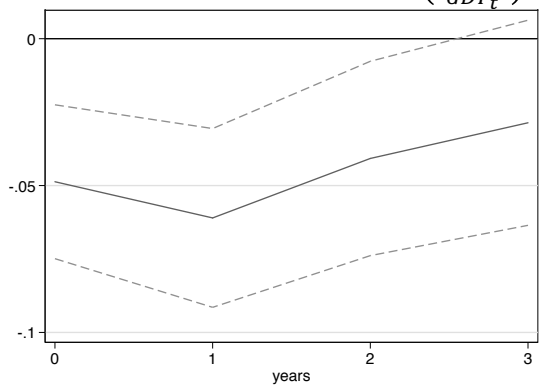
A. Effect of $Ln(\Delta GDP_t)$ on $Ln(\frac{RECPI_t}{GDP_t})$



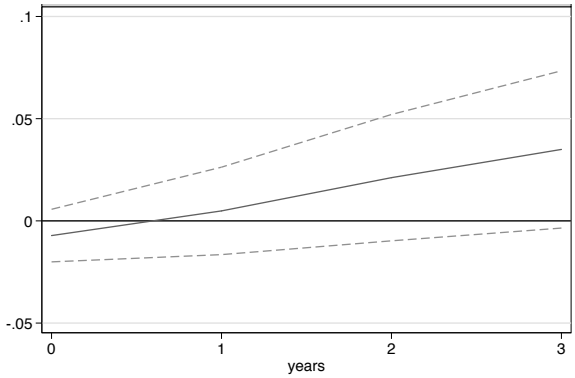
B. Effect of $Ln(\Delta GDP_t)$ on $Ln(\frac{Obs_t}{GDP_t})$



C. Effect of $1[US\_Recession_t]$ on $Ln\left(\frac{RECPI_t}{GDP_t}\right)$



D. Effect of $1[US\_Recession_t]$ on $Ln(\frac{Obs_t}{GDP_t})$



Notes: This figure reports the relationship of economic conditions to entrepreneurship production in the United States. ΔLn(GDP(t)) is the log change in US GDP between years t-1 and t. 1[US Recession(t)] is a dummy variable equal to 1 if the US was in a recession in that year. The years of recession are 1990, 2001, 2008, and 2009. Technically the 1990 recession ended until March of 1991, and the 2008 recession started in December of 2007, we do not include these years as the bulk of the year was not in recession. VAR models allow for lags from t-1 to t-n. The structural VAR also allows for a contemporaneous effect of GDP or Recessions on RECPI. GDP is estimated as the GDP only in the sample of interest. All lag structures are chosen as those that maximize the Akaike's Information Criterion (AIC) and the Schwarz's Bayesian Information Criterion (SBIC), which agree in all cases. All models pass a VAR stability test, with all eigenvalues within the unit circle. A Granger causality test rejects the null of no relationships ($p < 0.01$) for the ΔLn(GDP) models, and is marginally unable to reject it for the US Recession models ($p = 0.13$).

FIGURE 7A

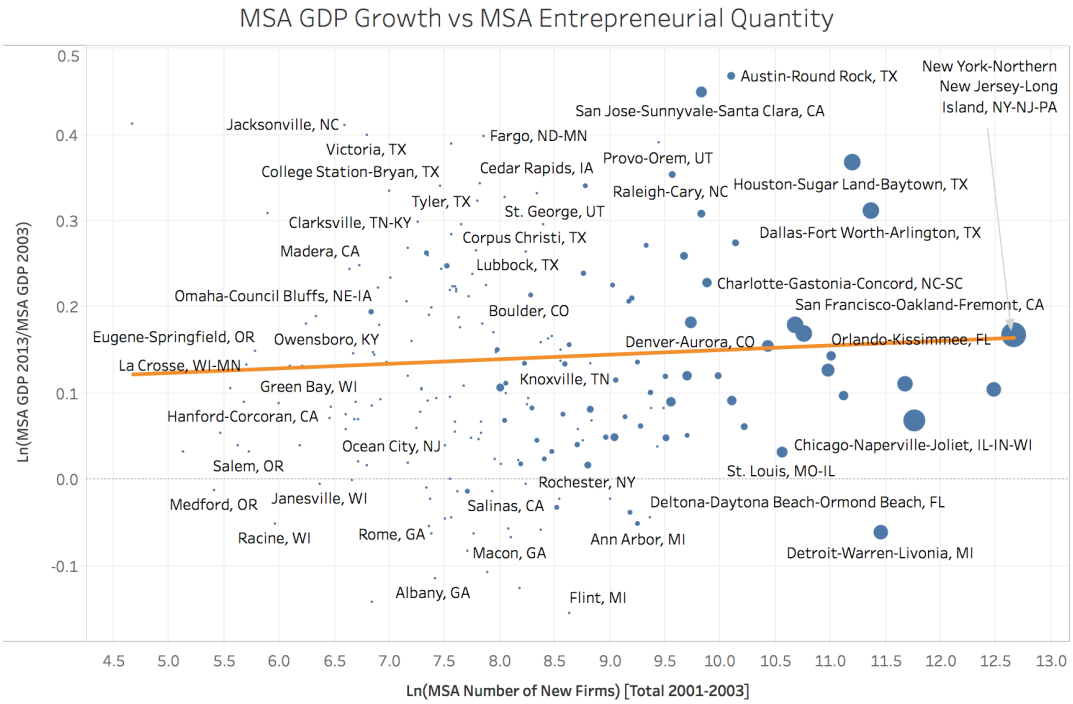
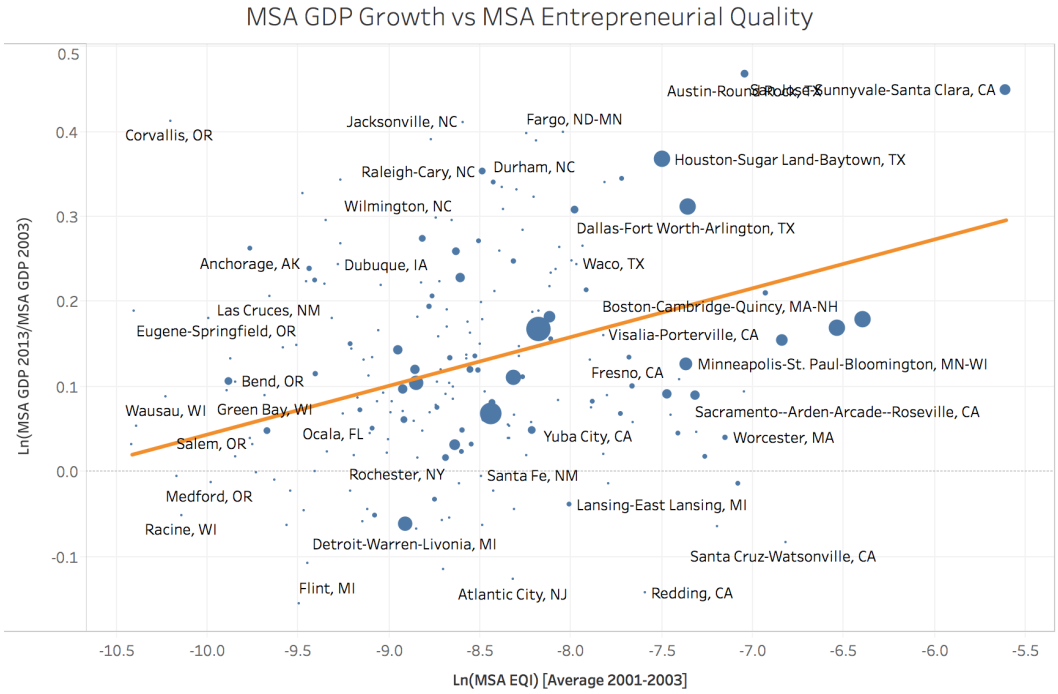
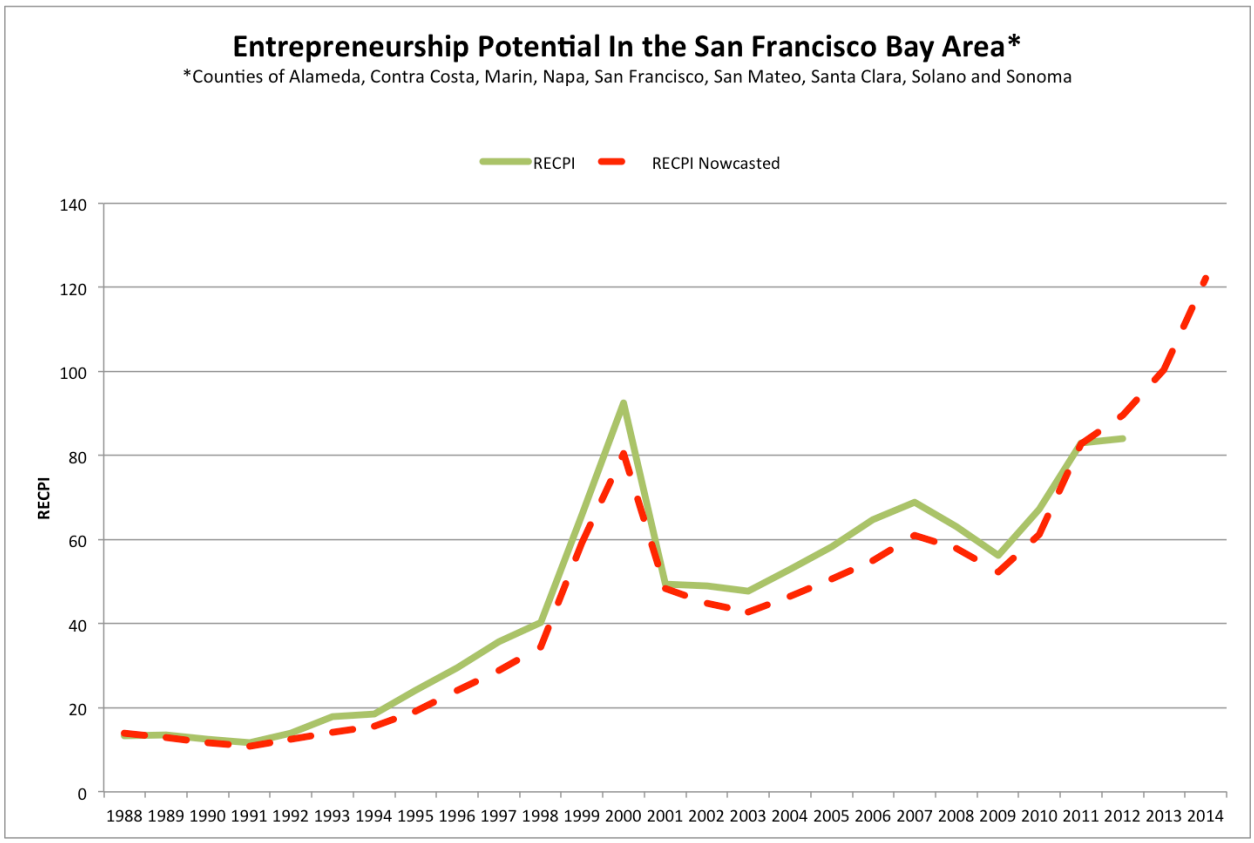


FIGURE 7B



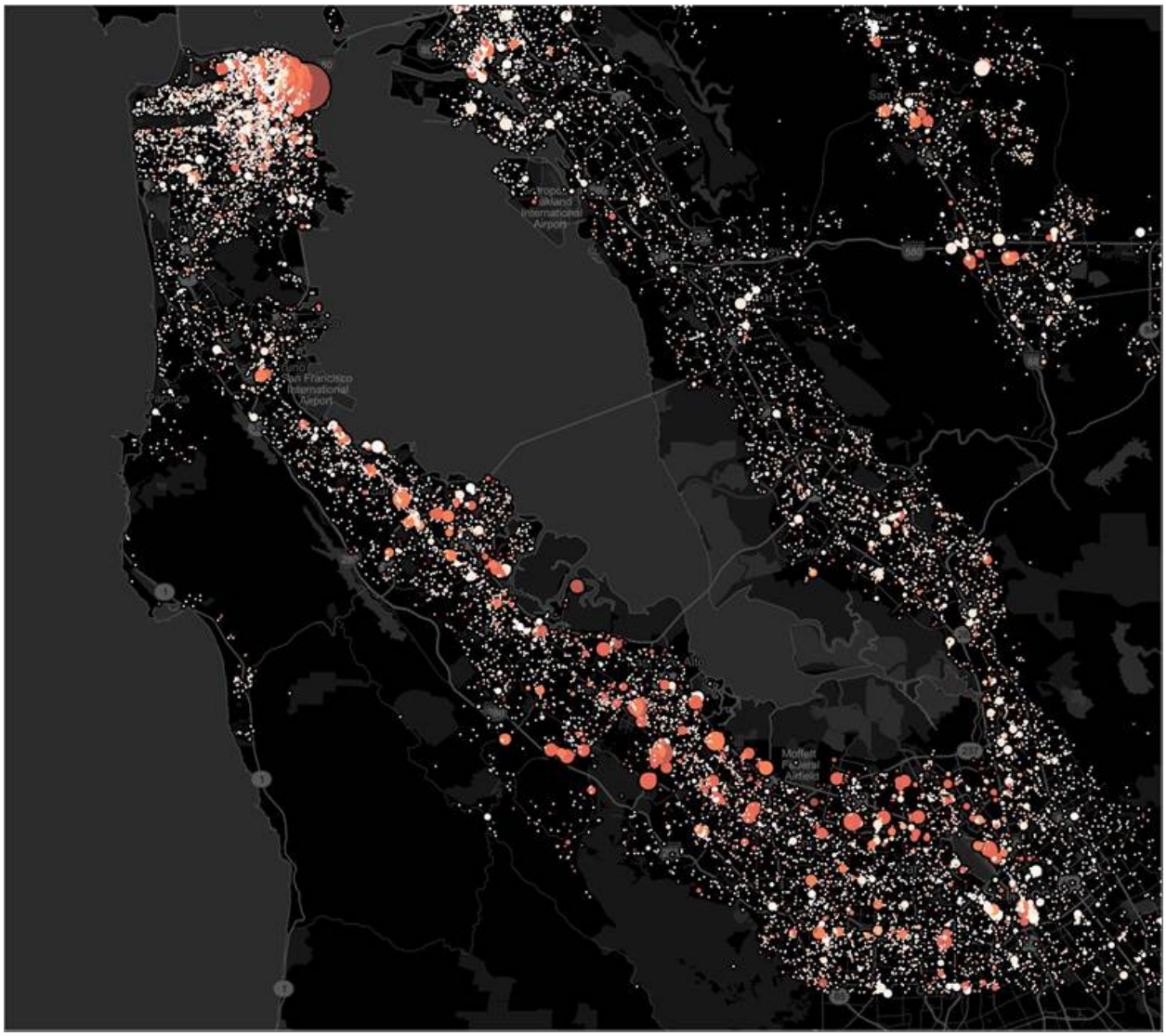
Note: All MSAs in our sample for which we could conclusively establish firm location included, except for two outliers (Midland, TX, and Odessa, TX). MSAs developed using firm ZIP Code and the 2013 US Census MSA Definitions. Size of point represents its GDP in 2003. Fitted line represents the fitted values of weighted regression by MSA GDP in 2003. MSA GDP is an equivalent of GDP estimated by the Bureau of Economic Analysis, formally called Gross Metropolitan Product.

FIGURE 8A



Notes: This figure shows RECPI for the San Francisco Bay Area (The combined counties of Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma) from 1988 to 2014.

FIGURE 8B



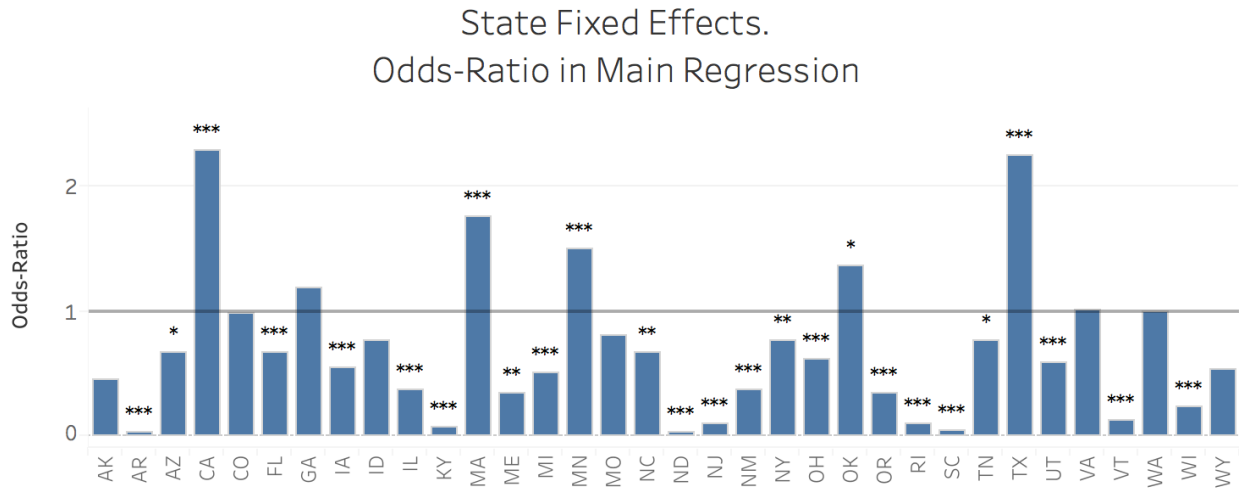
Notes: Figure represents all new firms founded in 2012 in the southern portion of the San Francisco Bay Area. Each point represents a location. The size of the point represents the number of firms and the color is the average quality of firms founded in that location.

APPENDIX A1

Data Coverage US States - Ranked by GDP

Rank in US GDP	*State*	*GDP*	*Share of GDP*
1	*California*	$2,287,021	13.0%
2	*Texas*	$1,602,584	9.1%
3	*New York*	$1,350,286	7.7%
4	*Florida*	$833,511	4.7%
5	*Illinois*	$742,407	4.2%
7	*Ohio*	$584,696	3.3%
8	*New Jersey*	$560,667	3.2%
9	*North Carolina*	$491,572	2.8%
10	*Georgia*	$472,423	2.7%
11	*Virginia*	$464,606	2.6%
12	*Massachusetts*	$462,748	2.6%
13	*Michigan*	$449,218	2.6%
14	*Washington*	$425,017	2.4%
17	*Minnesota*	$326,125	1.9%
18	*Colorado*	$309,721	1.8%
19	*Tennessee*	$296,602	1.7%
20	*Wisconsin*	$293,126	1.7%
21	*Arizona*	$288,924	1.6%
22	*Missouri*	$285,135	1.6%
25	*Oregon*	$229,241	1.3%
27	*Oklahoma*	$192,176	1.1%
28	*South Carolina*	$190,176	1.1%
29	*Kentucky*	$189,667	1.1%
30	*Iowa*	$174,512	1.0%
32	*Utah*	$148,017	0.8%
34	*Arkansas*	$129,745	0.7%
39	*New Mexico*	$95,310	0.5%
43	*Idaho*	$66,548	0.4%
45	*North Dakota*	$62,772	0.4%
46	*Alaska*	$60,542	0.3%
47	*Maine*	$56,163	0.3%
49	*Wyoming*	$48,538	0.3%
50	*Rhode Island*	$45,962	0.3%
52	*Vermont*	$30,723	0.2%
	Total GDP in Sample		
	Number of States	34	
	US GDP	$17,565,783	
	Share of GDP in Sample	**83%**	

FIGURE A1



Notes: This figure represents the fixed-effect coefficients for all state fixed-effects in the main predictive model of entrepreneurial quality (Table 3, Column 5). The bar represents the incidence rate ratio of this coefficient, while the stars above the bar are the significance of the coefficient. Including fixed-effects on a regression is a choice available to the researcher. Most of the coefficients are relatively close to 1 (no effect), and including or excluding them did not change any of our main results in a significant manner.