American Economic Journal: Economic Policy 2020, 12(4): 212–243
https://doi.org/10.1257/pol.20170498

The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014[†]

By Jorge Guzman and Scott Stern*

Assessing the state of American entrepreneurship requires not simply counting the quantity but also the initial quality of new ventures. Combining comprehensive business registries and predictive analytics, we present estimates of entrepreneurial quantity and quality from 1988 to 2014. Rather than a secular pattern of declining business dynamism, our quality-adjusted measures follow a cyclical pattern sensitive to economic and capital market conditions. Consistent with the role of investment cycles as a driver of high-growth entrepreneurship, our results highlight the role of economic and institutional conditions as a driver of both initial entrepreneurial quality and the scaling of new ventures over time. (*JEL* G24, G32, L25, L26, M13)

Over the past two decades, economists have made significant progress in advancing the measurement of entrepreneurship. The pioneering studies of Haltiwanger and various coauthors (Davis and Haltiwanger 1992; Davis, Haltiwanger, and Shuh 1996; Haltiwanger, Jarmin, and Miranda 2013; Decker et al. 2014) moved attention away from simply counting the density of *small and medium-sized firms* toward the measurement and growth dynamics of *young firms*. These studies established that a disproportionate share of new job creation is associated with new firms, and economic growth is grounded in business dynamics. A separate stream of research focusing on selective samples of high-performance entrepreneurial ventures and the institutions that surround them reinforces 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 that

*Guzman: Columbia Business School, Uris Hall, 7th Floor, 3022 Broadway, New York, NY 10027 (email: jorge.guzman@columbia.edu); Stern: MIT Sloan School of Management, 100 Main Street, E62-476, Cambridge, MA 02142, and NBER (email: sstern@mit.edu). Kate Ho was coeditor for this article. 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, the University of Toronto, the University of Virginia, the Kauffman Foundation New Entrepreneurial Growth Conference, and the NBER Pre-Conference on Entrepreneurship and Economic Growth, and to four anonymous referees. We also thank Open Corporates for providing data for New York and Michigan, and RJ Andrews for his development of the visualization approach. Sarah Andries, Jintao Chen, Ji Seok Kim, and Yupeng Liu 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.

[†]Go to https://doi.org/10.1257/pol.20170498 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

venture capital has a positive impact on aggregate income, employment, and new establishment formation.

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, b, c), researchers focused on venture capital and high-growth firms have documented a sizable 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 between the rate (relative to GDP) (US Bureau of Economic Analysis 2017) of firm births per year as measured by the Business Dynamics Statistics (US Census Bureau 2017) 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 initital public offering (IPO) or significant acquisition within 6 years of initial business registration) for the 32 states that will form the basis of our analysis. While the BDS shows a slow and steady decline of approximately 40 percent (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.[1] Moreover, 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.[2]

Simply put, alternative definitions of entrepreneurship suggest different assessments of the state of American entrepreneurship.

Not simply a matter of measurement, characterizing entrepreneurial quality by cohort allows for the empirical assessment of important economic and policy questions. For example, in line with the debt deflation theory suggested by Fisher (1933), Bernanke and Gertler (1989) suggest that high-quality entrepreneurship may be reduced during a recession due to structural financing constraints (while low-quality entrepreneurship may be unaffected). And this reduction in growth-oriented entrepreneurship can exacerbate a downturn through reduced business dynamism. Assessing this theoretical claim empirically requires the development of consistent measures for the quantity and quality of entrepreneurship at founding and observing

[1] 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, Decker et al. (2016) show that high growth in the high tech sector (as defined by a collection of NAICS codes) has followed a more cyclical pattern than the retail and services sectors, which account for the bulk of the decline in new firms.

[2] The challenge is fundamentally a 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, 2).

how these measures change at different points in the business cycle. From a policy perspective, fostering growth-oriented entrepreneurship may be fundamentally different than focusing on policies that enhance the environment for "Main Street" businesses (Aulet and Murray 2013, Mills and McCarthy 2016, Chatterji 2018), and so being able to differentiate between new ventures in terms of their growth potential can offer more targeted and effective entrepreneurship policy.

Building on Guzman and Stern (2015, 2017), this paper develops and implements a novel approach to the measurement of both the quantity and quality of entrepreneurship, which we then use to provide substantive insight into both theoretical and policy questions.[3] Our approach to measuring entrepreneurial quality combines three interrelated insights. First, a practical requirement for any growth-oriented entrepreneur is business registration (as a 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. Second, moving beyond simple counts of business registrants (Klapper, Amit, and Guillén 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, how it is named, and how the idea behind the business is protected. These start-up characteristics may reflect choices by founders who perceive their venture to have high potential. In other words, 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 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 lack thereof) has not yet had time to be observed.

We use this predictive analytics approach to propose three new statistics for the measurement of growth entrepreneurship: the Entrepreneurship Quality Index (EQI), the Regional Entrepreneurship Cohort Potential Index (RECPI), and the Regional Entrepreneurial Acceleration Index (REAI). EQI is a measure of *average quality* within any given group of firms and allows for the calculation of the probability of a growth outcome for a firm within a specified population of start-ups. RECPI multiplies EQI and the number of start-ups within a given geographical region (e.g., from a zip code or town to the entire 32-state coverage of our sample)

[3] 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 32 "representative" US states (comprising more than 80 percent of overall GDP) over 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).

and so is a measure of the quality-adjusted quantity of entrepreneurship. Whereas EQI compares entrepreneurial quality across different groups, RECPI allows the direct calculation of the expected number of growth outcomes from a given start-up cohort within a given regional boundary. 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 32 US states for the period 1988–2014. We document several key findings. 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.[4] 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 at a given quality level to realize 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 6 years of founding), these broad patterns of results also hold if one focuses on alternative definitions of equity growth (e.g., focusing only on IPOs) or alternative growth measures such as the realization of more than 500 employees within the first 6 years after founding.[5]

We use these measures to assess the long-standing theoretical debate regarding the relationship between entrepreneurship and the business cycle. While a long line of theoretical work emphasizes the potential for economic growth to stimulate and nurture the founding of new growth-oriented ventures, either by improving the balance sheet of investors (Fisher 1933, Bernanke and Gertler 1989, Carlstrom and Fuerst 1997, Rampini 2004) or the investor expectations of follow-on capital availability (Caballero, Farhi, and Hammour 2006; Nanda and Rhodes-Kropf 2014), others have emphasized the potential for a "cleansing effect" of recessions, whereby downturns instead enhance the potential of the firms that are founded ventures at that time (Schumpeter 1939, Cabarello and Hammour 1994). A key insight of this theoretical literature is that economic conditions shape the aggregate entrepreneurial potential of start-up cohorts and the overall incidence of high-growth start-ups but not the total number of start-ups founded for each cohort. We use our measurement

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

[5] 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 be conducted more naturally with a comprehensive administrative dataset such as the LBD.

of both quantity and initial entrepreneurial quality to test this hypothesis directly. Using a structural vector autoregression (SVAR) model, we provide the first evidence that quality-adjusted quantity (RECPI) is procyclical, while quantity is unrelated to economic conditions.

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 the form of business registration 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 that might allow for even more differentiation in quality or account 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 toward young firms (pioneered by Haltiwanger and coauthors) is enriched by accounting directly 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.[6] 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. 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 are increasing, there seems to be a reduction in the ability of companies to scale in a meaningful and systematic way.

Our approach holds promise for multiple areas of economics research and policy. For example, a predictive analytics approach to entrepreneurial quality allows for the assessment of the relative importance of passive versus proactive growth firms in the overall process of firm growth and the specific role of venture capital in enabling the growth of firms with high initial quality (Catalini, Guzman, and Stern 2019). In addition, it is possible to use our approach to examine the role of gender differences among founders in the process of attracting venture capital and overall firm growth (Guzman and Kacperczyk 2019). And it is possible to use this approach (which uses firm choices regardless of location) to assess the role of location (e.g., whether the firm is in Silicon Valley) in facilitating firm growth (Guzman and Stern 2015), shaping the incentives to migrate between locations (Guzman 2019), and in assessing the role of local institutions (such as universities) and policies (such as tax) in shaping

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

the process of firm founding and growth (Tartari and Stern 2018; Fazio, Guzman, and Stern 2019).

The rest of this paper is organized as follows. Section I provides an overview of entrepreneurial quality in economics and briefly outlines our theoretical intuition. Section II explains our methodology and Section III our dataset and the estimation of entrepreneurial quality for our sample. Sections IV and V describe the variation in our key statistics across geography and time. Section VI compares the relationship of our index to an alternative measure of economic growth using employment outcomes. In Section VII, we test empirically the relationship between GDP growth and changes in national entrepreneurship. Section VIII concludes.

I. Entrepreneurial Quality: Do Initial Differences Matter?

Economists have long sought to understand the role of firm-specific characteristics in industry dynamics. Gibrat (1931) provides the foundational benchmark in this area: Gibrat's Law proposes that the growth rate of firms (and the variance in that growth rate) is independent of firm size (Sutton 1997). Despite broad patterns consistent with Gibrat's Law, a large literature beginning with Mansfield (1962) instead emphasizes deviations from proportional growth. This literature first emphasized that smaller firms have both higher growth rates and lower probabilities of survival (Mansfield 1962, Acs and Audretsch 1988, among others), but over time additional research suggested that younger firms also had high average growth rates and lower survival probabilities (Evans 1987; Dunne, Roberts, and Samuelson 1988).[7]

Davis and Haltiwanger (1992) clarified this empirical debate by considering simultaneously the role of size and age and developing systematic evidence 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; Haltiwanger, Jarmin, and Miranda 2013; Akcigit and Kerr 2018). Building on these studies, Decker et al. (2014) extend this approach to document an overall decline in the rate of new firms that have 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. Using surveys and aggregate economic comparisons, some have suggested that these differences in growth are accounted for by underlying differences in the firms themselves (Hurst and Pugsley 2011, Kaplan and Lerner 2010, Schoar 2010). Yet, beyond broad industry effects, systematic studies of firm dynamics have yet to incorporate such ex ante differences in a way that ties closely to the highly skewed ex post distribution of firm growth.

[7]Not simply a set of empirical regularities, these findings formed the foundations for important theoretical work, notably Jovanovic (1982) and subsequent formal models of firm and industry dynamics (Ericson and Pakes 1995, Klepper 1996, Hopenhayn 1992, Klette and Kortum 2004).

Accounting for this skew requires confronting a measurement quandary: at the time that a company is founded, one cannot observe whether that particular firm will experience a skewed growth outcome (or not). This challenge is fundamental, since entrepreneurship involves a high level of uncertainty and luck. And some outsized successes certainly result from unlikely origins. Ben and Jerry's, for example, was founded with the intention to be a one-store, homemade ice cream shop. With that said, many of the most successful firms in the economy were founded with a strong growth orientation. For example, Jeff Bezos founded Amazon with the intention of first creating the "world's biggest bookstore" in order to take advantage of the nascent potential of electronic commerce (Stone 2013). To the extent that the new firms that ultimately contribute to the skew are drawn disproportionately from firms with significant growth ambitions and underlying potential at their time of founding, identifying these growth-oriented firms can contribute significantly to the understanding of firm dynamics.

Our key insight is that while it may be difficult to identify the potential for growth based on traditional economic metrics (e.g., profits during the first of operations), the founders themselves likely have information about the underlying quality of their idea and their personal level of ambition and make choices at the time of founding consistent with their objectives and potential for growth. Specifically, we can take advantage of the fact that entrepreneurs who assess the underlying quality of their venture to be higher are more likely to make choices that result in "digital signatures" associated with growth-oriented start-up firms, and that firms with these resulting digital signatures are themselves more likely to grow. In other words, we can map the realized performance of start-ups to the early-stage choices of founders. By mapping the relationship between growth outcomes and these founder choices, we are able to form an estimate of entrepreneurial quality at founding.

To understand this intuition, consider a simple model where all new firms have an underlying quality level q (e.g., the underlying value of the idea and 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 $H = \{h_1, \ldots, h_N\}$, 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 underlying quality, she is able to observe both the corporate governance choice bundle H^* as well as growth outcomes g. The proof in online Appendix B demonstrates that the 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, importantly that, this mapping is a monotonically increasing function of q.

II. The Measurement of Entrepreneurial Quality and Performance

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 $\hat{\theta}$ 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 creates a public record of all registered companies, allowing us to form a population sample of entrepreneurs "at risk" of growth at a similar (and foundational) stage. 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, 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. For firm i born in region r at time t with start-up characteristics $H_{i,r,t}$ and growth outcome $g_{i,r,t+s}$, we estimate

$$\theta_{i,r,t} = \Pr\left(g_{i,r,t+s} | H_{i,r,t}\right) = f\left(\alpha + \beta H_{i,r,t}\right). \tag{1}$$

This model allows us to predict quality as the probability of achieving a growth outcome given start-up characteristics at founding and so estimate entrepreneurial quality as $\hat{\theta}_{i,r,t}$. As long as the process by which start-up characteristics map to growth remains 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).

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.

A. *The Entrepreneurial Quality Index*

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* as an aggregate of quality at the region-year level by simply estimating the average of $\theta_{i,r,t}$ over that region:

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

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.

B. *The Regional Entrepreneurship Cohort Potential Index*

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

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

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

C. *The Regional Entrepreneurial 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:

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

A value of REAI above 1 indicates a region-cohort that realizes a greater than expected number of growth events (and a value below 1 indicates underperformance 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.

III. Data and Entrepreneurial Quality Estimation

Our analysis leverages business registration records, a potentially rich and systematic dataset for the study of entrepreneurship. Business registration records are public records created when an individual registers a new business as a corporation, LLC, or partnership. Online Appendix C of the Supplementary Materials in this paper provides a rich and detailed overview of this dataset, as do the Data Appendixes in our prior work (Guzman and Stern 2015, 2017).

We focus on 32 US states from 1988 to 2014 (see online Appendix C for a list), collected through the Start-up Cartography Project (Andrews et al. 2019). 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 equivalent) 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.

Concretely, our analysis draws on the complete population of firms satisfying one of the following conditions: (i) a for-profit firm in the local jurisdiction or (ii) 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 nonprofit organizations as well as companies whose primary location is not in the state. The resulting dataset contains 27,976,477 observations.[8] For each observation, we construct variables related to (i) a growth outcome for each start-up, (ii) start-up characteristics based on business registration observables, and (iii) 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 online Data Appendix.

[8]The number of firms founded in our sample is substantially higher than in the US Census Business Dynamics Statistics (US Census Bureau 2017), done from tax records. For example, for Massachusetts in the period 2003–2012, the BDS records an average of 8,615 new firms per year, and we record an average of 24,090 firm registrations. We have yet to explore the reasons for this difference. However, we expect that it may be explained in part by the: (i) partnerships and LLCs that do not have income during the year do not file a tax return and are thus not included in the BDS and (ii) firms that have zero employees and thus are not included in the BDS.

Growth.—The growth outcome utilized in this paper, *Growth*, is a dummy variable equal to 1 if the start-up achieves an IPO or is acquired at a meaningful positive valuation within six years of registration[9] as reported in the Thomson Reuters SDC database (Refinitiv 2018).[10] During the period of 1998 to 2008, we identify 13,406 firms that achieved growth, of which 1,378 are IPOs and 12,028 are acquisitions, representing 0.07 percent 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: (i) measures based on business registration data observable in the registration record itself and (ii) measures based on external indicators of start-up quality that are observable at or near the time of business registration.

A. *Measures Based on Business Registration Observables*

We construct 12 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.[11] 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 1 if the entire firm name has three or fewer words and 0 otherwise.[12]

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

[9] 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.

[10] Although the coverage of IPOs is likely to be nearly comprehensive, the SDC dataset 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 percent accuracy; Netter, Stegemoller, and Wintoki (2011) perform a similar review. While we know these data not to be perfect, we believe the data 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.

[11] Belenzon et al. (2014) and Belenzon, Chatterji, and Daley (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.

[12] 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.").

identified as local (*Local*), or traded (*Traded*), or traded within resource-intensive industries (*Traded Resource-Intensive*). The other five industry groups are narrowly defined high-technology industries that could be expected to have high growth, including whether the firm is associated with biotechnology (*Biotech Sector*), e-commerce (*E-Commerce*), other information technology (*IT Sector*), medical devices (*Medical Dev. Sector*), or semiconductors (*Semiconductor Sector*).

B. *Measures Based on External Observables*

We construct two measures related to start-up quality based on intellectual property data sources from the US Patent and Trademark Office (US Patent and Trademark Office 2018a, b, c). *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 summary statistics and sources. 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 (online Appendix C).

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.[13]

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 with a 3-times increase in the probability of growth and having an eponymous name with a 70 percent *lower* probability of growth. Corporate form measures are also significant. Corporations are 3.4 times more likely to grow, and firms registered under Delaware jurisdiction (instead of the local jurisdiction) are 24 times more likely to grow. These magnitudes are economically important and have strong

[13] 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. 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. We have also undertaken exploratory work investigating a nonparametric approach involving unstructured interactions of start-up characteristics. The results from such an exercise result in an even more skewed distribution of estimated entrepreneurial quality.

Table 1—Summary Statistics (1988–2014)

	Source	Mean	SD
Outcome variable			
Growth	SDC platinum	0.00071	0.02672
Corporate form observables			
Corporation	Bus. reg. records	0.48	0.50
Delaware	Bus. reg. records	0.024	0.153
Name-based observables			
Short name	Bus. reg. records	0.46	0.50
Eponymous	Bus. reg. records	0.0707	0.2563
Intellectual property observables			
Patent	USPTO	0.0019	0.0439
Trademark	USPTO	0.0014	0.0374
Industry measures (*US CMP clusters*)(2)			
Local	Estimated from name	0.19	0.39
Traded (3)	Estimated from name	0.537	0.499
Traded resource int.	Estimated from name	0.133	0.339
Industry measures (*US CMP high-tech clusters*)			
Biotech sector	Estimated from name	0.002	0.044
E-commerce sector	Estimated from name	0.050	0.218
IT sector	Estimated from name	0.022	0.147
Medical dev. sector	Estimated from name	0.028	0.166
Semiconductor sector	Estimated from name	0.000	0.020
Observations		27,976,477	

Notes: (2) US CMP cluster dummies are estimated by using a sample of 10 million 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 the online Appendix. (3) Note that there are also firms that we cannot associate with local nor traded industries.

Table 2—Logit Univariate Regressions

Variable	Coefficient	Pseudo R^2	Variable	Coefficient	Pseudo R^2
Firm name measures:			*Industry measures* (*US CMP clusters*):		
Short name	3.147	0.018	Local	0.206	0.011
	(0.0612)			(0.00848)	
Eponymous	0.299	0.003	Traded resource-intensive	0.952	0.000
	(0.0168)			(0.0243)	
			Traded	1.208	0.001
Corporate form measures:				(0.0212)	
Corporation	3.375	0.016	Biotech sector	12.22	0.004
	(0.0769)			(0.723)	
Delaware	23.72	0.088	E-commerce sector	1.823	0.002
	(0.427)			(0.542)	
			IT sector	5.463	0.012
IP measures:				(0.146)	
Patent	88.50	0.059	Medical dev. sector	3.486	0.006
	(2.225)			(0.102)	
Trademark	45.30	0.016	Semiconductor sector	12.61	0.001
	(1.882)			(1.517)	
Observations	18,764,856				

Notes: Logit univariate regressions of *Growth* (IPO or acquisition within six years) with each of the observables we develop for our dataset. Incidence rate ratios reported; standard errors are in parentheses.

explanatory power—the pseudo R^2 of a Delaware binary measure alone is 0.09—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 90 times more likely to grow, while firms with a trademark are 45 times more likely to grow. Finally, the set of US CMP cluster dummies, implied from firm name, is also informative. For example, firms whose names are associated with local industries (e.g., "Taqueria") are 79 percent less likely to grow, while firms whose names are associated with the biotechnology sector are 12 times more likely to grow. These coefficients highlight the value of early firm name choices as indicators of firm intentions and signals of a firm's relationship to an industry.

It is of course important to caution against causal interpretations of these findings (and our subsequent regression estimates). 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.[14] 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. We begin in the first three specifications by evaluating the joint role of related groups of measures.[15] Column 1 investigates the core corporate governance measures, indicating that corporations are 4.1 times more likely to grow and Delaware firms are 23 times more likely to grow.[16] Interestingly, both of these coefficients are actually larger than the odds ratio in the univariate analysis. In column 2, we focus on two measures based on firm name: firms with a short name are 3 times more likely to grow, while eponymous firms are 78 percent *less* likely to grow. Finally, in column 3, we study the relationship of intellectual property measures to *Growth*. Firms with a patent are 50 times more likely to grow, and firms with a trademark are 8 times more likely to grow.

We then estimate our core predictive analytics models in column 4 and column 5 by combining these measures alongside industry and sector controls (i.e., firm names indicating whether the firm is in a local versus traded industry or associated with a particular industry cluster). Our first specification (column 4) uses only business registration observables. The coefficients associated with each of the business

[14] While it is possible the firms might "game" the algorithm by selecting into signals of high quality (e.g., changing their name), 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 yearly expense (around $1,000) to maintain a registration in Delaware. As well, firms using names to signal that they serve a local customer base (e.g., "Taqueria") are unlikely to change their names in ways that affect their ability to attract customers. Finally, if firms with low underlying quality did choose to invest in signals associated with high quality, that would undermine the empirical correlation between start-up characteristics and firm growth.

[15] 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. All of our core findings are robust to the inclusion or exclusion of state fixed effects.

[16] Since these are incidence-rate ratios (odds-ratios), the joint coefficients can be interpreted multiplicatively: Delaware corporations are 94.3 times more likely to grow ($23 \times 4.1 = 94.3$).

Table 3—Growth Predictive Model—Logit Regression on IPO or Acquisition within Six Years

	Preliminary models			Nowcasting (up to real time)	Full (2-year lag)
	(1)	(2)	(3)	(4)	(5)
Corporate governance measures					
Corporation	4.070			3.276	3.061
	(0.0975)			(0.0788)	(0.0739)
Delaware	23.28			18.22	
	(0.451)			(0.363)	
Name-based measures					
Short name		2.815		2.487	2.263
		(0.0546)		(0.0491)	(0.0456)
Eponymous		0.218		0.296	0.315
		(0.0123)		(0.0168)	(0.0179)
Intellectual property measures					
Patent			49.57		
			(1.572)		
Trademark			7.772		3.964
			(0.484)		(0.219)
Patent–Delaware interaction					
Patent only					22.77
					(1.059)
Delaware only					15.18
					(0.335)
Patent and Delaware					84.08
					(3.320)
US CMP clusters				Yes	Yes
US CMP high-tech clusters				Yes	Yes
Observations	18,764,856	18,764,856	18,764,856	18,764,856	18,764,856
R^2	0.135	0.057	0.094	0.163	0.187

Notes: 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 six 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 are in parentheses.

registration measures are roughly equivalent, though the impact of each individual predictor is slightly attenuated.[17] We then extend this specification in column 5 to include observables associated with early-stage milestones related to intellectual property. While the coefficients on the business registration observables remain similar (though once again slightly reduced in magnitude), each of the intellectual property observables is highly predictive. Given the high correlation between *Delaware* and *Patent*, we separately allow for the estimation of firms with a patent and no

[17] Online Appendix Table A1 presents the complete set of coefficient estimates for the US CMP clusters and US CMP high-tech cluster dummy variables. Briefly, as indicated in (A1-2), firms whose names indicate inclusion in a local industry (such as "restaurant," "realtor," etc.) are 58 percent less likely to grow, firms associated with traded industries are not significant, and firms specifically associated with resource-intensive traded industries are 12 percent less likely to grow. Names associated with specific high-technology sectors are also associated with growth: firms related to biotechnology are 3 times more likely to grow, firms associated with e-commerce are 44 percent more likely to grow, firms associated with IT 2.5 times, firms associated with medical devices 54 percent, and firms associated with semiconductors 2.3 times more likely to grow.

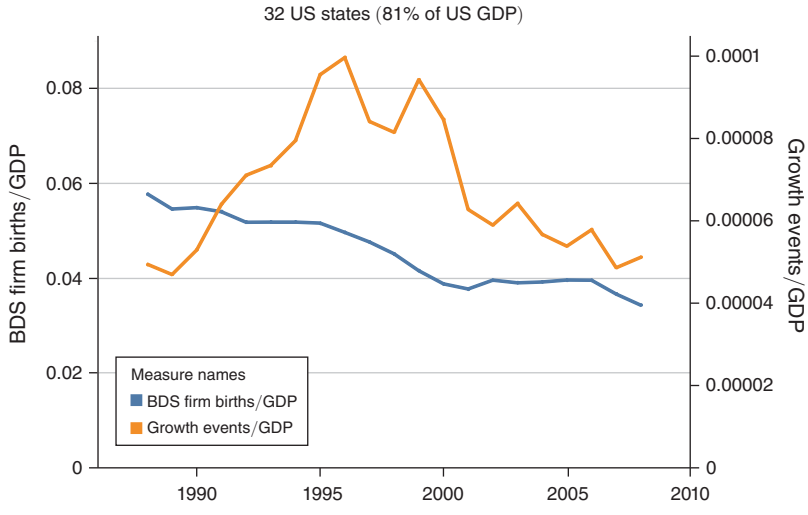

FIGURE 1. FIRM BIRTHS IN BUSINESS DYNAMICS STATISTICS VERSUS NUMBER OF GROWTH EVENTS PER COHORT

Delaware jurisdiction, firms with a Delaware jurisdiction and no patent, and firms with both.[18] In particular, receiving a patent is associated with a 23-times increase in the likelihood of growth for non-Delaware firms, and the combination of Delaware registration and patenting is associated with an 84-times increase in the likelihood of growth (simply registering in Delaware without a patent is associated with only a 15-times increase in the growth probability). Finally, firms applying successfully for a trademark in their first year after business registration are associated with a four-times increase in the probability of growth.[19]

These two models offer a trade-off. On the one hand, the "richer" specification of the model in Table 3, column 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). On the other hand, while including a more informative set of regressors, the model in Table 3, column 5 is not as timely as the model in Table 3, column 4. Indeed, specifications that rely exclusively on information encoded within the business registration record can be calculated on a near real-time basis and so provide the timeliest index for policymakers and other analysts. We will calculate indices based on both specifications; while our main historical analyses will be based off the results from the model in column 5, the model in column 4 can be used to provide our best estimate of changes in the last few years. We use the term

[18] 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* × *Patent* interaction rather than providing the effect directly.

[19] 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.

nowcasting in reference to the estimates related to Table 3, column 4 and refer to Table 3, column 5 as the "full information" model (Scott and Varian 2015).

D. *Prediction Quality and Robustness*

In Figure 2, we evaluate the predictive quality of our estimates by undertaking a tenfold cross-validation test (Witten and Frank 2005)[20] 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 percent of our estimated growth probability distribution ranges from 51 percent to 54 percent, with an average of 53 percent. The share of growth firms in the top 1 percent ranges from 34 percent to 38 percent, with 36 percent on average. Growth, however, is still a relatively rare event even among the elite: the average firm within the top 1 percent of estimated entrepreneurial quality has only a 2.4 percent chance of realizing a growth outcome.

In online Appendix Table A2, 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 (A2-1) we report a variation of our model after also including year fixed effects; (A2-2) includes state-specific time trends; and (A2-3) includes both year fixed effects and state-specific time trends. While there is some variation in the magnitude of the coefficients, these changes are relatively small, suggesting that the estimates are not driven by idiosyncratic variation across years or states.[21]

IV. The State of American Entrepreneurship

We now leverage these prediction models to calculate the centerpiece of our analysis: evaluating trends in entrepreneurial quality (EQI), entrepreneurial potential (RECPI), and regional economic performance (REAI) across the 32 states in our sample from 1988 through 2014. We estimate two RECPI indexes: a full information index based on Table 3, column 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 (Table 3, column 4), which we call Nowcasted RECPI. US RECPI, reported in Figure 3, is RECPI adjusted by the aggregate GDP of the 32 states in the sample.[22] Finally, we also include a

[20] Specifically, we divide our sample into ten random subsamples, using the first subsample as a testing sample and using the other nine to train the model. For the retained test sample, we compare realized performance with entrepreneurial quality estimates from the model resulting from the nine training samples. We then repeat this process nine 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.

[21] Online Appendix Table B1 assesses the robustness of the index across states by estimating the out-of-sample shares of firms in the top 5 percent and top 10 percent of quality by state and the correlation between entrepreneurial quality estimates performed individually for each state and the 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.

[22] 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.

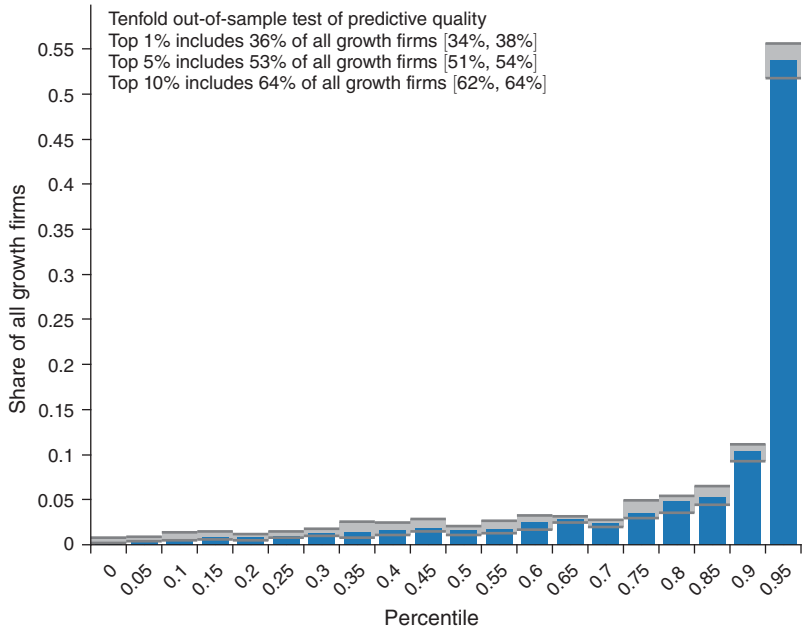

FIGURE 2

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. We use a tenfold cross-validation and plot the incidence of growth across each 5 percent bin.

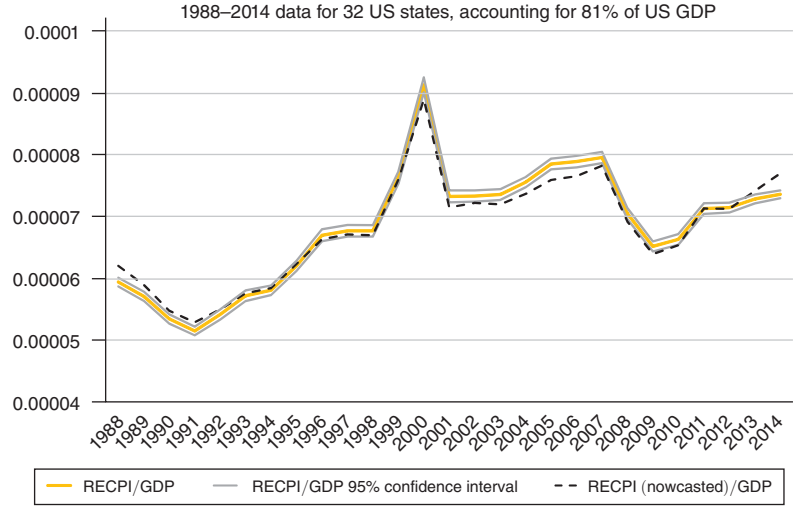

FIGURE 3. US AGGREGATE ENTREPRENEURSHIP REGIONAL ENTREPRENUERSHIP COHORT POTENTIAL INDEX (RECPI) BY YEAR

Note: RECPI/GDP represents the total, quality-adjusted entrepreneurship production in a region after controlling for the size of the economy in that year.

confidence interval estimated through a Monte Carlo process repeating our procedure for 30 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 US RECPI and Nowcasted US RECPI move very close to each other and that the confidence interval of US 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 during the 2000s, through 2008, 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.[23] Interestingly, Nowcasted US 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-ups 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 online Appendix Figure A1 as well as the number of growth firms in Figure 1 peak in the years 1995 and 1996 (respectively), US RECPI instead peaks in the year 2000. This offers insight into the possible 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 US 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).

US RECPI offers a new perspective on the "state" of entrepreneurship (at least for these 32 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, US 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 32 states in our sample. We present RECPI by zip code for all zip codes with at least 10 new firms (to avoid overcrowding the image), 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).

This map offers insight into the distribution of entrepreneurial quantity and quality across the United States. First, the most intense areas for entrepreneurial potential

[23] These broad patterns accord closely with the patterns we found for Massachusetts in Guzman and Stern (2017).

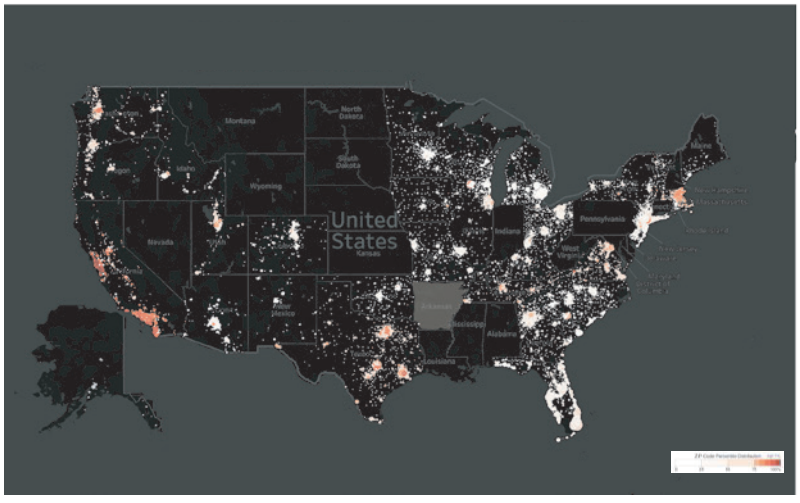

FIGURE 4. THE STATE OF AMERICAN ENTREPRENEURSHIP

Notes: This map represents the quality and quantity of entrepreneurship in 2012 across all 32 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.

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 other well-known locations, such as Seattle, northern Virginia (in the Washington, DC, area), and Denver register a high average EQI. At the same time, there are large areas of the United States that 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 have 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 as measured by the number of realized growth firms as well as market value created by firms in those cohorts. Finally, given the more gently sloped shape of the entrepreneurial boom of recent years, it may be the case that this accumulation of entrepreneurial potential is more sustainable than earlier periods.

V. Trends in the Effect of the US Entrepreneurial Ecosystem (REAI)

Entrepreneurship performance depends on not simply founding new enterprises but the scaling of 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 Entrepreneurial 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 32 states for which we have data), we plot REAI over time from 1988 to 2008, and we develop a projected measure of REAI for years 2009–2012.[24]

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 1.58 for the 1995 cohort. This is consistent with our evidence from Figure 1, in which the 1995–1996 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 a result of the environment (i.e., time) in which that start-up was trying to grow. From 2000 to 2007 there is a period of slight decline, with REAI moving from 0.95 down to 0.63. These differences are economically meaningful: a start-up at a given quality level is estimated to be three times more likely to experience a growth event in the six years after founding if it was founded in 1995 rather than in 2007. Finally, though still a preliminary estimate, we observe a resurgence in REAI for cohorts from 2007 to 2012, 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 ultimate impact once all growth outcomes are realized remains to be seen.

VI. Equity versus Employment Growth Outcomes

While equity growth is a measure of success for founders and investors, realizing significant employment growth is an alternative measure of entrepreneurial success more closely tied to broader economic performance (e.g., Krishnan, Nandy, and Puri 2015; Davis and Haltiwanger 1992). While a full analysis of the relationship

[24] Because our approach requires that we observe the *realized* growth firms, we can only measure our index with a six-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.

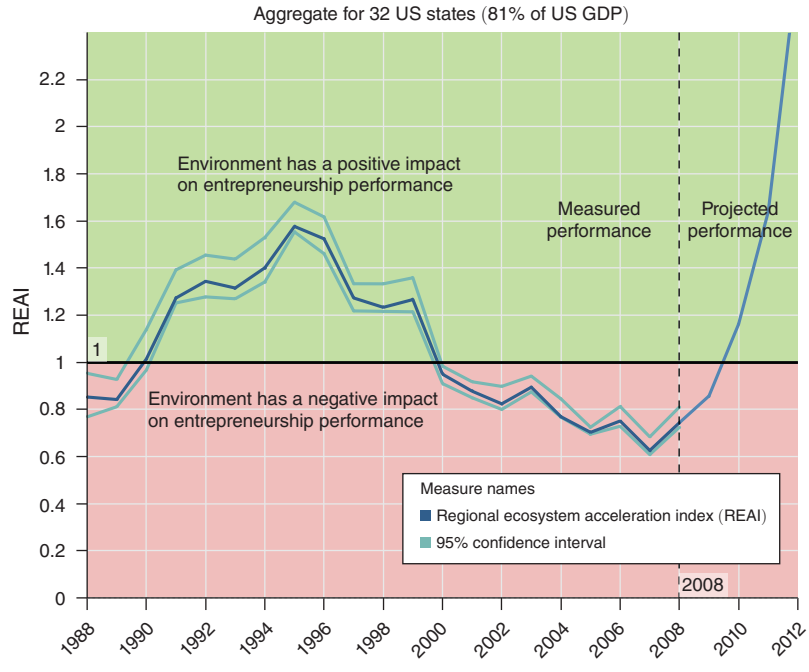

FIGURE 5. REGIONAL ENTREPRENEURIAL ACCELERATION INDEX (REAI)

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 30 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.

between business registration observables and comprehensive employment outcomes is beyond the scope of this paper (as such an analysis could be conducted more naturally in the context of an integrated longitudinal database such as the US Census's Longitudinal Business Database (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 between 1997 and 2014 (Infogroup 2014).[25] We construct two new outcome variables, *Employment Growth 500* and *Employment Growth 1000*, each equal to 1 for all firms recorded as having greater than 500 or 1,000 or more employees, respectively, within 6 years, and 0 otherwise. Though this measure does not capture the employment levels of

[25] Infogroup is a private sector business database similar to Dunn and Bradstreet. An overview of the dataset and our variable construction, as well as references to prior work using these data, is provided in online Appendix D. We utilize the annual snapshot data maintained by MIT Libraries from 1997 to 2014. We match firms by name and states and then examine, for each firm name/state combination, whether that firm achieves a given employment outcome (500 or 1,000) within 6 years of its business registration. In addition, to avoid duplicates, we focus only on headquarter locations (as indicated by Infogroup), deleting all nonheadquarter establishments. Infogroup reports employment for the entire company in the "headquarter" entry.

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. As context, the 500-employee threshold is used to differentiate small and medium-sized enterprises from large firms by the Small Business Administration, and so it is useful to consider this transition within six years from one category to the other.

We use these data to conduct three interrelated exercises. First, in Table 4, 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 500*. For example, firms with a trademark are 6.2 times more likely to get 1,000 employees (4 times for equity growth), firms with a patent 46.8 times (20.8 times for equity growth), firms registered in Delaware 13.3 times (14.0 for equity growth), and firms with both a patent and Delaware registration 131.4 times (80.56 for equity growth). 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 approach to form a quality estimate for each firm in our sample and compare our initial entrepreneurial quality estimates with this alternative. The correlation between a predictive analytic based on *Growth* versus *Employment Growth 500* is 0.84. 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 percent and 10 percent of quality. The results are striking: more than 48 percent of all measured employment growth outcomes occur within the top 10 percent of our entrepreneurial quality distribution, with around 40 percent in the top 5 percent.

While we emphasize that this analysis is incomplete insofar as our measures of employment growth may be incomplete, it suggests nonetheless 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.

VII. The Impact of the Business Cycle on Entrepreneurial Quantity and Quality

We now proceed to evaluate how the business cycle influences US entrepreneurial quality and quantity. To do so, we implement a SVAR regression that models the interdependent relationship between RECPI and the business cycle and allows us to estimate the impact of GDP growth on entrepreneurship. The impact of business cycles on entrepreneurship has been long debated in economics, most notably in the contrast between the debt deflation theory of Fisher (1933) and the "cleansing effect" of recessions emphasized by Schumpeter (1939).

Table 4—Entrepreneurial Quality Models with High Employment Growth Outcomes

Dependent variable	Equity growth (IPO or acquisition) (1)	Employment > 500 (2)	Employment > 1,000 (3)
Corporate governance measures			
Corporation	3.008	1.542	1.378
	(0.0860)	(0.0681)	(0.103)
Name-based measures			
Short name	2.248	1.568	1.279
	(0.0514)	(0.0635)	(0.0883)
Eponymous	0.304	0.675	0.781
	(0.0197)	(0.0595)	(0.112)
Intellectual property measures			
Trademark	3.984	7.194	6.243
	(0.268)	(0.750)	(1.053)
Patent–Delaware interaction			
Delaware only	14.01	12.61	13.43
	(0.354)	(0.626)	(1.149)
Patent only	20.83	26.52	46.79
	(1.101)	(2.607)	(6.684)
Patent and Delaware	80.56	95.87	131.4
	(3.645)	(9.064)	(19.86)
US CMP clusters	Yes	Yes	Yes
US CMP high-tech clusters	Yes	Yes	Yes
Observations	12,842,817	12,842,817	12,708,349
Pseudo R^2	0.184	0.103	0.100

Notes: 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 0 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 are in parentheses.

Bernanke and Gertler (1989) offer the first formal account of the debt deflation hypothesis, demonstrating how random macroeconomic shocks influence the balance sheets of would-be entrepreneurs and consequently change their ability to undertake the ambitious projects that represent new high-quality entrepreneurship (see also Carlstrom and Fuerst 1997 and Rampini 2004). Relatedly, business cycles also change investor expectations concerning the future availability of capital, which in turn can lead to a reduction in the riskiness of financed projects (Caballero, Farhi, and Hammour 2006), a dynamic with particular implications for the availability of early-stage venture capital (Nanda and Rhodes-Kropf 2016). On the other hand, a smaller theoretical literature has focused on the potential for more growth-oriented entrepreneurship to be founded in recessions due to a "cleansing effect," under the possibility that the bankruptcy or financial stress of marginal incumbent firms during a downturn might enhance entry opportunities for new productive firms (Schumpeter 1939, Caballero and Hammour 1994).

Existing empirical evidence has yet to support precisely one hypothesis or the other. For example, while Nanda and Rhodes-Kropf (2013) show that increases in

the supply of venture capital lead venture capitalists to invest in more innovative firms, and Moreira (2016) highlights that there are procyclical differences in the initial size of firms across the business cycle, which persist, Koellinger and Thurik (2012) do not observe any relationship between GDP growth and the subsequent quantity of entrepreneurship (using surveys of business ownership) in a panel of 22 OECD countries.[26] More generally, none of these papers provide a direct test of the key underlying hypotheses in Bernanke and Gertler (1989) or other theories. Do positive changes in the GDP growth rate predict no change in the overall quantity of entrepreneurship and a positive change in the quality-adjusted quantity of entrepreneurship? Conversely, do recessions result in a downward shift in the distribution of entrepreneurial quality-adjusted quantity?

To evaluate such relationship, we propose a structural vector autoregression model that models the interdependent nature of entrepreneurship and economic growth. Economic growth influences entrepreneurship contemporaneously and with a lag, while entrepreneurship only responds to growth with a lag, reflecting the time it takes to undertake an investment, which can be modeled as follows:

$$\ln(\Delta GDP_t) = a_1 \ln\left(\frac{RECPI_{t-1}}{GDP_{t-1}}\right) + \cdots + a_n \ln\left(\frac{RECPI_{t-n}}{GDP_{t-n}}\right) + b_1 \ln(\Delta GDP_{t-1}) + \cdots + b_n \ln(\Delta GDP_{t-n}) + u_t, \tag{5}$$

$$\ln\left(\frac{RECPI_t}{GDP_t}\right) = c_1 \ln\left(\frac{RECPI_{t-1}}{GDP_{t-1}}\right) + \cdots + c_n \ln\left(\frac{RECPI_{t-n}}{GDP_{t-n}}\right) + d_0 \ln(\Delta GDP_t) + d_1 \ln(\Delta GDP_{t-1}) + \cdots + d_n \ln(\Delta GDP_{t-n}) + v_t, \tag{6}$$

where $\ln(\Delta GDP_t)$ represents GDP growth (in log) at time t, and $\ln(RECPI_t/GDP_t)$ represents the quality-adjusted flow of entrepreneurship at t, and u_t and v_t are the idiosyncratic disturbances in the growth rate and the entrepreneurship rate, respectively. We fit a recursive SVAR to estimate d_n as the percent increase in quality-adjusted entrepreneurship from a percentage increase in the annual GDP growth rate.

Using the (admittedly small) sample of 27 annual observations for the United States from 1988 to 2014, online Appendix Table A4 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 (A4-1) and (A4-2) with a reduced-form single-lag VAR model. While (A4-1) reports a positive relationship between GDP growth and RECPI, (A4-2) 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.[27]

[26] Koellinger and Thurik (2012) do find, however, a relationship in the opposite causal direction, that entrepreneurship predicts (Granger causes) economic growth.

[27] The three-lag structure minimizes both the Akaike Information Criterion and Schwarz's Bayesian Information Criterion.

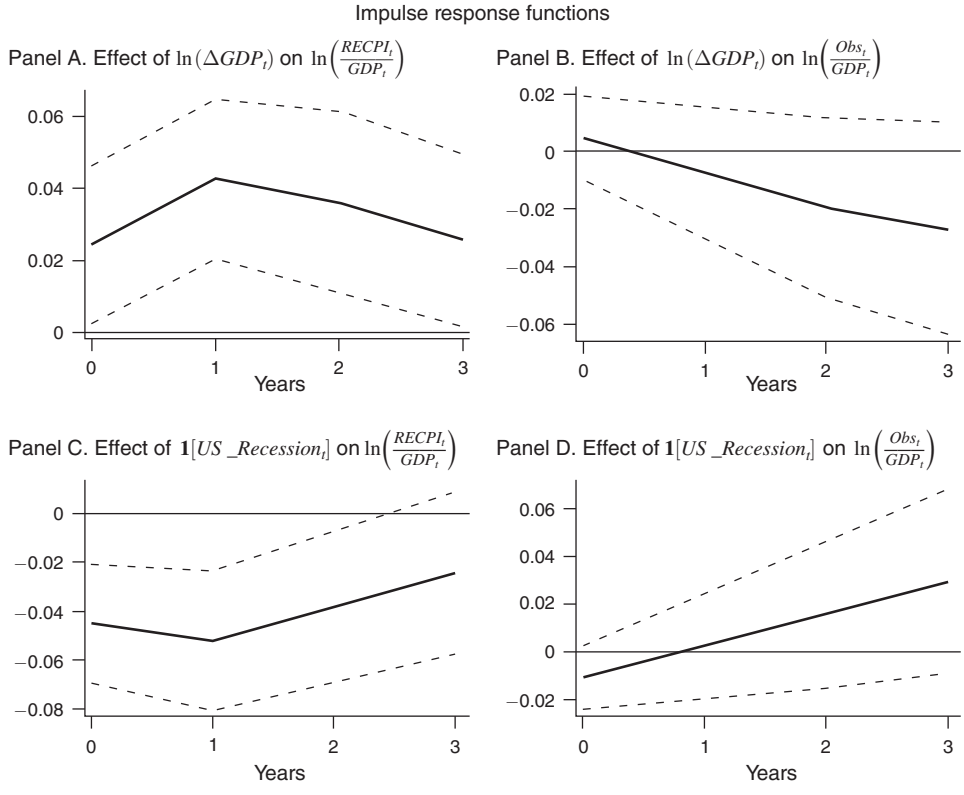

FIGURE 6. EFFECT OF ECONOMIC CONDITIONS ON ENTREPRENEURSHIP

Notes: This figure reports the relationship of economic conditions to entrepreneurship production in the United States. $\ln(\Delta GDP_t)$ is the log change in US GDP between years $t-1$ and t. $\mathbf{1}[US\_Recession_t]$ 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. All lag structures are chosen as those that maximize Akaike's Information Criterion (AIC) and 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 ln(ΔGDP) models and is marginally unable to reject it for the US Recession models ($p = 0.13$).

The parameter estimates are similar. Throughout, we observe a positive relationship between GDP growth and subsequent RECPI and no relationship between GDP growth and the raw quantity of entrepreneurship. We see this more clearly in panels A and B of Figure 6, which present the impulse response function for regression. The figures indicate that a doubling of the GDP growth rate leads to a 2 percent increase in $RECPI/GDP$ in current year t and a 4 percent increase in years $t+1$ and $t+2$, which then tapers off. In contrast, as illustrated in panel B of Figure 6, there is no net relationship between GDP growth and $\ln\left(Obs/GDP\right)$. We further this analysis in (A4-5) and (A4-6) by considering an alternative measure of business cycles, the presence of an economic recession as determined by the NBER Business Cycle Dating Committee. As shown in panels C and D of Figure 6, the onset of a recession decreases $RECPI/GDP$ by 5 percent in t and $t+1$ (with a subsequent tapering off) while having no net impact on $\ln\left(Obs/GDP\right)$.

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 between economic shocks and the propensity to start ventures with high growth potential at founding.

VIII. Conclusion

This paper develops a quality-based approach with business registration records for 32 states to create and evaluate novel indices of the quantity and quality-adjusted quantity of entrepreneurship. 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 economic fluctuation.

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. US RECPI reflects broad and well-known changes in the environment for start-ups, such as the dot-com 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 30 years. While the expected number of high-growth start-ups peaked in 2000 and then fell dramatically with the dot-com bust, starting in 2010 there has been a sharp upward swing in the expected number of successful start-ups 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 US RECPI indicate a net *upward* trend across the full time series of our sample. For example, the rate of expected successful start-ups 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 start-ups and instead more related to other dynamics or ecosystem concerns. In particular, while there is high cyclicality in RECPI/GDP, REAI—the likelihood of start-ups to reach their potential—declined in the late 1990s and did not recover through 2008. Relative to the mid-1990s, the 2000s were a period in which a lower level of entrepreneurial potential was realized. For example, conditional on the same estimated potential, a 1995 start-up was three times more likely to achieve a growth event in six years than a start-up founded in 2007.

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 toward entrepreneurial quality allows one to connect entrepreneurship and overall economic performance more directly. Using our measures in a structural VAR model, we find economic shocks are associated with a procyclical impact on the quality-adjusted quantity of entrepreneurship, while there is no relationship with quantity alone. These results provide novel empirical evidence on the way entrepreneurship is shaped by economic conditions and allow us to begin to adjudicate between competing theories of this relationship.

More generally, our analysis suggests that taking directly 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 entrepreneurial quality, the growth process, and even the migration of firms between regions is a promising area for future research (Guzman 2018). 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 start-ups after founding (such as the receipt of venture capital) impact that growth process. For example, in Catalini, Guzman, and Stern (2019), we examine both the selection into and impact of venture capital on start-up firms by exploiting this predictive analytics approach. 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., around 40 percent of all equity growth outcomes are contained within the top 1 percent 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. These benefits are likely to be enhanced by ongoing technological improvements in data storage and processing capabilities, which will likely improve the precision and applicability of these estimates. To this end, these estimates are an initial implementation pointing toward a more general approach using founding observables and ex post performance to estimate founding quality. For example, it may be possible to develop integrated datasets including measures based on firm founding statements, online job postings, media mentions, and the degree and nature of online or social media presence (e.g., presence or absence of a web page, functionality of that web page, etc.). Combined with more sophisticated predictive algorithms (e.g., machine learning approaches as developed in Guzman 2018), it may be possible to capture different types of performance and the linkage between the initial conditions at founding and different types of economic and social impact.

Appendix

Table A1—Data Coverage US States—Ranked by GDP

Rank in US GDP	State	GDP	Share of GDP (percent)
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
46	Alaska	$60,542	0.3
47	Maine	$56,163	0.3
50	Rhode Island	$45,962	0.3
52	Vermont	$30,723	0.2
	Total GDP in sample	$13,941,781	
	Number of states	32	
	US GDP	$17,565,783	
	Share of GDP in sample	81%	

REFERENCES

Acs, Zoltan J., and David B. Audretsch. 1988. "Innovation in Large and Small Firms: An Empirical Analysis." *American Economic Review* 78 (4): 678–90.

Akcigit, Ufuk, and William R. Kerr. 2018. "Growth through Heterogeneous Innovations." *Journal of Political Economy* 126 (4): 1374–1443.

Andrews, Raymond J., Catherine Fazio, Jorge Guzman, Yupeng Liu, and Scott Stern. 2019. "Startup Cartography Project [Data from 1988–2014]." https://www.startupcartography.com (accessed December 14, 2019).

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics—An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Aulet, Bill, and Fiona Murray. 2013. "A Tale of Two Entrepreneurs: Understanding Differences in the Types of Entrepreneurship in the Economy." https://www.issuelab.org/resources/15236/15236.pdf.

Axtell, Robert L. 2001. "Zipf Distribution of US Firm Sizes." *Science* 293 (5536): 1818–20.

Barnes, Beau Grant, Nancy L. Harp, and Derek Oler. 2014. "Evaluating the SDC Mergers and Acquisitions Database." *Financial Review* 49 (4): 793–822.

Belenzon, Sharon, Aaron K. Chatterji, and Brendan Daley. 2017. "Eponymous Entrepreneurs." *American Economic Review* 107 (6): 1638–55.

Belenzon, Sharon, Aaron Chatterji, and Brendan Daley. 2018. "Choosing between Growth and Glory." NBER Working Paper 24901.

Bernanke, Ben, and Mark Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review* 79 (1): 14–31.

Caballero, Ricardo J., Emmanuel Farhi, and Mohamad L. Hammour. 2006. "Speculative Growth: Hints from the U.S. Economy." *American Economic Review* 96 (4): 1159–92.

Caballero, Ricardo, Emmanuel Farhi, and Mohamad Hammour. 2006. "Speculative Growth: Hints from the US Economy"*American Economic Review* 96 (4): 1159-192.

Caballero, Ricardo J., and Mohamad L. Hammour. 1994. "The Cleansing Effect of Recessions." *American Economic Review* 84 (5): 1350–68.

Carlstrom, Charles T., and Timothy S. Fuerst. 1997. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis." *American Economic Review* 87 (5): 893–910.

Catalini, Christian, Jorge Guzman, and Scott Stern. 2019. "Passive versus Active Growth: Evidence from Founder Choices and Venture Capital Investment." NBER Working Paper 26073.

Chatterji, Aaron K. 2018. "The Main Street Fund: Investing in an Entrepreneurial Economy." https://www.hamiltonproject.org/assets/files/Chatterji_20180611.pdf.

Churchwell, Cynthia. 2016. "Q. SDC: M&A Database." Baker Library—Fast Answers. https://asklib.library.hbs.edu/faq/47760 (accessed January 17, 2017).

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

Davis, Steven J., John C. Haltiwanger, and Scott Shuh. 1996. *Job Creation and Destruction.* Cambridge, MA: MIT Press.

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

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2016. "Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S." *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, and Larry Samuelson. 1988. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries." *RAND Journal of Economics* 19 (4): 495–515.

Ericson, Richard, and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62 (1): 53–82.

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

Fazio, Catherine, Jorge Guzman, and Scott Stern. 2019. "The Impact of State-Level R&D Tax Credits on the Quantity and Quality of Entrepreneurship." NBER Working Paper 26099.

Fisher, Irving. 1933. "The Debt-Deflation Theory of Great Depressions." *Econometrica* 1 (4): 337–57.

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

Gornall, Will, and Ilya A. Strebulaev. 2015. "The Economic Impact of Venture Capital: Evidence from Public Companies." Stanford Graduate School of Business Working Paper 3362.

Guzman, Jorge. 2019. "Go West Young Firm: Agglomeration and Embeddedness in Startup Migrations to Silicon Valley." Columbia Business School Research Paper 18-49.

Guzman, Jorge, and Aleksandra (Olenka) Kacperczyk. 2019. "Gender Gap in Entrepreneurship." *Research Policy* 48 (7): 1666–80.

Guzman, Jorge, and Scott Stern. 2015. "Where Is Silicon Valley?" *Science* 347 (6222): 606–09.

Guzman, Jorge, and Scott Stern. 2017. "Chapter 2—Nowcasting and Placecasting Entrepreneurial Quality and Performance." In *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, edited by John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar, 63–110. Chicago: University of Chicago Press.

Guzman, Jorge. 2018. "Go West Young Firm: Agglomeration and Embeddedness in Startup Migrations to Silicon Valley." Unpublished.

Guzman, Jorge, and Scott Stern. 2020. "Replication Data for: The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E111705V1.

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

Hathaway, Ian, and Robert Litan. 2014a. "Declining Business Dynamism in the United States: A Look at States and Metros." https://www.brookings.edu/wp-content/uploads/2016/06/declining_business_dynamism_hathaway_litan.pdf.

Hathaway, Ian, and Robert E. Litan. 2014b. "Declining Business Dynamism: It's for Real." https://www.brookings.edu/wp-content/uploads/2016/06/final2_declining_business_dynamism_its_for_real_hathaway_litan.pdf.

Hathaway, Ian, and Robert Litan. 2014c. "The Other Aging of America: The Increasing Dominance of Older Firms." https://www.brookings.edu/wp-content/uploads/2016/06/other_aging_america_dominance_older_firms_hathaway_litan.pdf.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60 (5): 1127–50.

Hurst, Erik, and Benjamin Wild Pugsley. 2011. "What Do Small Businesses Do?" *Brookings Papers on Economic Activity* 41 (2): 73–142.

Infogroup. 2014. "ReferenceUSA Business Historical Data Files." Harvard Dataverse. https://doi.org/10.7910/DVN/GW2P3G (accessed summer 2016).

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

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

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

Klapper, Leora, Raphael Amit, and Mauro F. Guillén. 2010. "Chapter 4—Entrepreneurship and Firm Formation across Countries." In *International Differences in Entrepreneurship*, edited by Josh Lerner and Antoinette Schoar, 129–58. Chicago: University of Chicago Press.

Klepper, Steven. 1996. "Entry, Exit, Growth, and Innovation over the Product Life Cycle." *American Economic Review* 86 (3): 562–83.

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

Koellinger, Philipp D., and A. Roy Thurik. 2012. "Entrepreneurship and the Business Cycle." *Review of Economics and Statistics* 94 (4): 1143–56.

Kortum, Samuel, and Josh Lerner. 2000. "Assessing the Contribution of Venture Capital to Innovation." *RAND Journal of Economics* 31 (4): 674–92.

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

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

Mills, Karen Gordon, and Brayden McCarthy. 2016. "The State of Small Business Lending: Innovation and Technology and the Implications for Regulation." Harvard Business School (HBS) Working Paper 17-042.

Moreira, Sara. 2016. "Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles." Society for Economic Dynamics Meeting Paper 708.

Nanda, Ramana, and Matthew Rhodes-Kropf. 2013. "Investment Cycles and Startup Innovation." *Journal of Financial Economics* 110 (2): 403–18.

Nanda, Ramana, and Matthew Rhodes-Kropf. 2014. "Financing Risk and Innovation." Harvard Business School (HBS) Working Paper 11-013.

Nanda, Ramana, and Matthew Rhodes-Kropf. 2016. "Financing Entrepreneurial Experimentation." *Innovation Policy and the Economy* 16 (1): 1–23.

Netter, Jeffry, Mike Stegemoller, and M. Babajide Wintoki. 2011. "Implications of Data Screens on Merger and Acquisition Analysis: A Large Sample Study of Mergers and Acquisitions from 1992 to 2009." *Review of Financial Studies* 24 (7): 2316–57.

Pohlman, John T., and Dennis W. Leitner. 2003. "A Comparison of Ordinary Least Squares and Logistic Regression." *Ohio Journal of Science* 103 (5): 118–25.

Rampini, Adriano. 2004. "Entrepreneurial Activity, Risk, and the Business Cycle." *Journal of Monetary Economics* 51 (3): 555–73.

Refinitiv. 2018. "SDC Platinum." Refinitiv—Columbia Libraries. (accessed January 2019).

Samila, Sampsa, and Olav Sorenson. 2011. "Venture Capital, Entrepreneurship, and Economic Growth." *Review of Economics and Statistics* 93 (1): 338–49.

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

Schumpeter, Joseph A. 1939. *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*. Eastford, CT: Martino Publishing.

Scott, Steven, and Hal R. Varian. 2015. "Nowcasting Economic Time Series." In *Economic Analysis of the Digital Economy*, edited by Avi Goldfarb, Shane Greenstein, and Catherine Tucker, 119–135. Chicago: University of Chicago Press.

Stone, Brad. 2013. *The Everything Store—Jeff Bezos and the Age of Amazon*. New York: Little, Brown.

Sutton, John. 1997. "Gibrat's Legacy." *Journal of Economic Literature* 35 (1): 40–59.

Tartari, Valentina, and Scott Stern. 2018. "The Role of Universities in Local Entrepreneurial Ecosystems." https://pdfs.semanticscholar.org/a37f/59a7f020058b9b674594ba0beb85b67f98d4.pdf?_ga=2.91604186.1904163147.1580845066-790025412.1578061555.

US Bureau of Economic Analysis. 2017. "Real Gross Domestic Product [GDPCA]." Federal Reserve Economic Data (FRED), Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/GDPCA (accessed July 2017).

US Census Bureau. 2017. "Legacy Firm Characteristics Tables 1977–2014." Business Dynamics Statistics, US Census Bureau. https://www.census.gov/programs-surveys/bds.html (accessed on July 2017).

US Patent and Trademark Office (USPTO). 2018a. "Patent Assignment Dataset." United States Patent and Trademark Office. https://www.uspto.gov/learning-and-resources/ip-policy/economic-research/research-datasets (accessed on July 2018).

US Patent and Trademark Office (USPTO). 2018b. "Patent Examination Research Dataset (Public PAIR)." United States Patent and Trademark Office. https://www.uspto.gov/learning-and-resources/ip-policy/economic-research/research-datasets (accessed on July 2018).

US Patent and Trademark Office (USPTO). 2018c. "Trademark Case Files Dataset." United States Patent and Trademark Office. https://www.uspto.gov/learning-and-resources/ip-policy/economic-research/research-datasets (accessed on July 2018).

Witten, Ian H., and Eibe Frank. 2005. *Data Mining—Practical Machine Learning Tools and Techniques*. 2nd ed. Amsterdam: Elsevier.