NBER WORKING PAPER SERIES

HIDDEN IN PLAIN SIGHT:
VENTURE GROWTH WITH OR WITHOUT VENTURE CAPITAL

Christian Catalini
Jorge Guzman
Scott Stern

Working Paper 26521
http://www.nber.org/papers/w26521

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2019

We thank Theresa Amabile, Tania Babina, Iain Cockburn, Shane Greenstein, Josh Lerner, Ramana Nanda, Giorgia Piacentino, Manju Puri, Tim Simcoe, and Jane Wu for helpful comments, as well as participants at the Dartmouth Venture Capital and Private Equity Conference, the NBER Productivity Lunch, the Kauffman Scholars Conference, and seminar participants at Brandeis University, Michigan University, Kellogg School of Management, and Columbia Business School. We also acknowledge and thank the Kauffman Foundation for their support of our research agenda, including the Uncommon Methods and Metrics Grant and the Kauffman Dissertation Fellowship, as well as the Jean Hammond (1986) and Michael Krasner (1974) Entrepreneurship Fund and the Edward B. Roberts (1957) Entrepreneurship Fund at MIT. Yupeng Liu provided excellent research assistantship. All errors and omissions are of course our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w26521.ack

Hidden in Plain Sight: Venture Growth with or without Venture Capital
Christian Catalini, Jorge Guzman, and Scott Stern
NBER Working Paper No. 26521
December 2019
JEL No. G24,L26

ABSTRACT

The majority of IPOs and acquisitions are achieved without venture capital financing, yet research has focused mostly on VC backed firms. Using founding choices and a predictive analytics approach on virtually all US registered businesses, we shed light into these "missing" growth firms. Founding choices that predict raising venture capital also strongly predict equity exits without VC. Firms with growth potential are similar to each other, irrespective of funding source. Moreover, matching firms that are born with identical observables, but only differ in whether they receive venture capital, suggests an upper bound to the returns to venture capital of 600%.

Christian Catalini
MIT Sloan School of Management
100 Main Street, E62-480
Cambridge, MA 02142
and NBER
catalini@mit.edu

Jorge Guzman
Columbia Business School
Uris Hall, 711
116th St & Broadway
New York, NY 10027
jag2367@gsb.columbia.edu

Scott Stern
MIT Sloan School of Management
100 Main Street, E62-476
Cambridge, MA 02142
and NBER
sstern@mit.edu

I. INTRODUCTION

The skewed nature of firm growth outcomes is a striking feature of the process through which entrepreneurship influences broader economic performance. From a financial perspective, only a very small fraction of firms (less than 1 in 2000) reaches a successful financial exit in the form of an IPO or successful acquisition. Most of what we know about these growth firms comes from carefully constructed samples of firms funded by venture capitalists and angel investors (Lerner, 1995; Hellman and Puri, 2000; Chemmanur, Nandy, and Krishnan, 2011; Lerner et al 2015; Puri and Zarutskie, 2012). By following startups from their earliest funding rounds to an exit, this stream of research surfaced the central role professional investors play in enabling and accelerating startup growth. Although VCs only fund a very small number of startups each year (approximately a thousand in the US), they account for a disproportionate share of growth events: Kaplan and Lerner (2010) estimate that venture-backed companies account for an impressive 30% to 70% of "startup" IPOs (1995-2009)[2], and, more recently, Ritter (2016) traces back 37% of startup IPOs to a VC funding event (1980-2015).

Whereas these shares are a testament to the role VCs play in the selection and nurturing of high potential startups, they also indirectly highlight how little we know about the sizable share of firms that achieve growth without ever being associated with a venture capital firm. Of course, given that VC activity is concentrated within a few regions and sectors, it is possible that firms that grow without VC are simply coming from areas and industries that have not yet developed a thriving venture capital ecosystem. Under this hypothesis, we would anticipate that the firms that ultimately growth without venture capital would be in many respects similar at founding to the firms that growth with venture capital, although their growth trajectories could be somewhat different. For example, it is possible that growth without venture capital would be concentrated among firms that are of even higher quality at founding, since a less favorable funding environment – either within a non-hub region or during a VC downturn – would select out many ventures that could succeed if only venture capital were available. In

[2] Startup IPOs are all IPOs after excluding financial IPOs, blank check companies, re-listings, reverse LBOs, real estate investment trusts (REIT), and special purpose acquisition companies (SPAC).

the absence of VC, growth may also take longer to materialize, as firms have to slowly bootstrap their development through alternative sources of capital such as revenues from sales, loans, government grants etc.

An alternative hypothesis is that despite regional, industry and economic cycle differences, multiple routes to equity growth exist, and that the broader availability of data on VC-backed firms has skewed researchers' focus towards just one of the possible paths to growth. Conditional on alternative paths to growth actually existing, this raises the question of how they may differ (if at all) from the venture capital one, and what types of firms are more likely to select into one versus the other. The underlying, key welfare question is one of how society allocates capital to novel, high potential ideas and encourages their development from concept to market.

By design, the study of selection into alternative paths to growth requires first defining the full population of firms at risk of growth, and then following their outcomes *independent* of funding source and path chosen. This has prevented previous studies from systematically examining this process, as most research either: a) starts from a selected sample (e.g. the set of firms that raise venture capital, qualify for a government grant, etc.) and then matches it to controls along idiosyncratically chosen dimensions; or b) directly compares VC-funded firms to the general population of firms, the vast majority of which is never really at risk of growing in the first place. Whereas the first approach typically misses firms with growth potential that do not fit the venture capital 'playbook', the second one overestimates the role of VC on growth because it confounds selection and treatment.

The objective of this paper is to characterize the differences between firms that achieve a significant growth outcome with versus without venture capital. To identify the full set of firms with growth potential – irrespective of future funding source – we extend Guzman and Stern's (2015, 2017, 2019) predictive analytics approach, and estimate a *'VC-likelihood'* for all incorporated firms based on information that is available at the time of their founding. We then use this estimate to match VC-funded firms to comparable control firms from the non-VC-funded part of the sample. Our empirical approach follows three steps, which we describe in more detail below.

In the first step, we train a model on a random subsample of all incorporated firms[3] to learn as much as possible, using historical data on VC funding events, from the selection process performed by venture capitalists. In this step, our objective is to extract key observable dimensions VCs select on when trying to predict the future growth potential of a firm, and then use the results from this predictive analytics exercise to calculate a *VC-likelihood* for all the firms in our sample (irrespective of them receiving VC funding or not). This measure allows us to replicate elements of the screening process VCs perform on every firm in the economy, including firms VCs may have turned down or never had a chance to evaluate in the first place (e.g. because they do not operate in the region the firm is located, were not aware of the deal, or the company did not look for venture capital funding). Of course, VCs collect substantially more information than we do when deciding to invest in a startup or not through face-to-face meetings, due diligence, etc. At the same time, as long as some of the dimensions they care about are captured by our data, then our method should be able to replicate at least part of their screening heuristics.

We train our model on venture capitalists because their objective is to maximize the chances of an equity growth event: i.e., by studying the observables that correlate with their decision to invest, we are able to identify firm characteristics that VCs *believe* can predict future growth. Whether or not these observables have any actual predictive power (beyond the self-fulfilling component resulting from the VCs 'treatment effect' on the firms) is an empirical question. It is also not clear, a priori, if the same dimensions VCs select on would be predictive of growth within the sample of non-VC-funded firms. If VCs endogenously match with firms that they know would benefit the most from their approach to scaling startups, then non-VC-funded firms that grow could be fundamentally different than VC-funded ones. If instead there is a single playbook for firm growth, and VCs are able to capture some of the early signals of

[3] 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. Moving beyond simple counts of business registrants (Klapper, Amit, and Guillen, 2010), we are able to measure characteristics related to entrepreneurial quality at or close to the time of registration. These characteristics include how the firm is organized (e.g., as a corporation, partnership, or LLC, and whether the company is registered in Delaware), how it is named (e.g., whether the owners name the firm eponymously after themselves), and how the idea behind the business is protected (e.g., through an early patent or trademark application). These startup characteristics may reflect choices by founders who perceive their venture to have high potential. As a result, though observed startup characteristics are not causal drivers of startup performance, they may nonetheless represent early-stage "digital signatures" of high-quality ventures.

a firm's future potential, then we would expect non-VC-funded firms that grow to be similar, on at least some of the dimensions VCs care about, to VC-funded firms. Our paper tests these competing hypotheses.

To do so, in the second step of our approach, we use the estimates resulting from our prediction of the *VC-likelihood* to explore the growth process among firms that *did not receive venture capital*.[4] One can think of our *VC-likelihood* as a proxy—based on firm observables—for the probability that a VC firm would have invested in a focal firm based on what was known about it around the time of its birth. The intuition behind this step is to check if the determinants of venture capital financing are similar to the determinants of growth *outside* of the venture capital sample. If venture growth in the absence of VC is fundamentally different from growth with VC (i.e. if these two paths to growth have little in common), then this analysis would surface the observables that are associated with VC funding, but are not associated with growth within the non-VC sample. If instead the two paths to growth are similar, then we would expect many of the determinants across the two models to be the same, and the *VC-likelihood* would be a good proxy for growth potential also in the absence of venture capital (since it distills some key predictors of growth VCs select on).

In the third step, since the *VC-likelihood* can be calculated for all firms independent of funding source, we explicitly use it to identify, for each firm that raised VC, a comparable, 'VC-type' firm among the firms that did not receive venture capital. This allows us to perfectly match each VC-funded firm with a control firm from the non-VC sample of the same observable estimated quality (at least from our model). We use this last step to describe both the process of selection into venture capital starting from the full population of new firms, and to estimate an upper bound to the returns from VC.

We apply this three-step approach to a dataset covering all business registrants in 49 US states and Washington D.C. (comprising 99.6% of US GDP) from 1995-2005. This data is part of the Startup Cartography Project (Andrews, et al, 2019). Relative to other work in this agenda, our paper is the first to consider the determinants of venture capital selection in the population of firms, rather than equity alone growth (Guzman and Stern, 2015, 2017, 2019),

[4] We leverage the fact that, although rare, we observe both the receipt of venture capital (though data on the precise amount and valuation is somewhat noisy) and meaningful growth outcomes for those firms that realize such outcomes (e.g., for equity growth, we can observe firm for example IPO or high-value acquisitions).

or the interaction of financing and gender (Guzman and Kacperczyk, 2019). An additional improvement upon this prior work is an expanded dataset from 32 US states to now include data from 49 states (and Washington, DC), as well as improved measures of venture capital financing across these firms.

Our analysis delivers several novel findings about the process of venture growth with versus without venture capital. While our estimates of the incidence of venture-backed IPOs are similar to prior estimates in the literature, they nonetheless highlight the important role that non-VC-backed companies play in economic growth: about 85% of all firms that achieve an equity growth event do so without venture capital financing (69% of IPOs, in our data). Furthermore, our results show that the process of selection into venture capital, similar to the process of equity growth (Guzman and Stern, 2015, 2017, 2019), is highly skewed, and can be characterized through a small number of firm observables at birth.[5]

When we use our estimates from the predictive analytics approach to understand growth within the non-VC sample, we find that a doubling in the estimated *VC-likelihood* more than doubles the probability of an equity growth outcome, with almost 45% of all non-VC-backed equity growth outcomes estimated to be in the top 5% of the *VC-likelihood* distribution. Among firms that did not raise venture capital, a firm in the top 0.5% of our estimated *VC-likelihood* distribution is 118X more likely to achieve an equity growth outcome than a firm in the bottom 50% of the distribution. The results highlight the striking similarity between the determinants of venture capital and the determinants of equity growth in the absence of venture capital. With the exception of trademark, which is more salient among non-venture-backed firms, all other startup characteristics are comparable across the two models. The relationship between our estimated *VC-likelihood* and equity growth within the non-VC-backed sample is also quite stable across time periods where venture capital was more versus less abundant, and across geographies (startup hubs versus not).

The stability of these estimates supports the view that our predictive analytics approach is able to capture fundamental firm characteristics that are predictive of growth irrespective of funding source and VC presence in a region or sector. The same estimates also allow us to

[5] Firms that have short names are 356% more likely to receive venture capital, while eponymous firms are more than 80% less likely to receive VC; firms that register in Delaware and receive or apply for a patent within a year of founding are 140X more likely to receive venture capital.

revisit the question of the role of venture capital in the process of equity growth, as they can be used to carefully match each venture-backed firm with a control firm with similar growth potential from birth. The method provides us with an upper-bound estimate of the returns to venture capital investment on equity growth, as VCs also select firms based on characteristics that are unobservable to us. We find that relative to a "naïve" estimate where venture capital is associated with a 400X increase in the probability of equity growth, our matching results suggest up to a 6X boost to equity growth from venture capital. Interestingly, in our data, the returns to venture capital are lower in the upper tail of the estimated *VC-likelihood* distribution (e.g., firms in the top 0.05% of the distribution receive only a 2.3X increase in their probability of growth). The estimated VC effect is also lower within startup hubs and during the .com market crash.

The contributions of this paper are two-fold. First, by using a predictive analytics approach to estimate the probability of receiving venture capital investment, we provide a more direct comparison of firms that grow with and without venture capital. Though prior work has highlighted the mechanisms that lead to incomplete venture capital investment markets (Hochberg et al, 2010; Ewens et al, 2013; Ewens and Townsend, 2019; Piacentino, 2019), and assessed the incidence of venture capital in initial public offerings (Ritter, 2016; Kaplan and Lerner 2010), our paper is the first to document the process of growth without VC systematically for all firms in the economy. A critical contribution in this respect is considering the founding observables of firms, and then be able to characterize the similarity of companies from founding to exit across these two financing paths. Second, we are able to use this comparison to offer a novel complementary approach to a growing literature on the "returns" to venture capital investment (Chemmanur, et al, 2011; Puri and Zarutskie, 2012). Our upper bound estimate is consistent with that of Puri and Zarutskie (2012), and provides additional evidence on the returns to VC, during our time period.

More generally, our results evidence important facts on the sources of venture capital selection and follow-on performance. They suggest that not only multiple financing paths to growth exist, but that there are strong similarities between firms that grow through either of these routes. And, once the estimated *VC-likelihood* is accounted for, the gap between VC-funded and other firms of comparable potential is much smaller than simpler approaches would lead to believe. In our companion paper, Catalini et al (2019), we build on the results in this paper to consider more carefully the relative incidence of "passive" versus "active" growth in the context of a structural model considering the interplay between the early-stage choices of founders and venture capitalists.

The paper proceeds as follows. Section 2 discusses the process of selection into venture capital and equity growth. Section 3 develops our predictive analytics approach. Section 4 introduces the data and descriptive statistics, before turning to the main empirical findings in Sections 5 and 6. Section 7 concludes.

II. VENTURE QUALITY, SELECTION INTO VENTURE CAPITAL, AND GROWTH

Over the past decade, there has been increasing appreciation for the skewed nature of entrepreneurial outcomes, and for the disproportionate impact high quality new ventures have on innovation, employment and productivity growth. Starting from founding, firms exhibit substantial heterogeneity in quality, and only a very small fraction of successful startups is responsible for the economy-wide benefits from entrepreneurship (Kerr, Nanda, and Rhodes-Kropf, 2014, Guzman and Stern, 2019). While there is increasing understanding of the importance of accounting for such heterogeneity in the measurement and impact of entrepreneurship on the economy (Schoar, 2010; also see Hurst and Pugsley, 2010, Lerner, 2009, and Decker, Haltiwanger, Jarmin, and Miranda, 2014, Guzman and Stern, 2015, 2019), systematic measurement of new venture quality has been challenging. In the area of entrepreneurial finance, researchers often rely on samples of firms that have reached rare milestones such as raising venture capital. While this facilitates the examination of the dynamics of high-potential firms, it also creates a disconnect between these small, selected samples of firms and the overall population of new ventures.[6] As emphasized by Hathaway and Litan, the challenge in directly incorporating heterogeneity 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. This is true even with venture-capital backed firms…"* (Hathaway and Litan, 2014).

[6] One notable and insightful exception is the positive relationship between organizing your firm as a corporation and entrepreneurial income highlighted by Levine and Rubinstein (2017).

Though it is certainly the case that entrepreneurship is a highly uncertain activity, it is nonetheless also the case that entrepreneurs and investors make (somewhat) informed decisions at a relatively early stage of the life of a firm, based on their best assessment of the growth potential of that firm. For example, given the objective of predicting and enabling startup growth, venture capital firms explicitly seek to identify new ventures that have a higher likelihood of achieving an equity growth outcome over a relatively bounded period of time (i.e., typically the lifetime of the fund). After selecting a firm, they also do not rely on a passive investment strategy, but actively support the ventures they add to their portfolio in order to accelerate their path to growth. At the same time, as shown by multiple studies in this area, their search and investment activities are concentrated not only in specific industries and geographies, but also disproportionately focus on specific types of firms, founding teams and technology trends. VCs are more likely to fund ventures that have secured (or are in the process of securing) formal intellectual property, and do not currently have a sizable stream of revenues (Hellman and Puri, 2000). Their investments tend to be focused on a narrow range of industries (Gans and Stern, 2003), on firms located in close proximity to their offices (Lerner, 1995) and startup hubs (Chen et al., 2014), on sectors where they have previous investment experience (Sorenson and Stuart, 2001), and where they expect follow-on capital to be available (Nanda and Rhodes-Kropf, 2013). They also prefer to invest in teams with a strong track-record (Gompers et al., 2016) and in serial entrepreneurs (Gompers, Lerner, and Sharfstein, 2005; Gompers et al, 2010). Interestingly, in a recent paper, whereas Nanda, Samila and Sorenson (2018) find evidence of VCs being able to select good investments, they do not find evidence of them being able to correctly identify, ex-ante, the very top right tail outcomes. This speaks to both VCs' ability to identify key predictors of future firm growth, but also to the presence of residual uncertainty about the prospects of the high potential candidates that enter their portfolios.

On the other hand, while it is possible that venture capitalists do identify many of the firms with growth potential, the full population of growth firms may be quite different (as emphasized, among others, by Bhide (2000)). Differences between the process of selection into venture capital and the overall process of firm growth might be driven both by supply and by demand-side factors. On the supply side, it is possible that differences in regional and industry composition may result in a relatively lower rate of entrepreneurial activity for certain types of businesses. As well, if VCs face higher search costs outside of regional startup hubs or specific industries they have experience in, then some firms with high growth potential might be excluded from venture capital investment because they do not fall within the traditional VC 'search space'. On the demand side, firms that can bootstrap through other means and generate enough cash flow to sustain their growth may have little demand for venture capital to begin with and may want to avoid the loss of equity and control that is associated with raising external funding.

Understanding the process of selection into venture capital and how it relates to the broader process of firm growth—even in the absence of VC funding—matters for estimating how efficiently society allocates resources to new ventures, and for regional policies targeted at sustaining entrepreneurship and economic growth outside of startup hubs. If VC-backed firms and non-VC-backed firms are similar, but VC funding drastically increases the odds of firm growth, then from a policy perspective it is useful to examine the barriers to venture financing in regions or industries where it is lacking, and what can be done to remove them. It would also suggest that the prior literature's focus on venture capital has not overlooked an alternative, critical path to firm growth, but that instead venture capital is a critical accelerant of growth within a single growth 'playbook'. If instead VC-funded and non-VC-funded firms that achieve growth are fundamentally different (and need different types of resources, investors and policies), then efforts targeted at expanding venture capital to these different types of firms, sectors, and regions, may be completely ineffective at accelerating them, and different types of interventions may be needed to support their alternative path to an equity growth outcome. Empirically, to adjudicate between these competing hypotheses, we need to develop a methodology which allows us to identify the growth potential of firms at founding – irrespective of future funding source – and systematically compare firm characteristics and growth outcomes between these possibly different paths to growth. Our next section presents in detail a predictive analytics approach which, by leveraging the information contained in VCs funding decisions, helps us make progress in this direction.

III. A PREDICTIVE ANALYTICS APPROACH FOR STUDYING THE PATHS TO FIRM GROWTH

To break through this impasse, we develop a predictive analytics approach that allows us to take advantage of the process of selection into venture capital to identify firm characteristics that are predictive of future growth. We then use the resulting 'VC-likelihood' estimate to study growth within the sample of firms that *do not* receive venture capital. Our goal is to estimate the relationship between an informed signal of growth potential (i.e., receiving venture capital), early firm characteristics and founder choices, and the resulting probability of growth for all firms in the economy.

Building on Guzman and Stern (2015, 2017, 2019), our approach takes advantage of three interrelated insights. First, a practical requirement for any entrepreneur trying to achieve a growth outcome is business registration (as a corporation, partnership, or limited liability company). This practical requirement allows us to form a population sample of entrepreneurs "at risk" of growth at a similar (and foundational) stage of the entrepreneurial process. Second, we are able to distinguish among different types of business registrants through the measurement of characteristics related to entrepreneurial quality observable *at or close to the time of registration.* For example, we can capture firm characteristics such as whether the founders name the firm after themselves (eponymy), whether the firm is organized in order to receive 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 both a signal of an informed investor's willingness to invest in a firm with growth potential (receipt of venture capital) as well as equity growth outcomes such as IPOs or acquisitions for all firms in our sample.

We combine these insights to develop a predictive analytics model that leverages the fact that venture capital is an informed (although imperfect) signal of growth potential to characterize the potential of firms that *do not* receive venture capital. In particular, we begin by estimating a predictive analytics model of the process of selection into venture capital. Specifically, for a firm *i* at time *t*, with startup characteristics $H_{i,t}$, we observe the receipt of venture capital $VC_{i,t+s}$ s years after founding and estimate:

$$\phi_{i,t} = P\left(VC_{i,t+s} \middle| H_{i,t}\right) = f\left(\alpha + \beta H_{i,t}\right) \tag{1}$$

This model allows us to *predict* quality as the probability of receiving venture capital given the focal startup characteristics at founding, and estimate a 'VC-likelihood' (a proxy for quality and potential as assessed by venture capitalists) as $\hat{\phi}_{i,t}$. We use these estimates to characterize whether the same startup characteristics of firms that *do not* receive venture capital are similarly informative for achieving an equity growth exit within the non-VC-backed sample. Specifically, from (1), we are able to form an estimate of the 'VC-likelihood', $\hat{\phi}_{i,t}$, and then consider how informative this estimate is within a regression where we estimate the probability of growth among firms *that do not receive venture capital:*

$$g_{i,t+s} = \alpha + \beta\hat{\phi}_{i,t} + \epsilon_{i,t} \text{ if } VC_{i,t+s} = 0 \tag{2}$$

To the extent that the estimate in (2) is informative (i.e. to the extent that determinants of growth within VC-backed firms are similar to the determinants of growth within the non-VC sample), we can also use our 'VC-likelihood' estimates to construct matched sample control groups to evaluate the returns to venture capital on equity growth itself, and separate the role of selection into venture capital (based on observables), from treatment:

$$g_{i,t} = h\left(VC_{i,t+s} \middle| \hat{\phi}_{i,t}\right) \tag{3}$$

IV. DATA AND DESCRIPTIVE STATISTICS

Our analysis uses business registration records, which are public records created when an individual registers a new business as a corporation, LLC or partnership (Guzman and Stern, 2015; 2017; 2019).[7] We rely on all registrations from 1995 to 2005 in 49 US states and Washington D.C,[8] representing virtually the totality of the United States venture capital. We

[7] This section draws heavily from this prior work, where we introduce business registration records and many of the measures used in this paper.
[8] We only exclude firms that are local to the state of Delaware which, due to the unique role Delaware jurisdiction plays in the United States, we are not able to differentiate statistically from high growth startups.

build this dataset from the underlying data in the Startup Cartography Project (Andrews et al, 2019), a systematic effort to measure entrepreneurship in the United States. Portions of this dataset have also been used in our prior work (e.g., Guzman and Stern, 2019; Guzman, 2018; Fazio et al, 2019; Guzman and Kacperczyk, 2019). Relative to these prior studies, we have expanded the data in two dimensions. First, we increased the coverage from 32 U.S. states to 49 U.S. states plus Washington D.C., allowing us to virtually cover the whole U.S. economy, and, second, we have added venture capital databases allowing us to develop measures of venture capital financing for all firms within our data.

While it is possible to found a new business without appearing in these data (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 customers. Furthermore, all corporations, partnerships, and limited liability companies must register with a Secretary of State[9] in order to take advantage of these benefits, as 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. Our analysis draws on the complete population of firms satisfying one of the following conditions: (a) a for-profit firm in the local jurisdiction or (b) a for-profit firm whose jurisdiction is Delaware but whose principal office address is in the local state. In other words, our analysis excludes non-profit organizations as well as companies whose primary location is not in the state. The resulting dataset contains 13,231,305 observations.[10] For each observation we construct variables related to: (a) growth outcomes (IPO or significant acquisition); (b) venture capital financing events; (c) firm characteristics based on business registration observables; and (d) firm characteristics based on external data that can be directly linked to the firm (e.g. patents, trademarks). We briefly review each one in turn.

Growth Outcomes. The growth outcome used in this paper, *Growth*, is a dummy variable equal to 1 if the firm has an initial public offering (IPO) or is acquired at a meaningful

[9] Or Secretary of the Commonwealth.

[10] The number of firms founded in our sample is substantially higher than the US Census Longitudinal Business Database (LBD), done from tax records. For example, for Massachusetts in the period 2003-2012, the LBD records an average of 9,450 new firms per year and we record an average of 24,066 firm registrations. On the other hand, our number is lower than the total number of tax-paying entrepreneurs reported when including non-employer firms by Fairlie, Miranda and Zolas (2019). The number of business registrants thus seems to strike a middle ground between only employer companies (missing many startups), and all tax paying self-employed (which includes many non-startups).

positive valuation within 10 years of registration as reported in the Thomson Reuters SDC database. Between 1995 and 2005, we identify 17,494 firms that achieve growth, representing 0.13% of the total sample of firms.

Venture Capital Financing. We collect information on Series-A venture capital financing events from multiple databases: AngelList, CapitalIQ, Preqin, and Thompson Reuters VentureXpert, from which we create two measures., *Gets Venture Capital*, is a dummy equal to 1 if a firm receives financing and 0 otherwise, and *Gets Venture Capital in 2 Years* is only equal to 1 if the firm raises this financing in the first two years[11]

Firm Characteristics. We develop two types of firm characteristics: (a) those based on business registration data, and (b) those based on external indicators of quality that are observable at or near the time of business registration.

a. Measures based on business registration data. In the first category, we first create two binary measures that relate to how the firm is registered: *Corporation*, which captures whether the firm is a corporation rather than an LLC or partnership, and *Delaware*, equal to one if the firm is registered in Delaware. We then create five additional measures based directly on the name of the firm. *Eponymous* is equal to 1 if the first, middle, or last name of the top managers is part of the name of the firm itself.[12] Our last measure relates to the structure of the firm name. Based on our review of naming patterns of growth-oriented startups 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 the organizational form, e.g. "Inc."). We define *Short Name* to be equal to one if the entire firm name has three or less words, and zero otherwise.[13] We then create several measures based on how the firm name reflects the industry or sector within which the firm is operating, taking advantage of the industry categorization of the US Cluster Mapping Project ("US CMP") (Delgado, Porter, and Stern, 2016)

---

[11] As shown in Figure A1, about 70% of all firms that raise venture capital financing do so within the first two years from their founding date.

[12] Belenzon, Chatterji, and Daley (2017, 2019) perform a more detailed analysis of the interaction between eponymy and firm performance finding an important negative relationship between an intent to use equity financing and eponymy.

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

and a text analysis approach. We develop seven such measures. The first three are associated with broad industry sectors and include whether a firm can be identified as local (*Local*), traded (*Traded*) or resource intensive (*Resource Intensive*). The other five industry groups are narrowly defined high technology sectors that are typically associated with high growth firms, including whether the firm is within the biotech (*Biotech Sector*), e-commerce (*E-Commerce*), other information technology (*IT*), medical devices (*Medical Devices*) or semiconductors (*Semiconductor*) space.

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

Descriptive Statistics. Table 1 reports summary statistics. There are 13,231,305 firms in our data: 0.13% of these firms achieve an equity growth outcome within ten years from incorporation, and 0.08% receive venture capital (0.05% within 2 years)[14]. 0.2% of firms have a patent (within 1 year from birth), and 0.08% have a trademark. 57% of firms are corporations, 47% have a short name, 10% are eponymous and 2.6% are registered in Delaware.

In Table 2, we directly compare the share of firms in our sample that grow with versus without VC: although only 0.15% of non-VC-funded firms achieve growth, this share is 34% for VC-funded firms (Panel A). If we only focus on IPOs, whereas 1 out of 22 VC-funded

[14] This number of investments is not comparable with the number of investments in these states within those years for at least three reasons. First, we only include firms registered after 1995, but investments occurring in the early part of our sample could be on firms registered earlier than 1995, which we do not observe. Second, we only include *local* firms, but some regions such as Silicon Valley or Boston, have a history of firms that are not local but instead move to these locations after receiving venture capital financing, and might receive follow-on financing in these regions. For example, many Israeli firms move to the United States after receiving their first round of financing. Our dataset is designed to exclude these firms. Finally, naturally, our matching cannot be perfect. While we have applied to matching improvements developed by Balasubramanian and Sivadasan (2008) and Kerr and Fu (2008), our focus has intently been on avoiding as many false-positives as possible. We have high confidence that the investments we observe reflect the true investment as stated in Thompson Reuters VentureXpert. Through manual checks, we do not believe the number of false-positives to be many.

firms achieves this milestone, among the remaining firms it is only 1 every 10,000 firms. Similarly, while approximately 1 out each 3 VC-funded firms is successfully acquired, only 1 out of 700 non-VC-funded firms does so.

V. WHAT THE PROCESS OF SELECTION INTO VENTURE CAPITAL REVEALS ABOUT THE PROCESS OF GROWTH IN THE ABSENCE OF VENTURE CAPITAL

We now proceed to use predictive analytics to characterize the process of selection into venture capital, and then estimate the likelihood of receiving venture capital financing for all business registrants (including firms that never received VC funding). Following the approach outlined in Section III, in Table 3 we estimate a predictive model that relates observables at founding to ex-post VC financing. The observables we use in the logit regressions are the same as Guzman and Stern (2015, 2017, 2019), although the dependent variable (*Gets Venture Capital)* in this case is equal to 1 if the firm receives VC funding, and 0 otherwise. To avoid overfitting and ensure the robustness of our estimates of 'VC-likelihood', in this first step we use a random, 50% subsample of the data. This allow us to calculate an 'out of sample' VC-likelihood for the excluded, 50% of firms (including those that do not receive VC-funding), which we will base the rest of our analysis on. For ease of interpretation, coefficients are presented as incidence rate ratios.

Table 3 explores the correlation between business registration observables and selection into venture capital financing within a random, 50% sample of all firms (our 'training set' for studying the determinants of venture capital). Column 1 does not control for intellectual property variables nor industry characteristics, which are respectively introduced in Columns 2 and 3. In Column 1, firms registered as corporations are 5.6 times more likely to receive VC funding, firms with a short name are 5 times more likely to reach the same milestone, eponymous firms are 87% less likely to be VC-backed, and Delaware registered firms are 30 times more likely to get VC capital. Column 2 relates only intellectual property measures to venture capital: Consistent with VCs selecting ventures of high quality and that have secured (or are in the process of securing) intellectual property protection, firms with a trademark are 3 times more likely to attract VC, and firms with a patent are 78 times more likely to do so.

Column 3 represents our main predictive model, which brings all of our observables together. While we could use a more flexible and complex functional form (and improve our predictive performance), we opt for a simple one to allow for easy interpretation of our next sets of results, and a higher degree of transparency on what the predictive model is based on. In Column 3, corporate form observables are quite informative of whether a firm will receive VC or not: Corporations are 4.3 times more likely to raise venture capital than non-corporations[15], and Delaware firms are 22 times more likely to raise VC. The naming choices of the firms are also predictive of future funding source. Firms with a short name are 3.6 times more likely to raise VC, and eponymous firms are 84% *less* likely to raise VC[16]. Firms with intellectual property are also more likely to raise VC. Firms with a trademark are 102% more likely to raise VC, and firms with a patent are 38 times more likely to receive VC financing. Firms that hold both a patent and are Delaware incorporated are significantly more likely to receive venture capital - 140 times more than other firms. The name-based industry coefficients are also significant, and sectors typically associated with the VC 'search space', such as IT, biotechnology, e-commerce, medical devices, and semiconductors are all more likely to receive venture capital financing.

We define *VC Likelihood* as the predicted probability resulting from this regression. This estimate is a useful summary statistic for how similar a specific firm is to other 'VC-type' firms.It is also a proxy of firm quality from a venture capitalist's perspective. Since we have the observables it is based on for the full set of incorporated firms, we can calculate the *VC Likelihood* also for firms that never received VC (either because they were rejected by professional investors, or because they never tried to raise from a VC firm in the first place). Furthermore, the measure is independent from the relative availability of VC in the region or time period the firm is created in.

Of course, VCs observe substantially more information than us when screening candidates for investment, as our approach only captures quality on dimensions that are public around the time of incorporation. If VCs predominantly select firms on measures of quality that are unobservable to us, then our *VC Likelihood* estimate would not be able to perfectly

[15] Though it might seem counter-intuitive that *any* venture-backed firm is not a corporation, the data during the late 1990's does include several LLCs that received venture capital financing.

[16] The negative effect of eponymy in the financing dynamics of firms is explored more systematically by Belenzon, Chatterji, and Daley (2017).

separate, at birth, VC-backed firms from other firms, as our observables would be too noisy of a predictor for future VC investment. We would still expect to see more VC-funded firms for higher levels of observable *VC Likelihood*, but the relationship could be possibly very noisy. Notwithstanding these limitations, the measure, is highly informative: In Figure 1, we plot the share of venture-backed firms in each of twenty, 5 percent bins in the distribution of predicted *VC Likelihood*. To avoid overfitting, we estimate this through a 10-fold cross validation approach: we separate our sample into 10 random groups, and calculate this summary statistic ten times, each time with one of these 10 groups as the out of sample group and the other 9 as the ones with which the model is built. This is the preferred testing approach in machine learning applications, since it also allows all data-points to be included in the test only once. The maximum, minimum, and mean of this statistic across the quality distribution are reported inside each bar. The distribution is highly skewed and the predictive capacity of our estimate is significant: 77% of all venture-backed firms are in the top 5% of the *VC Likelihood* distribution, and 57% are clustered in the top 1%.

Together, these results highlight three key findings. First, when looking at population-level data, VC activity is disproportionately concentrated on the right tail of the observable quality distribution that can be built, from the perspective of a venture capitalists, using firm characteristics at the time of founding. Whereas it is known that VCs invest in high quality firms, from a policy perspective it is interesting to benchmark these firms to the broader population in their respective regions and sectors. Second, our simple model based only on observables around birth is clearly able to effectively separate firms with some possibility of raising VC from the vast majority of incorporated firms. Third, *some* VC investment takes place all the way down to the 50th percentile of the quality distribution, suggesting that professional investors may be screening on dimensions that are sometimes not visible to the econometrician.

The *VC Likelihood* – a measure of firm quality and potential from the VC perspective – can also be used to identify how likely it would have been for a firm to raise VC, independent of the funding actually taking place. To begin testing the relationship between growth within the non-VC-backed firm sample and the *VC Likelihood*, we repeat the out of sample cross validation procedure but use non-VC-backed growth outcomes as our dependent variable. The resulting estimates, reported in Figure 2, tell us where in the distribution of *VC Likelihood* are the non-VC-funded firms that ended up achieving an IPO or significant acquisition. Similar to the findings illustrated in Figure 1, the relationship between *VC Likelihood* and growth outcomes is substantial: 43% of non-VC growth firms are in the top 5% of VC likelihood, and 52% in the top 10%.

We further explore this relationship within a regression framework in Table 4, where we compare the role of the *VC Likelihood* in predicting non-VC-funded growth in the 50% test sample that was not used to build the original *VC Likelihood* model (Table 3). The regression uses *Growth* as the dependent variable and includes state and year fixed effects. All VC-backed firms are excluded from the sample. Standard errors are clustered at the level of state-year pairs to better account for unobserved local factors in cohorts of firms. In Column 1, we introduce the *VC Likelihood* in an OLS regression. The coefficient is 0.6 and significant. Given a mean of *Growth* of 0.0013, and a standard deviation of *VC Likelihood* of 0.008, the effect is substantial. Increasing the predicted value of VC by one standard deviation increases the probability of growth by about 3X. In Column 2 we instead consider the log VC likelihood. Columns 3 and 4 instead consider a logit model and report the incidence rate ratios of the coefficients. Column 3 shows an IRR of 2.0, implying there is a one-to-one mapping of sorts between the two: a doubling of the likelihood of VC relates to a similar 100% increase in the probability of firm growth. Finally, we consider differential probability of growth across our distribution in Column 4 by using indicators for different ranges of the VC likelihood. We observe the that the big difference between the likelihood of growth occurs at the very top end of our distribution. While firms between the 95th and 99th percentile are only 10 times more likely to growth, firms between the 99th and 99.5th are 36 times more likely to grow, those in the 99.5th to 99.9th range are 83 times more likely to grow, and those in the top 0.1% are an impressive 355 times more likely to grow. While considering the top the top 0.1% of our test sample could seem like a slim group, this group still represents 13,232 firms. This is about the same as the total number of venture backed companies during our time period.

To further unpack the relationship between the *VC Likelihood* and growth in the non-VC sample, in Table 5 we compare the role the different observables play in predicting growth (Columns 2 to 4) relative to the role they play in predicting venture capital financing (Column 1). The main comparison of interest is between Columns 1 and 4. In general, we see the same sign and statistical significance across coefficients, though we do observe them attenuated for the equity grow outcomes. Perhaps the largest differences are in the predictive role of being a corporation and of holding a trademark. The importance of being a corporation is significantly lower for explaining non-VC growth, suggesting that, though firms may benefit from the stronger corporate governance tools offered by this incorporation form, a large portion of the benefit might be related to the ability to sell shares to investors. We also see a much higher importance of having an early trademark (an indicator that the firm is planning to commercialize a product or service) for growth without venture capital, an effect consistent with these firms bootstrapping through sales. The role of naming appears to be different, with short names predicting VC financing much more closely than equity growth, though it is unclear if this reflects differences in VC preferences or in the underlying types of firms and industries represented by each group. More interestingly, the role of Delaware jurisdiction and patenting — the two indicators with the most predictive power in both regressions — is surprisingly similar across specifications. Firms with ideas that can be protected through intellectual property rights, and firms that seek the more flexible (but also more expensive) protection of Delaware incorporation are substantially more likely to both receive venture capital financing, and achieve equity growth even in the absence of VC funding, reflecting large similarities in the at-birth observables of firms across these two groups.

Last, in Table 6 repeats our estimate of the association between the log-odds of *VC Likelihood* and equity growth outcomes for non-VC-backed firms across different geographies and time-periods. The coefficient is stable and similar across all columns, suggesting that the relationship we have identified between observables at the time of incorporation, how VCs interpret them, and ultimately firm growth within a sample of firms that never raised VC, holds across very different types of regions and time periods.

VI. THE RETURNS TO VENTURE CAPITAL FINANCING

We now turn to studying the full population of firms and the role venture capital financing plays in their ability to reach a growth outcome (IPO or acquisition). In Table 7, our objective is to partially separate VC selection (on observables) from treatment using the estimates from our predictive analytics approach. Raising venture capital is an informative signal of quality: as we have seen in Section V, VCs select startups that are on the extreme right tail of the observable *VC Likelihood* distribution. They also contribute to the success of the firms they invest in by providing capital, offering mentorship, performing monitoring, helping or replacing founding teams, connecting firms to possible customers and suppliers etc. Hence, in the absence of exogenous variation, any estimate of the correlation between VC funding and growth will always be a composition of the VCs' role in the selection of higher quality firms as well as in increasing their chances of success. Both effects will also vary with the underlying quality of the investors involved, as high quality VCs will not only see better deals, but may also provide better support to their portfolio companies (e.g. through their networks, etc.).

To account for selection on observables, we rely on our approach to deliver us a proxy for a firm's quality and potential — from the perspective of VCs — around the time of birth. As we are unable to fully control for firm differences (since VCs also select firms based on variables that are unobservable to us), accounting for such a measure when estimating the association between VC and growth should return an upper bound on the VC treatment effect (as we are likely underestimating selection).

Before introducing the summary measure directly, in Table 7 we progressively add the controls we used so far in Columns 1 to 4. In each column, we perform logit regressions with *Gets VC in 2 Years* (a binary measure indicating whether a firm receives VC financing within the first two years)[17] as the main independent variable, and our binary outcome measure *Growth* — achieving an equity growth outcome in 10 years— as the dependent variable. Results are reported as incidence rate ratios, and standard errors are clustered at the state-year pair level. All the remaining tables in the paper only use the 50% random test subsample we did not use to develop our predictive approach.

Column 1 of Table 7 compares VC-funded firms to non-VC-funded firms within the subsample. The probability of growth for firms that raise venture capital is 405times higher than that of a random firm in the sample. Selection is obviously a major concern here, as the vast majority of firms in the sample have an extremely low probability of achieving an IPO or

[17] As documented in Appendix Figure A2, the majority of firms that eventually raise VC do so within 2 years (about 25% receives financing within 3 months, 56% within a year, 75% within 2 years). The short time-frame between firm birth and VC financing motivates our choice to focus the rest of our analyses on receiving a series A investment within two years, which has the additional benefit of allowing us to evaluate firms across time without running into truncation issues.

acquisition in the first place, and therefore are not a credible control group for VC funded firms. Adding state and year fixed effects, and controlling for traditional proxies for quality such as the presence of patents and trademarks reduces this estimate by an order of magnitude in Column 2. Nonetheless, VC-funded firms are still 78 times more likely to grow than non-VC-funded firms. Interestingly, the introduction of our basic firm observables revealed at incorporation in Column 3 leads to a sizable reduction in the coefficient, bringing VC firms fairly close to firms of comparable characteristics in terms of outcomes. Once all our measures are accounted for in Column 4, VC-funded firms are only 16 times more likely to grow than other firms. Column 5 instead includes the predicted *VC Likelihood* as a fourth order polynomial directly, rather than the observables themselves. The coefficient is reduced slightly and VC-funded startups are now 13 times more likely to grow.

Column 6 is our preferred estimate. In this column we extend approach for separating selection on observables from treatment by performing an exact matching procedure.[18] For each VC-backed firm, we randomly select a non-VC-backed firm founded in the same year and geographic region, with the same exact value of observable *VC Likelihood*. The matching is at the same zip code level for 86% of firms, with remaining firms matched at the MSA and state level. After matching, we estimate the differences in the odds of achieving an equity growth outcome between our 'treated' firms (i.e. the firms that received VC funding) and our 'control' firms (i.e. firms that did not raise VC funding, but that have exactly the same *VC Likelihood* of doing so at birth). The incidence rate ratio drops from 13.8 to 7.0. This estimate is significant: conditional on the *VC Likelihood*, firms that raise VC are still 6 times more likely to achieve an equity growth outcome than non-VC-funded firms. However, the coefficient is also two orders of magnitude lower than the original, naïve estimate from Column 1. While the implied role of venture capital on firm performance is meaningful, 98% of the difference in outcomes between VC-backed and non-VC-backed firms is accounted for by characteristics that are observable at founding. Interestingly, our estimate is comparable to those of Chemmanur et al. (2011) and Puri and Zarutskie (2002), even though there are important differences in the specifications and samples we use since we start from the full population of incorporated firms.

[18] Imbens and Rubin (2015) recommend using exact matching on propensity score estimators to improve balance.

Table 8 extends the previous table by introducing a series of additional fixed effects to control for regional and microgeographic heterogeneity in our matching estimator. Consistent with the idea that unobservables may be less of a concern after we perform our matching on *VC Likelihood*, adding state-year pair fixed effects, MSA fixed effects, or controls for the average quality of the zip code level neighbors of the focal firm does not change our estimates: VC-funded firms continue to be approximately 6 times more likely to grow than their counterparts, irrespective of which controls we introduce.

It is important to stress that the estimate based on matching is still likely to be an upper bound on the true effect of VCs on firm growth, as the firms in our sample are still likely to differ on unobservable quality. Nevertheless, given the informational imbalance between the VC partners actually making the investment decisions and our regressions, it is surprising to see how much of the variance in outcomes we are able to explain. Furthermore, our *VC Likelihood* measures are defined many years before the actual acquisition or IPO takes place, i.e. when the uncertainty surrounding a startup is still extremely high.

Taken together, results from Tables 7 and 8 highlight just how much of the initial difference in the probability of growth between VC-funded-firms and other firms is driven by selection. Whereas in the most naïve estimation VC-funded firms are 405 times more likely to grow than other firms, this premium is reduced to only 6 times using our matching approach. This is consistent with our descriptive results on selection presented in Figures 1 and 2, and confirms that VCs select firms that are already of very high quality based on observables.[19] Our exercise places an upper bound on how much value, on average, VCs may be adding to the firms they invest in.

In Table 9, we re-estimate our model for firms on the right tail of the observable *VC Likelihood* distribution (firms in the top 5%, 1%, 0.1% and 0.05%). As we move up our quality distribution, the marginal contribution of VCs to growth is drastically reduced, possibly because for these right tail firms we do have a better measure of actual growth potential (i.e. our information gap relative to the VCs is smaller). For firms that exhibit extremely high, observable quality at incorporation (Column 5), VC-funded firms are only 2.3 times more

[19] In terms of the type of growth outcomes we observe, VCs are associated with a larger increase in the probability of an acquisition than in the probability of an IPO, which is consistent with them supporting their portfolio firms in the search for potential buyers through their professional network.

likely to grow than similar firms that do not receive VC funding. This group, however, represents a sizable share of all VC funded firms (34%).

Last, in Table 10 we divide the sample by startup hubs versus not (Columns 2 to 4), and over economic cycles (Columns 5 to 7). Estimates for the role of VC are higher outside of hubs, where VC-funded firms are 9 times more likely to grow than their counterparts, and when follow-on capital is more likely to be available (as in the .com boom period). The first effect suggests that the marginal VC-funded company in a non-hub region may be of higher quality than the marginal VC-funded firm in a hub, and is consistent with the results Catalini and Hui (2017) find when looking at US equity crowdfunding investments.

VII. CONCLUSION

Our results support the presence of multiple alternative paths to startup growth, but also emphasize a common profile for high potential firms which is independent of funding source. Though a large portion of firms grow without venture capital, many characteristics of these startups are strikingly similar to the characteristics of startups that are typically selected by VCs. Almost 50% of the firms that never raise VC are in the top 5% of our estimated *VC Likelihood* distribution, and non-VC-backed firms in the top 1% of the same estimate are over 350 times more likely to achieve an equity growth outcome (compared to the bottom 50%). Our estimates of the 'VC-effect', while inherently imperfect because of our inability to capture many of the firm and founder characteristics VCs observe through their due diligence and screening process, place an upper bound on the contribution of VCs to growth. In our matched sample estimates, VC funded firms are 6 times more likely to grow than non-VC-funded firms of comparable quality. Furthermore, when we focus on the right tail of the estimated *VC Likelihood* distribution, the 'VC-effect' is substantially reduced: firms in the top 0.05% of our quality measure at birth, are only 2.3 times more likely to grow with VC funding than without it.

Our effort to have comprehensive coverage of all firms in the economy has necessitated some tradeoffs. For example, our study has focused on relatively coarse binary measures of financing and equity outcomes. We fully recognize that the terms of the financing events, the underlying firm valuations, and the size of each equity outcome, are all also relevant margins of study. We expect follow-on work to focus on these issues more clearly.

Finally, an obvious follow-on question from our results, given the striking similarity between firms that grow with and without, is to ask what share of all growth outcomes in the economy are from startups that fit a common growth profile. In our companion paper (Catalini et al, 2019) we target this issue by developing a simple structural model that ties growth intention to founding choices, and use these to assess the relative importance of two modes of startup growth. The first, a passive growth model—such as random growth (Gibrat, 1931) and passive learning (Jovanovic, 1982)—assumes all startups start with similar growth orientation, and startup outcomes are the result of self-selection based on realized performance. The second, an active growth model, instead considers entrepreneurs different in their intention and ability at founding—such as Ericson and Pakes (1995)—and growth is the result of active conscious investment by those startups that have the ability and intention to grow at founding. We find that at least half of all IPOs and acquisitions is active growth, a share that increases for locations with more VC-oriented firms and when we limit only to larger outcomes, such as valuations over $100 million dollars.

Overall, our findings highlight the importance of selection in accounting for the perceived contribution of venture capital to startup growth. Given how simple the observables from our prediction model are, their public nature, and the fact that they are collected many years before an exit event, it is striking to see how much they explain of the process of selection into VC and startup growth both for VC-backed and non-VC-backed firms. Further exploring how these alternative paths to growth differ, should be a fruitful research area for scholars interested in how society allocates capital to novel, high potential ideas and converts them from ideas to massively scalable businesses.

REFERENCES

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. “The Startup Cartography Project”. Working Paper.

Balasubramanian, N, and J Sivadasan. 2008. “NBER Patent Data-BR Bridge: User Guide and Technical Documentation.” *Working Paper*. ftp://tigerline.census.gov/ces/wp/2010/CES-WP-10-36.pdf (May 18, 2014).

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 Chatterji, and Brendan Daley. 2017. “Eponymous Entrepreneurs.” *American Economic Review* 107(6): 1638-55.

Belenzon, Sharon, Aaron Chatterji, and Brendan Daley. 2019. “Choosing Between Growth and Glory”. *Management Science.* Forthcoming.

Bhide, A., 2000. *The Origin and Evolution of New Businesses*. Oxford University Press on Demand.

Catalini, Christian, and Hui, X. 2017 “Can Capital Defy the Law of Gravity? Investor Networks and Startup Investment”, Working Paper, MIT.

Catalini, Christian, Jorge Guzman, and Scott Stern. 2019. “Passive versus Active Growth: Evidence from Founder Choices and Venture Capital Investments”. *NBER Working Paper #26073.*

Chemmanur, Thomas J., Karthik Krishnan, and Debarshi K. Nandy. 2011. “How Does Venture Capital Financing Improve Efficiency in Private Firms? A Look beneath the Surface.” *Review of Financial Studies* 24(12): 4037–90.

Chen, Henry, Paul A. Gompers, Anna Kovner, and Josh Lerner. "Buy Local? The Geography of Successful Venture Capital Expansion." *Journal of Urban Economics* 67, no. 1

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

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

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

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jacknife." *The Annals of Statistics* 7(1): 1–26.

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

Ewens, Michael, Charles Jones, and Matthew Rhodes-Kropf. 2013. "The price of diversifiable risk in venture capital and private equity" *Review of Financial Studies.* 26 (8): 1854-1889

Ewens, Michael and Richard R. Townsend. 2019. "Are early stage investors biased against women?" *Journal of Financial Economics*. forthcoming

Fairlie, R. W., Miranda, J., & Zolas, N. 2019. "Measuring job creation, growth, and survival among the universe of start-ups in the United States using a combined start-up panel data set." *ILR Review*, *72*(5), 1262-1277.

Fazio, C., Guzman, J. and Stern, S., 2019. "*The Impact of State-Level R&D Tax Credits on the Quantity and Quality of Entrepreneurship.*" National Bureau of Economic Research.

Gibrat, Robert. 1931. *Les inégalités économoques.* Paris. Libraririe du Recueil Sirey.

Gans, J., & Stern, S. 2003. "When does funding research by smaller firms bear fruit?: Evidence from the SBIR program." *Economics of Innovation and New Technology*, *12*(4), 361-384.

Gompers, Paul, William Gornall, Steven N. Kaplan, and Ilya A. Strebulaev. 2016. "How do Venture Capitalists Make Decisions" *NBER Working Paper Series .#22587*

Gompers, P., Lerner, J. and Scharfstein, D., 2005. "Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999." *The journal of Finance*, *60*(2), pp.577-614.

Gompers, Paul A., Josh Lerner, David Scharfstein, and Anna Kovner. "Performance Persistence in Entrepreneurship and Venture Capital." *Journal of Financial Economics* 96, no. 1 (April 2010): 731–764.

Guzman, Jorge. 2018. "Go West Young firm: Agglomeration and Embeddedness in Startup Migrations to Silicon Valley". *SSRN Working Paper #3175328*

Guzman, Jorge, and Aleksandra Kacperczyk. 2019. "Gender Gap in Entrepreneurship" *Research Policy.* 48 (7): 1666-1680

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

Guzman, Jorge, and Scott Stern. 2017. “Nowcasting and Placecasting Entrepreneurial Quality and Performance.” *NBER/CRIW Measuring Entrepreneurial Businesses: Current Knowledge and Challenges conference* (December): 67.

Guzman, Jorge, and Scott Stern. 2019. “The State of American Entrepreneurship: New Estimates of The Quantity and Quality of Entrepreneurship for 32 US States, 1988-2014.” *American Economic Journal: Economic Policy*. Forthcoming`

Hellmann, Thomas, and Manju Puri. 2000. “The Interaction Between Product Market and Financing Strategy: The Role of Venture Capital.” *The Review of Financial Studies* 13(4): 959–84.

Hochberg, Yael, Alexander Ljungqvist, and Yang Lu. 2010. “Networking as a barrier to entry and the competitive supply of venture capital.” *The Journal of Finance* 65 (3): 829-859.

Hsu, DH. 2004. “What Do Entrepreneurs Pay for Venture Capital Affiliation?” *The Journal of Finance* LIX(4). http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2004.00680.x/full (September 2, 2013).

Hurst, Erik, and BW Pugsley. 2011. “What Do Small Businesses Do?” *Brookings Papers on Economic Activity* (Fall): 73–128. http://www.nber.org/papers/w17041 (May 8, 2014).

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jovanovic, Boyan. 1982. “Selection and the Evolution of Industry”. *Econometrica*. 50 (3):649-670.

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.

Kaplan, Steven N, and Per Strömberg. 2003. “Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts.” *The Review of Economic Studies* 70(2): 281–315. http://restud.oxfordjournals.org/lookup/doi/10.1111/1467-937X.00245.

Kaplan, Steven N, and Per Strömberg. 2004. “Characteristics, Contracts, and Actions: Evidence from Venture Capital Analyses.” *Journal of Finance* LIX(5): 2173–2206.

Klapper, Leora, Raphael Amit, and MF Guillén. 2010. “Entrepreneurship and Firm Formation across Countries.” *International Differences in Entrepreneurship*. Eds Josh Lerner and Antoinette Schoar. http://www.nber.org/chapters/c8220.pdf (May 18, 2014).

Kerr, William R., and Shihe Fu. 2008. “The Survey of Industrial R&D—patent Database Link Project.” *The Journal of Technology Transfer* 33(2): 173–86. http://link.springer.com/10.1007/s10961-007-9078-3 (July 8, 2014).

Kerr, W. R., Lerner, J., & Schoar, A. (2014). The consequences of entrepreneurial finance: Evidence from angel financings. *Review of Financial Studies*, *27*(1), 20-55.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1998. "Law and Finance." *The Journal of Political Economy* 106(6): 1113–55.

Lerner, Josh. 1995. "Venture Capitalists and the Oversight of Private Firms." *The Journal of Finance.* 50 (1): 301-318

Lerner, Josh. 2009. *Boulevard of Broken Dreams: Why Public Effort to Boos Entrepreneurship and Venture Capital Have Failed-and What to Do about It.* Reprint Ed. Princeton University Press.

Lerner, Josh, et al. 2015. *The globalization of angel investments: Evidence across countries*. No. w21808. National Bureau of Economic Research, 2015.

Marx, M., D. Strumsky, and L. Fleming. 2009. "Mobility, Skills, and the Michigan Non-Compete Experiment." *Management Science* 55(6): 875–89.

Nanda, R., & Rhodes-Kropf, M. 2013. "Investment cycles and startup innovation." *Journal of Financial Economics*, *110*(2), 403-418.

Nanda, Ramana, Sampsa Samila, and Olav Sorenson. 2018. "The persistent effect of initial success: Evidence from venture capital". No. w24887. National Bureau of Economic Research.

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): 2317–57.

Piacentino, Giorgia. 2019. "Venture Capital and Capital Allocation." *The Journal of Finance*. 74 (3): 1261-1314

Puri, Manju, and Rebecca Zarutskie. 2012. "On the Life Cycle Dynamics of Venture-Capital- and Non-Venture-Capital-Financed Firms." *The Journal of Finance* 67(6): 2247–2293.

Ritter, Jay. 2016. "Initial Public Offerings: VC-backed IPO Statistics Through 2015". Working Paper.

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

Sorenson, O., & Stuart, T. E. 2001. "Syndication networks and the spatial distribution of venture capital investments." *American journal of sociology*, *106*(6), 1546-1588.

SSTI. 2015. *Useful Stats: Venture Capital Investment Dollars, Deals by State, 2009-2014.* Url: http://ssti.org/blog/useful-stats-venture-capital-investment-dollars-deals-state-2009-2014. Accessed on: January 17, 2017.

Sorensen, Morten. 2007. "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital." *The Journal of Finance* 62(6): 2725–62.

TABLE 1
Summary Statistics

Measure	Source	Description	Mean	Std. Dev.
Firm Outcomes				
Gets Venture Capital	Multiple	Whether the firm receives any financing.	0.0008	0.028
Gets Venture Capital in 2 years	Multiple	Whether the firm receives any financing within 2 years of founding.	0.0005	0.021
Equity Growth (IPO or Acquisition)	SDC Platinum IPO and M&A.	Whether the firm has an equity growth event in the first 10 years.	0.0013	0.036
Business Registration Observables				
Corporation	Business Reg.	1 if a firm is a corporation (not an LLC or partnership)	0.565	0.496
Delaware	Business Reg.	If the firm's jurisdiction is Delaware	0.026	0.160
Short Name	Business Reg.	If the firm's name length is 3 words or less (including firm type (e.g. "inc."))	0.466	0.499
Eponymous	Business Reg.	Business Reg. If the president or CEO share the name of the firm.	0.097	0.296
Intellectual Property Observables				
Trademark	USPTO	If the firm acquires for a trademark within 1 year of founding.	0.0008	0.029
Patent	USPTO	If the firm acquires for a patent application within 1 year of founding.	0.0023	0.048
USCMP Name Based Industry Measures				
Industry Dummies	Business Reg.	If firm name is associated to an industry group (see Appendix for details).		
Observations			13,231,305	

This table represents our full dataset, comprised of all registered firms registered within the years 1995 and 2005 in 49 US states. These states account for 99.6% of US GDP and 95% of US venture capital investments in 2015. All measures defined in detail in Section III of this paper. Venture capital outcomes are taken for all firms reported in Thompson Reuters VentureXpert, Prequin, Capital IQ, and AngelsList. Business registration records are public records created endogenously when a firm registers as a corporation, LLC, or partnership. IP observables include both patents and trademarks filed by the firm within a year of founding, as well as previously filed patents assigned to the firm close to founding. All business registration observables, IP observables, and industry measures are estimated at or close to the time of firm founding. Further information on all measures can also be found in Guzman and Stern (2015), Guzman and Stern (2016), and Guzman and Stern (2017). Growth IPOs include only 'true' startup IPOs, we exclude all financial IPOs, REITs, SPACs, reverse LBOs, re-listings, and blank check corporations.

TABLE 2

Distribution of Equity Outcomes with and without VC

Panel A. Growth with and without VC

	Firms with VC	Firms without VC
Firms without Growth	6,597	13,200,904
(Share)	(65.6%)	(99.85%)
Firms with Growth	3,454	20,350
(Share)	(34.4%)	(0.15%)
Growth Split by IPO and Acquisition		
Share that IPO	4.50%	0.01%
Share that are Acquired	29.9%	0.14%

We perform an analysis of all firms that achieve IPO or acquisition (at any point) vs those that do not. IPOs are taken from SDC Capital and exclude all re-listings, reverse LBOs, SPACs, REITs, blank check companies, and financial IPOs.

TABLE 3

Determinants of Venture Capital Financing
Training Sample (50% Random Sub-Sample)
Logit model. Incidence Rate Ratios Reported

	Gets Venture Capital		
	(1)	(2)	(3)
Business Registration Observables			
Corporation	6.557***		5.277***
	(0.588)		(0.480)
Short Name	5.821***		4.559***
	(0.367)		(0.287)
Eponymous	0.128***		0.157***
	(0.0178)		(0.0223)
Delaware	30.96***		
	(3.461)		
Intellectual Property			
Trademark		4.291***	2.024***
		(0.771)	(0.338)
Patent		78.86***	
		(6.638)	
Patent Delaware Interactions			
Delaware Only			23.40***
			(2.215)
Patent Only			38.83***
			(2.742)
Patent and Delaware			141.4***
			(18.70)
USCMP Industry Dummies			
Local Industry			0.299***
			(0.0316)
Traded Industry			0.722***
			(0.0330)
Resource Intensive Industry			0.658***
			(0.0455)
IT			2.971***
			(0.155)
Biotechnology			3.785***
			(0.421)
E-Commerce			1.501***
			(0.0969)
Medical Devices			1.314***
			(0.0992)
Semiconductor			2.131***
			(0.381)
State F. E.	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	6,606,756	6606756	6606756
pseudo R-sq	0.292	0.190	0.349

This table reports a logit model estimating the determinants of firm growth for all firms without venture capital in a 50% random subsample of our data. Using firms without VC allows us to measure the possibility of growth independent of this input. We use this as a training sample for our predictive analytics model of entrepreneurial quality. Incidence rate ratios reported. Robust standard errors in parenthesis clustered at the state-year level.** $p < .01$, *** $p < .001$

TABLE 4

Does VC Likelihood Predict Equity Growth for non-VC Firms?
Dependent Variable: Equity Growth
50% Test Sample, excluding all VC-backed firms.

	(1) OLS	(2) OLS	(3) Logit	(4) Logit
VC Likelihood ($\hat{\mu}$)	0.602***			
	(0.0322)			
Log VC Likelihood ($\ln(\hat{\mu})$)		0.00101***	2.013***	
		(0.0000795)	(0.0205)	
Distribution of VC Likelihood				
Baseline: < 50%. ($\hat{\mu}$< .0001)				
50% - 90%				2.294***
($\hat{\mu} \in$ [.0001 , .0009])				(0.0863)
90% - 95%				5.260***
($\hat{\mu} \in$ [.0009 , .0020])				(0.286)
95% - 99%				12.09***
($\hat{\mu} \in$ [.0020 , .0072])				(1.181)
99% - 99.5%				38.66***
($\hat{\mu} \in$ [.0072 , .017])				(3.22)
> 99.5%				83.98***
($\hat{\mu} \in$ [.017 , .097])				(6.991)
> 99.5%				355.65***
($\hat{\mu} \in$ [.097 , .89])				(68.67)
State F.E.	No	Yes	Yes	Yes
Year F.E.	No	Yes	Yes	Yes
Observations	6610610	6601590	6601590	6610610
R-squared	0.017	0.003		
Pseudo R-squared			0.131	0.129
Log-Likelihood			-50283.8	-50420.3

VC Likelihood is the predicted probability that a firm gets venture capital given its characteristics at founding, predicted from (3-3). It has a mean of .0008 and a standard deviation of .0083. Standardized VC Likelihood changes this measure to have a standard deviation of 1. The mean value of the outcome variable is 0.0013. Robust standard errors in parenthesis clustered at the state-year level. *** p < .001.

TABLE 5

Determinants of Growth without Venture Capital
Training Sample (50% Random Sub-Sample)
Logit model. Incidence Rate Ratios Reported

	Gets Venture Capital	Equity Growth without Venture Capital Financing (VC-Backed firms Excluded)		
	(1)	(2)	(3)	(4)
Business Registration Observables				
Corporation	5.277***	2.040***		1.807***
	(0.480)	(0.0869)		(0.0724)
Short Name	4.559***	2.055***		1.873***
	(0.287)	(0.0717)		(0.0621)
Eponymous	0.157***	0.289***		0.328***
	(0.0223)	(0.0163)		(0.0185)
Delaware		15.24***		
		(0.805)		
Intellectual Property				
Trademark	2.024***		12.41***	6.030***
	(0.338)		(1.483)	(0.628)
Patent			31.47***	
			(1.648)	
Patent Delaware Interactions				
Delaware Only	23.40***			11.04***
	(2.215)			(0.611)
Patent Only	38.83***			14.28***
	(2.742)			(1.250)
Patent and Delaware	141.4***			72.48***
	(18.70)			(5.355)
USCMP Industry Dummies				
Local Industry	0.299***			0.395***
	(0.0316)			(0.0219)
Traded Industry	0.722***			1.174***
	(0.0330)			(0.0422)
Resource Intensive Industry	0.658***			0.926
	(0.0455)			(0.0457)
IT	2.971***			1.820***
	(0.155)			(0.0924)
Biotechnology	3.785***			1.966***
	(0.421)			(0.277)
E-Commerce	1.501***			1.198***
	(0.0969)			(0.0621)
Medical Devices	1.314***			1.228***
	(0.0992)			(0.0750)
Semiconductor	2.131***			1.837**
	(0.381)			(0.359)
State F. E.	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	6606756	6604792	6604792	6604792
pseudo R-sq	0.349	0.117	0.078	0.146

This table reports a logit model estimating the determinants of firm growth for all firms without venture capital in a 50% random subsample of our data. Using firms without VC allows us to measure the possibility of growth independent of this input. We use this as a training sample for our predictive analytics model of entrepreneurial quality. Incidence rate ratios reported. Robust standard errors clustered at the state-year level in parenthesis. ** p < .01 , *** p < .001

TABLE 6

Growth without Venture Capital
Logit Regression.
50% Test Random Sample

		Place Heterogeneity			Time Heterogeneity		
	All Firms	Silicon Valley	Startup Hubs	Non Startup Hubs	.com Boom 1995-Sept, 1999	.com Crash Sept, 1999-2001	Recovery 2001-2005
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log VC Likelihood ($\ln(\hat{\mu})$)	2.013***	2.082***	2.131***	1.949***	1.971***	2.098***	2.015***
	(0.0205)	(0.0517)	(0.0328)	(0.0208)	(0.0229)	(0.0420)	(0.0344)
State F. E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of Incorporation F. E.	Yes	Yes	Yes	Yes			
N	6601590	102185	732775	5867050	1709191	1078218	3795782
pseudo R-sq	0.131	0.211	0.226	0.105	0.124	0.159	0.115

Robust standard errors in parenthesis clustered at the state-year level. *** p < .01

TABLE 7

Venture Capital and Growth Outcomes Controlling for Observables and VC Likelihood
DV: Equity Growth: 1 if Firm Achieves IPO or Acquisition

	All Firms					Exactly Matched Sub-sample
	(1)	(2)	(3)	(4)	(5)	(6)
Gets VC in 2 Years	405.6***	79.09***	26.40***	16.64***	13.89***	7.011***
	(30.53)	(8.716)	(3.177)	(1.741)	(1.424)	(0.893)
Business Registration						
Corporation			2.111***	1.876***		
			(0.0874)	(0.0748)		
Short Name			2.156***	1.990***		
			(0.0733)	(0.0645)		
Eponymous			0.306***	0.338***		
			(0.0221)	(0.0250)		
Delaware			13.77***	9.754***		
			(0.773)	(0.543)		
Intellectual Property						
Patent		18.56***		6.226***		
		(1.197)		(0.349)		
Trademark		15.10***		6.150***		
		(1.334)		(0.540)		
USCMP Industry Dummies						
Local Industry				0.425***		
				(0.0247)		
Traded Industry				1.142***		
				(0.0444)		
Resource Intensive Industry				0.929		
				(0.0449)		
IT				1.695***		
				(0.0785)		
Biotechnology				1.830***		
				(0.242)		
E-Commerce				1.228***		
				(0.0553)		
Medical Devices				1.320***		
				(0.0741)		
Semiconductor				1.199		
				(0.259)		
Entrepreneurial Quality Controls						
Log VC Likelihood ($\ln(\hat{\mu})$)					0.995	
					(0.150)	
Log VC Likelihood ($\ln(\hat{\mu})$)^2					0.803***	
					(0.0430)	
Log VC Likelihood ($\ln(\hat{\mu})$)^3					0.978***	
					(0.00540)	
Log VC Likelihood ($\ln(\hat{\mu})$)^4					0.999**	
					(0.000194)	
State F. E.	No	Yes	Yes	Yes	Yes	Yes
Incorporation Year F.E.	No	Yes	Yes	Yes	Yes	Yes
N	6615653	6615653	6615653	6615653	6606633	5276
pseudo R-sq	0.070	0.136	0.176	0.200	0.194	0.140

Robust standard errors in parenthesis clustered at the state-year level *** p < .01

TABLE 8

VC Financing and Equity Growth Outcomes
Logit Regression on Matched Sample.

DV: Equity Growth Outcome: 1 if firm achieves IPO or Acquisition

	Baseline Model	Extra Controls		
	(1)	(2)	(3)	(4)
Gets VC in 2 Years	7.011***	7.504***	7.144***	7.001***
	(0.893)	(1.038)	(0.927)	(0.885)
Year F. E.	Yes		Yes	Yes
State F. E.	Yes			
State X Year F. E.		Yes		
MSA F. E.			Yes	
Control for Average Neighbor Quality				Yes
N	5276	4827	4683	5229
Pseudo R-sq	0.140	0.157	0.154	0.153

Matching approach uses exact quality values to match firms. All regressions run only on the 50% test sample not included in training the entrepreneurial quality model in Table 3. Some observations dropped when including State X Year Fixed Effects, MSA Fixed Effects, and average neighbor quality. Control for neighbor quality is natural log of the average quality of the ZIP Code excluding the focal firm. Matching algorithm matches each company that gets VC finance to another company with the same quality, born in the same year and ZIP Code. In about 20% of the sample, we do not find a match in the same ZIP Code and use a match in the same MSA instead. Incidence rate ratios reported.Robust standard errors in parenthesis clustered at the state-year level. *** p < .001

TABLE 9

Logit Regression, Matched Firms.
DV: Equity Growth Outcome. 1 if firm achieves IPO or Acquisition.

	All Firms	Within the Quality Distribution			
		Top 5%	Top 1%	Top 0.1%	Top 0.05%
	(1)	(2)	(3)	(4)	(5)
Gets VC in 2 Years	7.011***	6.506***	5.344***	4.004***	3.284***
	(0.893)	(0.836)	(0.785)	(0.771)	(0.699)
N	5276	4400	3313	1310	797
Pseudo R-sq	0.140	0.129	0.113	0.085	0.070

State fixed effects and year fixed effects included in all regressions. Matching algorithm matches each company that gets VC finance to another company with the same quality, born in the same year and ZIP Code. In about 20% of the sample, we do not find a match in the same ZIP Code and use a match in the same MSA instead. Incidence rate ratios reported. Robust standard errors in parenthesis clustered at the state-year level. *** p < .001

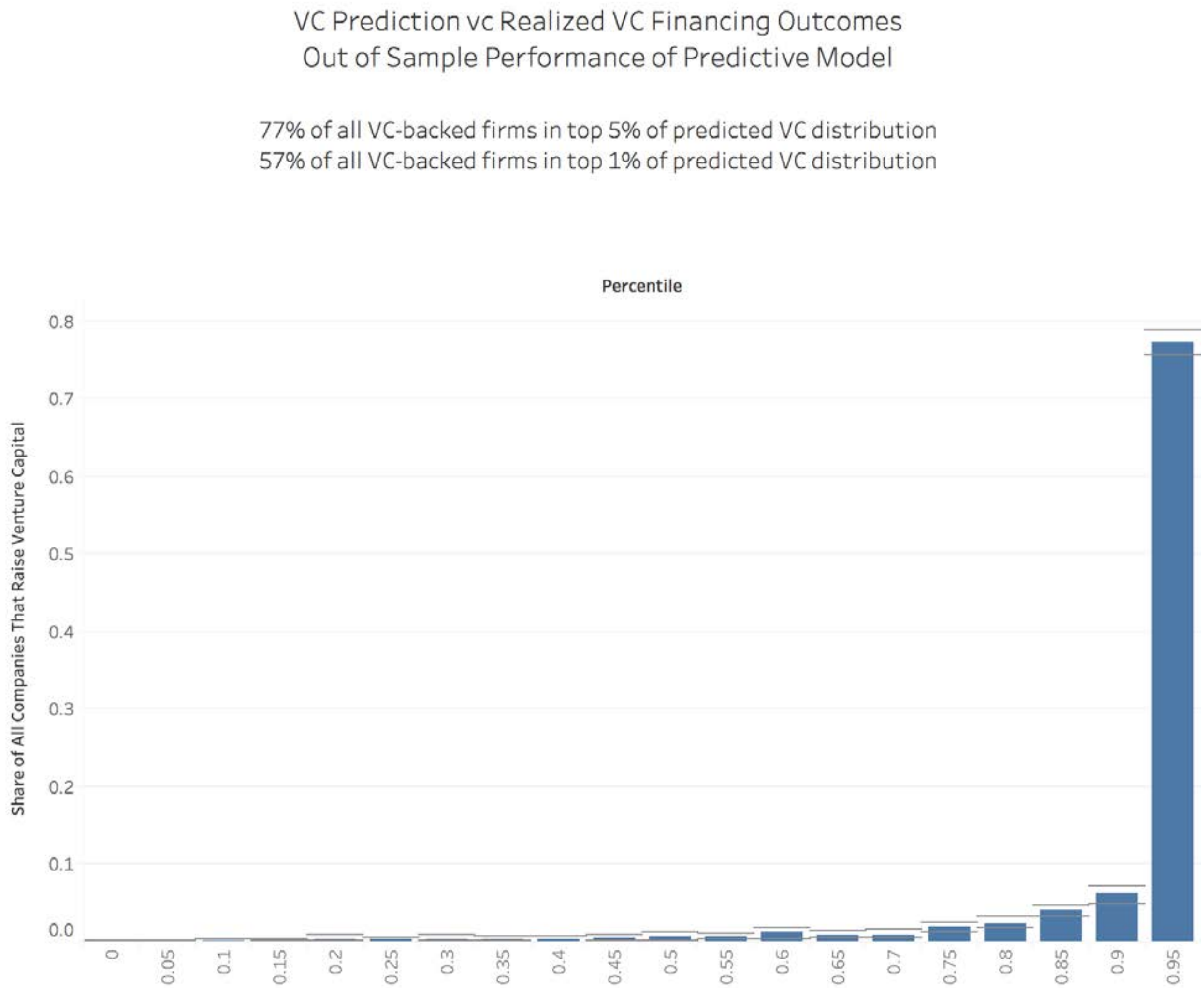
TABLE 10

VC Financing and Equity Growth Outcomes.
Logit Regression, Odds Ratios Reported.
Matched Sample. Heterogeneous Effects.
DV: Equity Growth. 1 if firm achieves IPO or acquisition

		Place Heterogeneity			Time Heterogeneity		
	All Firms	Silicon Valley	Startup Hubs	Non Startup Hubs	.com Boom Born: 1995-Sept, 1999	.com Crash Born: Sept, 1999-2001	Recovery Born: 2001-2005
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gets VC in 2 Years	7.011***	6.153***	6.463***	9.019***	9.539***	5.090***	8.691***
	(0.893)	(1.124)	(0.925)	(1.501)	(2.516)	(0.673)	(1.440)
State F.E.	Yes				Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes			
N	5276	1796	2778	2621	1402	2290	1447
Pseudo R-sq	0.140	0.125	0.125	0.134	0.171	0.109	0.157

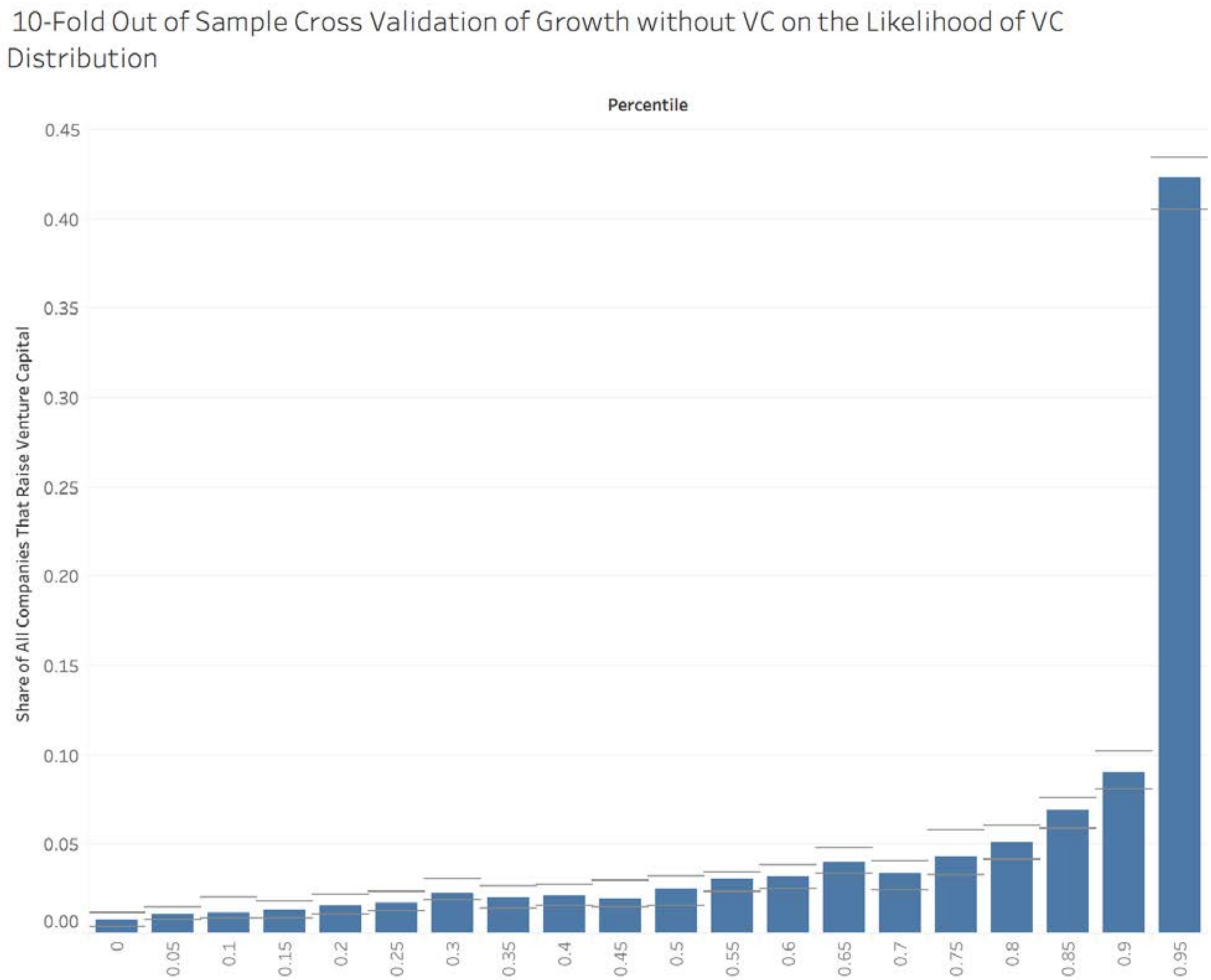
Robust standard errors in parenthesis. State fixed effects excluded from regressions that vary location, year fixed effects excluded from regressions that vary time, to allow differences in each dimension to show in the coefficient Matching algorithm matches each company that gets VC finance to another company with the same quality, born in the same year and ZIP Code. In about 20% of the sample, we do not find a match in the same ZIP Code and use a match in the same MSA instead. VC Quality only observed for California, Massachusetts, New York state, Texas, and Washington state. Incidence rate ratios reported. Robust standard errors in parenthesis clustered at the state-year level. *** p < .001

FIGURE 1



Note: We perform a 10 fold out of sample cross validation procedure to study the predictive capacity of our VC likelihood estimate. Bars indicate the average share of all out of sample VC-backed firms in different points of the predicted VC distribution, by 5 percent bins. Lines indicate the minimum and maximum estimate in this test.

FIGURE 2


10-Fold Out of Sample Cross Validation of Growth without VC on the Likelihood of VC Distribution
Percentile
Share of All Companies That Raise Venture Capital
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95

APPENDIX

Appendix A – Business Registration Requirements and Data

Business registration is the act of forming a firm as a corporation, limited liability company (LLCs), or partnership. In the process of providing financing, venture capitalists invest in registered businesses[20] by providing capital in exchange for ownership in the company[21]. The rights and obligations between the firm and the VC firm are then governed by the entity type, the jurisdiction, and the by-laws of each company. Being a shareholder in a corporation provides several benefits to venture capitalists relative to partnerships. In particular, minority shareholders have stronger rights in the corporation, which can be further augmented through provisions in the by-laws of the company, its operating agreements, or other contracts with the VCs. It also allows stricter governance. Finally, only corporations can be publicly traded companies, hence only corporations can exercise an initial public offering (IPO)—one of the main exit strategies for VCs. In the United States, the ability to exercise specific rules governing the VC contract depends on the state jurisdiction under which the firm operates. Due to a historical accident, there is a precedent of strong predictability of corporate law in the state of Delaware, and venture capitalist (as well as over half of all public companies) have a strong preference for firms registered under Delaware corporate law, even when this comes at an extra cost to the firm.

Entrepreneurs (and their lawyers) take these trade-offs into account as they convert their intentions for the firm into a legal structure. For example, they might prefer to register in Delaware if they expect to grow or seek VC financing. They also need to choose a name for the firm, whether to file for intellectual property protection through trademarks and patents, whether to be a corporation, partnership or LLC, etc. These choices are of strategic importance, and self-reveal part of the entrepreneur's ambition and own signal about the potential of the firm.

The timing of registration, while flexible, is influenced by similar considerations. While the cost of registration itself is low ($100 in California) and the process can usually be completed online in less than two hours, founders might struggle to register if they are not ready to choose a governance structure. As such, registration represents a moment in time when the core idea of the firm is developed enough to make these choices. Last, business registration is extremely useful in building a population-level dataset, as it is comprehensive and a necessary condition for equity

[20] This is an empirical fact rather than a theoretical requirement.

[21] In the case of LLCs and partnerships, purchasing ownership effectively make venture capital firms partners of the target firm. In the case of corporations, they become shareholders. While most venture capital investment occurs through the purchase of corporation shares, there are a few LLC companies invested on during the 1980s as well as the dot-com boom that where not corporations.

financing. This allows us to build a complete population of firms at risk of venture capital without selecting firms along idiosyncratic dimensions.

TABLE A1

Summary Statistics of industry measures

Measure	Source	Description	Mean	Std. Dev.
USCMP Name Based Industry Measures				
Local Industry	Business Reg.	If firm name is associated to a local industry.	0.191	0.393
Traded Industry	Business Reg.	If firm name is associated to a traded industry.	0.544	0.498
Resource Intensive Industry	Business Reg.	If firm name is associated to a resource intensive industry.	0.127	0.333
IT	Business Reg.	If firm name is associated to the IT industry cluster.	0.025	0.155
Biotechnology	Business Reg.	If firm name is associated to the Biotechnology industry cluster.	0.002	0.042
E-Commerce	Business Reg.	If firm name is associated to the E-Commerce industry cluster.	0.051	0.221
Medical Devices	Business Reg.	If firm name is associated to the Medical Devices industry cluster.	0.029	0.169
Semiconductor	Business Reg.	If firm name is associated to the Semiconductor industry cluster.	0.0004	0.020
Observations			13,231,305	

This table represents our full dataset, comprised of all registered firms registered within the years 1995 and 2005 in 49 US states. These states account for 99.6g% of US GDP and 95% of US venture capital investments in 2015. All measures defined in detail in Section III of this paper. Venture capital outcomes are taken for all firms reported in Thompson Reuters VentureXpert, Prequin, Capital IQ, and AngelsList. Business registration records are public records created endogenously when a firm registers as a corporation, LLC, or partnership. IP observables include both patents and trademarks filed by the firm within a year of founding, as well as previously filed patents assigned to the firm close to founding. All business registration observables, IP observables, and industry measures are estimated at or close to the time of firm founding. Further information on all measures can also be found in Guzman and Stern (2015), Guzman and Stern (2016), and Guzman and Stern (2017). Growth IPOs include only 'true' startup IPOs, we exclude all financial IPOs, REITs, SPACs, reverse LBOs, re-listings, and blank check corporations.

TABLE A2

Comparison of Means Between Growth, No Growth, VC Backed and Non VC Backed firms.

	No Equity Growth	Equity Growth	All		No Equity Growth	Equity Growth	All
Corporation				*Eponymous*			
No VC Financing	0.564	0.713	0.565	No VC Financing	0.097	0.042	0.097
VC Financing	0.869	0.950	0.897	VC Financing	0.015	0.012	0.014
All	0.564	0.748		All	0.097	0.038	
Delaware				*Patent*			
No VC Financing	0.025	0.254	0.026	No VC Financing	0.002	0.088	0.002
VC Financing	0.471	0.611	0.520	VC Financing	0.199	0.302	0.235
All	0.026	0.306		All	0.002	0.119	
Short Name				*Trademark*			
No VC Financing	0.466	0.652	0.466	No VC Financing	0.001	0.039	0.001
VC Financing	0.849	0.897	0.866	VC Financing	0.030	0.050	0.037
All	0.466	0.688		All	0.001	0.041	

TABLE A3

Share of firm in IPO and Acquisition Samples that Raise Venture Capital		
	IPO	Acquisition
Firms with VC Financing	452	3,002
(Share)	(31%)	(13%)
Firms without VC Financing	1,028	19,322
(Share)	(69%)	(87%)

Our estimates are based on firms *founded* between 1995 and 2005 in our sample of states that eventually IPO. Reitter (2015) estimates that the average VC incidence for firms that IPO between 1990 and 2015 as 37%. Kaplan and Lerner (2010) show this highly fluctuates through time.

TABLE A4

Relationship of VC Likelihood to Equity Growth Outcomes for all firms
Dependent Variable: Equity Growth (IPO or Acquisition)
50% Test Sample.

	All Firms that Achieve Growth	
	(1)	(2)
Log VC Likelihood ($\ln(\hat{\mu})$)	2.013***	
	(0.0205)	
VC Likelihood (Standardized)		1.116***
		(0.00605)
State F. E	Yes	Yes
Year F E	Yes	Yes
N	6601590	6601590
pseudo R-sq	0.131	0.067

Robust standard errors in parenthesis clustered at the state-cohort level. VC likelihood is the estimated likelihood of raising venture capital given the at-birth characteristics of a company, it is estimated in a separate training sample, showing in Table 4.

FIGURE A1
Time to VC Financing

Notes: This graph shows the time to financing by years (in pink) and months (in blue) for all firms that receive VC, estimated as the number of months between incorporation date and date of first VC investment.