PASSIVE VERSUS ACTIVE GROWTH:
EVIDENCE FROM FOUNDER CHOICES AND VENTURE CAPITAL INVESTMENT

Christian Catalini
Jorge Guzman
Scott Stern

Passive Versus Active Growth: Evidence from Founder Choices and Venture Capital Investment
Christian Catalini, Jorge Guzman, and Scott Stern
NBER Working Paper No. 26073
July 2019
JEL No. L25,L26

ABSTRACT

This paper develops a novel approach for assessing the role of passive learning versus a proactive growth orientation in the entrepreneurial growth process. We develop a simple model linking early-stage founder choices, venture capital investment and skewed growth outcomes such as the achievement of an IPO or significant acquisition. Using comprehensive business registration data from 34 US states from 1995-2004, we observe that firms that register in Delaware or obtain intellectual property such as a patent or trademark are far more likely to ultimately realize significant equity growth, and these choices also predict early-stage venture capital investment. Moreover, the estimated probability of receiving venture capital as reflected in early-stage founder choices predicts growth even for firms that do not receive venture capital. We use these findings to estimate bounds on the fraction of proactive versus passive firms among firms that ultimately achieve significant equity growth. While nearly half of all firms that achieve modest equity growth (> \$10M) are consistent with passive learning (as they neither make early-stage founder choices nor receive venture capital), 78% of firms experiencing an equity growth event greater than \$100M are associated with active founder choices and/or venture capital investment, and these firms are concentrated in geographic hubs such as Silicon Valley. Finally, our approach offers a novel approach for estimating the private returns to venture capital, matching on founder choices rather than demographics; consistent with prior studies, venture-backed firms are approximately 5X more likely to grow, with heterogeneity across location and time period.

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

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

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

I.   INTRODUCTION

The skewed nature of firm growth outcomes is a striking feature of the process through which entrepreneurship shapes broader economic performance. Only one in 1000 firms achieves a successful financial exit such as an IPO or high-value acquisition, and a very small fraction of firms is ultimately responsible for the vast majority of growth in revenues and employment (Decker, et al, 2014). This striking heterogeneity in firm outcomes reflects two distinct sources of heterogeneity among firms. First, firms differ markedly at their moment of founding in their potential and ambition to grow. Most founders intend to operate a small business, and only a small fraction of founders state an intention to operate as a growth-oriented venture at the time of founding (Hurst and Pugsley, 2011). Second, regardless of the founders' intentions at founding, entrepreneurship is associated with significant uncertainty that is only resolved over time through market experience, including the realized level of demand and productivity, and more informal feedback through consumers and suppliers. The resolution of uncertainty over time induces some small business-oriented entrepreneurs to learn from positive market feedback about their underlying growth potential (and the potential financial returns from business expansion), while many (indeed most) growth-oriented entrepreneurs are unable to achieve significant growth despite their founding motivation.

Consider Amazon and Chipotle, both founded in the mid-1990s, and each now employing well over 50,000 employees. On the one hand, when Jeff Bezos left the financial firm D.E. Shaw to start Amazon, his intention was to build a high growth startup by taking advantage of the nascent Internet to support electronic commerce (Stone, 2013). Not simply hindsight bias, the proactive growth orientation of Bezos is reflected in choices made in 1994 (the time of founding), including early registration of copyrights, patent filings, and organizing the firm as a Delaware corporation to facilitate external investment (Amazon in fact attracted venture capital financing from Kleiner Perkins in 1996). In contrast, Steve Ells, a recent cooking school graduate and sous chef at a local restaurant, received a small loan in 1993 (from his parents) to open a single Mexican grill restaurant, which he hoped would be successful enough to allow him to found a follow-on single-location fine-dining establishment. However, the unanticipated market success of the first Chipotle prompted Ells to focus his efforts on establishing a chain of "fast casual" Mexican restaurants, growing to more than 2000 locations by 2017.

This distinction between passive versus active growth is central to our theoretical understanding of the nature of business dynamics. Starting with Gibrat (1931), economists have grappled with the empirical relationship between the distribution of firm-level growth and the resulting (skewed) distribution of firm size (Iriji and Simon, 1977; Gabaix, 2009; see Sutton, 1997, for a synthetic review). On the one hand, skewness in the firm size distribution can result directly from initial heterogeneity among firms in their productive efficiency, and learning through market interactions about relative costs over time (Jovanovic, 1982). Under this conceptualization, the distribution of firm size and profits results not from active choice by firms or strategic investments, but from market selection pressure that leads inefficient firms to exit and more productive firms to attract larger market shares and a higher share of industry rents. Alternatively, if founders understand their relative potential at the time of founding (and have the ambition to grow), skewness can result from differences among firms in their strategic investment choices (Ericson and Pakes, 1995; Klette and Kotrum, 2004). As emphasized by Akcigit and Kerr (2018), firms with heterogeneous innovations choose to invest more or less in growth based on the underlying quality of their idea; as such, it is the combination of initial differences in potential and active choice by entrepreneurs (and investors) that results in the skewed distribution of growth outcomes.

Not simply a matter of economic theory, the relative importance of "active" versus "passive" entrepreneurs for overall business dynamism has implications for policy. If business dynamism results from the choices of proactive growth firms at the time of founding, then the economic case for targeted opt-in interventions (such as entrepreneurial accelerators) or favorable tax treatment (e.g., for venture capital investment) would be enhanced, since such focused policies might have a disproportionate impact on improving overall economic activity. On the other hand, if the bulk of realized growth arises from firms whose founders were unaware of the prospects for growth at founding, then policy might be oriented towards simply encouraging a higher level of entrepreneurial experimentation (e.g., through simplified and low-cost business registration and subsidized provision of small business loans).

Despite its importance for theory and policy, the empirical nature of firm growth has remained elusive. The principal barrier is a fundamental measurement challenge: among the millions of new firms founded each year, how would one "assign" firms to be classified as active versus passive at founding? Systematic administrative data capturing the population of businesses, such as the Longitudinal Business Database, tend to have limited information about the initial

growth orientation of new firms, and by and large infer growth through observed employment, revenue or productivity dynamics (Davis and Haltiwanger, 1992). While administrative data provide direct insight into the consequences of firm heterogeneity (and how dynamism varies across time, place and sector), ex post observations of growth cannot in isolation identify the relative importance of a proactive growth orientation on the part of founders at the time of founding. A separate stream of research focuses more directly on growth-oriented firms by abstracting away from the full population of establishments to focus on carefully constructed samples of firms, most notably those in receipt of venture capital (Lerner, 1994; Hellman and Puri, 2000; Chemmanur, Nandy, and Krishnan, 2011). Tracing the impact of such firms offers significant insight into the role of active growth firms on the economy. For example, although the venture capital industry only funds approximately one thousand ventures per year in the United States, firms linked to venture capital account for between 30% to 70% of start-up IPOs (Kaplan and Lerner, 2010; Ritter, 2016). While offering significant insight into the incidence and nature of growth among active firms, these careful studies of venture-financed companies throw into sharp relief how little we know about those firms that grow without receiving venture capital.

This paper breaks through this impasse by developing and estimating a model linking early-stage founder choices (across the full population of founders), venture capital investment, and growth outcomes in a way that allows us to estimate bounds on the relative importance of active versus passive firms in the growth process. Our approach takes advantage of three interrelated insights. First, building on Guzman and Stern (2015, 2016, 2017), if firm governance choices at the time of founding reflect founder beliefs about the growth potential of the firm, then these choices should themselves predict growth (though with significant noise). For example, if one sign of an "active" growth firm is investment in formal intellectual property protection such as a trademark or patent, then, across the full population of new firms, the probability of growth should be higher for firms that invest at founding in formal intellectual property protection relative to those that do not. While this growth prediction is of course not causal, the observation that choices at founding consistent with the ambition to grow also predict growth is consistent with the hypothesis that founder choices are informed by (true) growth potential. Second, relative to non-equity investors such as banks whose primary concern is the likelihood of default (i.e., the likelihood of failure), equity-oriented investors such as venture capitalists are explicitly selecting firms based on their potential to grow (and their own ability to enhance such potential). As such,

those founder choices that predict firm growth should, all else equal, predict the likelihood of attracting venture capital investment, and, the likelihood of receiving venture capital should be positively associated with the probability of growth even for firms that do *not* receive venture capital. Moreover, given that there are also some firms that do ultimately grow through passive learning (and so would have been unlikely to have made early-stage choices consistent with a growth orientation at birth), the empirical relationship between early founder choices and venture capital is likely to be tighter than the relationship between founder choices and overall growth. Finally, to the extent that we are able to establish a linkage between founding choices that predict growth and venture capital, then we can also develop a categorization of firms in terms of their growth orientation: simply put, a lower bound on the fraction of "active" growers is equal to the fraction of firms that ultimately grow that (a) make founding choices consistent with a growth orientation, (b) receive venture capital, or (c) both. In addition, this approach allows us to establish an estimate for an upper bound for the returns to venture capital, where we can match firms based not on demographic information such as industry or location, but on the observed choices of their founders at birth.

We put these ideas to work using data covering the full population of new US companies registered between 1995 and 2005 in 34 US states (covering 95% of US VC financing). Extending the data developed in the Startup Cartography Project (Guzman and Stern, 2016), we consider three core measures reflecting early-stage founder choices, including the choice to register as a Delaware corporation (a more expensive form of corporate governance offering advantages to growth-oriented firms), apply for or acquire a US patent or apply for a US trademark within the first year after founding. We then combine these founder choice data with information about significant firm equity growth, including the realization of an IPO or significant acquisition within 10 years of founding. In line with the theoretical model, a small number of firm choices consistent with active growth offer a sharp predictive analytic for equity growth: firms whose founders register in Delaware, and acquire both patent and trademark protection are 278 times more likely to achieve an equity growth event than firms that are associated with none of these choices. Moreover, this measure of entrepreneurial quality is also predictive of the receipt of venture capital: 65% of all venture-backed firms are in the top 5% of its distribution. Analogously, the same founder choices predict venture capital receipt directly (e.g., firms whose founders register initially in Delaware are more than 8 times more likely to receive venture capital), and, more

importantly, the likelihood of receiving venture capital (as reflected by founder choices) is predictive of growth even for those firms that do *not* receive venture capital. Furthermore, those firms that grow with venture capital are drawn from a distribution of firms with a more active growth orientation: the distribution of entrepreneurial quality for venture-backed growth firms first-order stochastically dominates the distribution of growth firms without venture capital.

We then use this information to assess the relative role of proactive versus passive learning in the overall growth process in the economy. On the one hand, when we simply examine firms experiencing modest equity growth (IPO or > $10M acquisition), we find that approximately half of all growth firms make choices and experience early-stage investment consistent with proactive growth. However, when we examine firms experiencing "extreme" equity growth (IPO or > $100M acquisition), we find that at least 78% are associated with a proactive growth orientation, and that these firms are particularly concentrated in entrepreneurial hubs such as Silicon Valley.

We use these findings to provide insight into the returns to venture capital, using an approach where we match on founder choices rather than demographics: consistent with earlier estimates such as Puri and Zarutskie (2012), venture-backed firms are approximately 5X more likely to grow. Interestingly, in our data, the returns to venture capital are lower in the upper tail of the *VC-likelihood* distribution (e.g., firms in the top 0.05% of the distribution are only 2.4X more likely to grow after receiving venture capital), within startup hubs, and when follow on capital is more likely to be scarce (returns are highest during the early stages of the Internet boom).

The paper proceeds as follows. Section II develops our theoretical model. Section III introduces the data on US business registrations and the venture capital industry. Sections IV through VI discuss the core empirical findings. Section VII concludes.

II. PICKING WINNERS: SELECTION INTO VENTURE CAPITAL AND THE DISTRIBUTION OF ENTREPRENEURIAL PERFORMANCE

Our analysis is motivated by the simple (but often implicit) insight that the key distinction between alternative models of firm growth reflects differences in the ability of founders to forecast their growth potential at (or near) the time of founding. Under the theory of *passive growth*, founders are relatively uninformed about the growth potential of their venture, and so any choice they make at founding is unlikely to be predictive of growth itself. Not simply an informational constraint of founders, early-stage investors are also uninformed about growth potential; if early-

stage investing is governed by (essentially) random assignment, then the returns to venture capital would be equal to the difference in outcomes between firms that rely on their funding, and those that do not. In other words, when founders and investors cannot forecast the potential for growth, founder choices should be uninformative about future growth, and inferring the returns to early-stage investment will be uncontaminated by selection bias.

In contrast, under a *proactive growth* model, entrepreneurs observe early signals of growth potential and make informed decisions based on them. These early-stage choices include the firm's corporate form (i.e., governance), investment (i.e., whether to seek out and accept risk capital investment), and whether to establish formal intellectual property protection over their ideas through patents and trademark. If founders receive informative signals about their growth potential, their choices would reflect those signals, and so will be predictive of growth itself. Investors, likewise, would be able to extract some information about the true, underlying quality of a venture and allocate risk capital based on those signals; as a result, the observed returns on investment would reflect the combination of selection (as only higher potential firms receive their capital) and the impact of capital investment on the growth process.[2]

To clarify the economic implications of these alternative perspectives on growth, we build a simple model linking early-stage governance and investor choices to long-term firm performance. Our analysis first focuses on the interplay between founding choices, the choices of venture capital investors, and equity growth outcomes. In particular, we use the model to decompose the prediction problem facing founders and venture capitalists (and the econometrician), and evaluate the potential to predict growth based on founder choices at or just after the time of founding.

We start by considering a newly formed value-maximizing and risk-neutral firm that observes a signal of its underlying quality $\theta \in \{\theta^L, \theta^H; \theta^H > \theta^L\}$, where the firm is of quality θ^L with probability λ, and θ^H with probability $1 - \lambda$. The model proceeds in three stages. First, the startup chooses a level of corporate governance, $K \in \{0,1\}$, where a higher level of corporate

---

[2] If ventures are heterogeneous not only in their underlying growth potential but also in their ability to observe this potential at founding, firm growth patterns may reflect the presence of both proactive investment and passive learning. To the extent that some ventures and investors are able to observe growth potential at an early stage, investors such as venture capitalists would likely invest primarily in firms where high potential quality is observable, while some portion of firms that grow without active investors would reflect the passive growth trajectory. In the extreme, if the venture capital industry were efficient in allocating capital to all firms with observable high-growth potential, then all firms that grow without venture capital would be drawn from the pool of passive learners.

governance enhances the growth potential of the firm, is costly to adopt, and is a complement to the underlying quality of a firm (i.e. better forms of governance are disproportionately helpful for higher quality firms). The next stage involves a choice by a venture capitalist to invest or not in the firm ($V \in \{0,1\}$), where we focus only on the extensive margin of whether the VC industry chooses to invest in the firm.[3] Similar to the corporate governance choice, we assume that venture capital investment is complementary to underlying quality and involves an idiosyncratic fixed cost (reflecting search or transaction costs). Last, conditional on initial quality, governance, and investor choices, the firm achieves an equity growth outcome through the realization of an independent stochastic component, ϵ, distributed log-logistic (i.e., $\ln(\epsilon) \sim Logistic; E(\epsilon) = 1$). Hence, startup performance is simply:

$$Y = \theta f(K, V)\epsilon \tag{1}$$

where $f(\cdot, \cdot)$ accounts for the impact of governance and venture capital financing choices on expected performance, and is (weakly) monotonically increasing in K and V. To the extent that most of the variance in performance among firms results from the realization of ϵ, equation (1) is fully consistent with the passive learning model (as in Jovanonic), while if a significant fraction of the variance arises from differences in θ, then equation (1) would be more consistent with a proactive model of firm growth (as in Ericson and Pakes).

Stage 1: The governance choice. The founders choose whether to adopt a higher level of corporate governance at the time they set up the firm. Their choice is based on their perception of the underlying quality of the firm, their expectation about whether they are likely to receive venture capital or not, the adoption cost of the governance form ($c_K > 0$), and the cost venture capitalists face when making investments ($c_V > 0$). Both costs are randomly allocated.[4] Entrepreneurs know c_K, and the distribution of c_V, though not its realized value. c_V influences the startup choices by shaping their assessment of their probability of raising venture capital. The startup chooses K by solving:

$$\max_{K=\{0,1\}} \theta f(K, E[V|K, \theta]) \, E[\epsilon] - Kc_K = \theta f(K, E[V|K, \theta]) - Kc_K \tag{2}$$

---

[3] We therefore abstract away from the level of investment and the distribution of equity, and assume that positive VC investment reflects a willingness to invest, a willingness by the founders to accept the terms of the investment, and terms that depend on the idiosyncratic bargaining positions of the VC and the founders.

[4] Specifically, the draws of c_K and c_V are idiosyncratic to each firm i, but we suppress the index i for expositional clarity.

To sharpen our focus on the role of proactive investment choices given a firm's self-perceived growth potential, we assume that firms that receive a signal that they have extremely low potential will not find it worthwhile to adopt K regardless of adoption costs or venture capital financing ($\theta^L f(1,1) < \min(c_k)$).[5] However, to account for heterogeneity in corporate governance choices among firms of high quality, we also assume that firms that receive a signal that they are of high potential will only adopt K when adoption costs are not too high ($\theta^H f(1,1) < \max(c_k)$). Hence, the fraction of firms choosing $K=1$ conditional on $\theta = \theta^H$ is ρ_K:

$$\rho_K = \int_{\min(c_k)}^{c^*} dF(c_K), c^* = \theta^H f(1, E[V|1, \theta^H]) \tag{3}$$

Stage 2: The venture capitalist's choice. The (profit-maximizing) venture capitalist observes the same signal of θ as the startup. It also observes the startup's choice of K, and the transaction cost of investment $c_V > 0$. Analogous to the startup governance choice, the venture capitalist chooses whether to invest a fixed amount (normalized here to 1) in the company by solving:

$$\max_{V=\{1,0\}} y = \theta f(K,V) - (1 + c_V)V \tag{4}$$

We assume that it is never worthwhile to invest in a firm unless the potential of the firm is high ($\theta^L f(1,1) - 1 < \min(c_V)$), but that even when the signal is high it is not worthwhile to invest if transaction costs are high ($\theta^H f(\cdot,1) - 1 < \max(c_V)$). As mentioned earlier, we abstract away from the precise level, staging or terms of the investment, and assume that when the cost of transacting is low and observable quality is high, the VC and the firm can reach an agreement to maximize value. The fraction of firms receiving venture capital is therefore ρ_V:

$$\rho_V = \rho_K \int_{\min(c_V)}^{c_V^{K=1}} dF(c_V) + \left(1 - \rho_K \int_{\underline{c_V}}^{c_V^{K=0}} dF(c_V)\right), c_V^{K=1} = \theta^H f(1,1) - 1 \tag{5}$$

$$, c_V^{K=0} = \theta^H f(0,1) - 1$$

Stage 3: Firm Performance. Conditional on firm quality, corporate governance, and VC investment, the startup's financial performance is realized according to (1). Given the skewed nature of firm growth outcomes, we focus on whether firms achieve an "extreme" growth outcome (in our empirical work, this will be whether the firm achieves an IPO or is acquired above a

---

[5] This type of behavior would be consistent with the empirical findings of Hurst and Pugsley (2011), in which the most founders neither anticipate nor invest in growth.

threshold value). Specifically, we will consider the observation of performance in a latent variable framework, where g is equal to 1 if y is greater than some threshold y^* and 0 otherwise.

We develop four propositions from this setup. First, we show that entrepreneurial quality can be partially estimated from founding choices, and that the resulting estimate also predicts the ability to raise venture capital. Second, we show that the likelihood of receiving VC funding can also be partially estimated from the same founding choices, and that the resulting VC likelihood in turn predicts follow-on performance even among firms that never take VC. Third, the quality distributions of firms that grow with and without venture capital are different, with the quality of growth with VC being higher (first-order stochastic dominance) than growth without VC.

We then leverage these insights to show how the observability of founding choices at birth and the availability of data on realized venture capital investments allow us to conceptually differentiate between passive and proactive firms at time of founding. To do so, we extend the model by: 1) allowing θ to be continuous rather than binary (thus enabling firms to differ significantly between themselves, as in Guzman and Stern, 2015); and 2) by considering n founding choices $\boldsymbol{K} = \{K_1, \dots, K_n\}$ rather than a single one. Using this framework, we propose a method to empirically estimate the share of proactive growers in the economy, and highlight how downward bias makes this measure a lower bound of the true share. Finally, for our fourth proposition, we show that this approach also allows us to use the founding choices to account for selection (on observables) into venture capital, and estimate an *upper* bound on the returns to this source of funding.

From Entrepreneurial Quality to Selection into Venture Capital. We build on Guzman and Stern (2015, 2016, 2017) and explore how founding choices can reveal valuable information about a firm's underlying quality even when quality itself is unobservable.

Lemma 1. Startup Choices Inform Entrepreneurial Quality. Define $I_X(\theta)$ *as the information of some variable on θ. Then, there exists a function $h(\boldsymbol{K}) \rightarrow \hat{\theta}$ such that*

$$I_{\hat{\theta}}(\theta) > I_{\bar{\theta}}(\theta), \hat{\theta} = h(\boldsymbol{K}), \bar{\theta} = \bar{g} \tag{8}$$

Proof. In appendix.

Lemma 1 indicates that there is information about a firm's underlying quality that can be estimated from founder choices. The intuition for this comes from the startup maximization function. As discussed in equation (2), K is only adopted by startups of high quality. Therefore $K = 1$ is informative about a startup's underlying quality (i.e. we learn that $\theta = \theta^H$ even though

quality itself is not observed). Lemma 1 directly links the choices startups make at founding with their underlying unobservable quality, and, gives us an empirical model for estimating $\hat{\theta}$. Since VCs also care about entrepreneurial quality, this empirical measure should also predict the receipt of venture capital.

Proposition 1. Estimated Quality Predicts Venture Capital.

$$\frac{\partial P(V = 1|\hat{\theta})}{\partial \hat{\theta}} > 0 \tag{9}$$

Proof. In appendix.

Proposition 1 creates an empirical linkage of a theoretical relationship in the model: firms of higher quality are more likely to raise VC. The logic is simple. VCs invest in high quality firms to maximize the returns of their portfolio. The value of Proposition 1 is that it moves from simple quality, to an empirical measure estimated from founding choices (that is, from θ to $\hat{\theta}$). The key assumption needed for this link is that $\hat{\theta}$ itself is informative (which we show in Lemma 1).

'VC Likelihood' and its Role in Predicting Performance Even in the Absence of VC. We now take an alternative perspective and consider the role of founding choices in directly predicting VC financing.

Lemma 2. Startup Choices Inform Venture Capital. Define $\mu = P(V = 1)$, and $I_X(\mu)$ as the information of some variable on μ. Then, there exists a function $h'(\boldsymbol{K}) \to \hat{\mu}$ such that

$$I_{\hat{\mu}}(\mu) > I_{\bar{\mu}}(\mu), \hat{\mu} = h'(\boldsymbol{K}), \bar{\mu} = E[V] \tag{10}$$

Proof. In appendix.

Lemma 2 constructs a parameter, μ, which we refer to as the '*VC Likelihood*'. It is the underlying probability that a firm will receive VC financing given its founding fundamentals (quality and governance). This summary statistic is positive for all firms, but could vary widely across them. Using a similar logic to the one used in Lemma 1, Lemma 2 shows that an empirical version of the VC likelihood, $\hat{\mu}$, can be estimated from founding choices $\boldsymbol{K}$. Better governance both indicates a higher underlying quality, and is useful in and of itself. We use Lemma 2 to consider whether the VC likelihood predicts growth among startups that do not receive VC. One of the motivating stylized facts for our paper is that most IPOs and acquisitions occur without VC (Ritter, 2017), and a key question is whether these are simply passive or proactive growers. Firms that attract VC capital — i.e. that have high μ — due so because of their underlying fundamentals. At the same time, in the VC choice equation (eq. (4)), venture capitalists can get draws of

investment costs that are too high for them to invest even if the startups are intrinsically of high quality. When this happens, these high potential firms will not receive VC, but will still be more likely to perform well because of their high quality (although they may face a penalty because of the lack of VC funding, mentorship and support).

Proposition 2. VC Likelihood Predicts Growth in Firms Without VC.

$$\frac{\partial P(g = 1|\hat{\mu}, V = 0)}{\partial \hat{\mu}} > 0 \tag{11}$$

Proof. In appendix.

Proposition 2 states that when a company is attractive to VCs and therefore has high VC-likelihood, it will be high performing even if it does not attract VC funding. This links the process of selection into venture capital to a more fundamental model of proactive growth, independent of funding source.

The Quality Distribution of Firms that Grow With versus Without VC. It is important to note that, among the growers in the non-VC sample, some are drawn from the group of passive growers (i.e., those who received the low-quality signal). For these companies, growth is completely random, and therefore would not be reflected through variation in μ. This bounds the informativeness of μ in terms of the potential for growth without VC:

Proposition 3 Among firms that grow, the initial distribution of quality of firms with VC is higher (in the sense of first order stochastic domination) than the distribution of quality of firms that do not receive VC.

$$P(\theta > x|g = 1, V = 1) > P(\theta > x|g = 1, V = 0) \ \forall \ x \tag{12}$$

Proof. In appendix.

Whether quality is higher or lower for VC growers than non-VC growers depends on the balance of two opposing effects. On the one hand, there is positive selection into VC. Since higher quality firms are more likely to get VC, it follows that within growth firms the probability that a firm has received VC increases as we move up in the quality distribution. However, there is also a positive treatment of VC. Because VCs directly improve the performance of venture-backed firms, it is possible that VCs move lower quality firms above a 'growth threshold'. Positive VC treatment could lead us to observe a lower founding quality for firms that grow with VC. However, under our assumption that VC financing is complementary to firm quality, the value-added of VC would be higher as quality increases. As a result, VC funding would have a relatively small effect on

moving low quality ventures above the growth threshold, and would have a larger impact on helping already high quality firms increase their chances of growth. Due to this complementarity, the overall distribution of quality of VC-backed growth firms should be more selective in initial quality than the distribution of quality of firms that do not receive VC.

Estimating the Share of Passive and Proactive Growth in the Economy. Building on this structure, we can distinguish the growth of startups across several distinct groups depending on their quality, their choices of governance at founding, and their use of venture capital. Normalizing θ^L to 1 for simplicity, and defining s^L as the share of low-quality startups, and s^H_{KV} as the share of high-quality startups with specific governance choices K and financing V, we can define the share of startups that grow under each group as a function of the logistic conditional distribution function F, as:

$$\alpha^L = \frac{s^L}{\bar{g}} P(g = 1|\theta^L) = \frac{1}{\bar{g}} \lambda \times \left(1 - F(y^*)\right)$$

$$\alpha^H_{00} = \frac{s^H_{00}}{\bar{g}} P(g = 1|\theta^H, K = 0, V = 0) = \frac{1}{\bar{g}}[(1-\lambda)(1-\rho_K)(1-\rho_V)] \times \left(1 - F\left(\frac{y^*}{\theta^H}\right)\right)$$

$$\alpha^H_{10} = \frac{s^H_{10}}{\bar{g}} P(g = 1|\theta^H, K = 1, V = 0) = \frac{1}{\bar{g}}[(1-\lambda)\rho_K(1-\rho_V)] \times \left(1 - F\left(\frac{y^*}{\theta^H f(1,0)}\right)\right) \quad (13)$$

$$\alpha^H_{01} = \frac{s^H_{01}}{\bar{g}} P(g = 1|\theta^H, K = 0, V = 1) = \frac{1}{\bar{g}}[(1-\lambda)(1-\rho_K)\rho_V] \times \left(1 - F\left(\frac{y^*}{\theta^H f(0,1)}\right)\right)$$

$$\alpha^H_{11} = \frac{s^H_{11}}{\bar{g}} P(g = 1|\theta^H, K = 1, V = 1) = \frac{1}{\bar{g}}[(1-\lambda)\rho_K\rho_V] \times \left(1 - F\left(\frac{y^*}{\theta^H f(1,1)}\right)\right)$$

It is useful to make some observations on how these five groups capture different parts of the process of firm growth. The model suggests that firms who receive a signal that they are of the low type (or equivalently do not receive a signal that they are of the high type), will not engage in costly investments in corporate governance, nor will they attract venture capital. However, despite the lack of a positive signal or any proactive growth choice on their part, a small fraction of these firms will still grow because of random luck through ϵ (the firms in α^L). Among the firms that receive positive signals, only a fraction will invest in corporate governance or receive venture capital ($\alpha^H_{00}, \alpha^H_{01}, \alpha^H_{10}$). Overall, the potential for growth will be highest among firms that are not only of high intrinsic quality, but also receive favorable cost draws for both K and V (i.e., α^H_{11}).

We use these α's to structure our understanding of the observability of the passive and proactive growth models of startup performance. Passive growth is the growth that is reflected in the share α^L, while proactive growth in α^H. The incidence of proactive growth is exactly

$$\zeta = \alpha_{00}^H + \alpha_{10}^H + \alpha_{01}^H + \alpha_{11}^H \tag{14}$$

This value is not estimable directly by the econometrician, because the econometrician cannot empirically distinguish between α_{00}^H (high quality startups with no founding choices) and α^L (low quality startups). What can be estimated instead is a lower bound of the incidence of passive growers by assuming that α_{00}^H is zero:

$$\text{Lower Bound: } \underline{\hat{\zeta}} = \alpha_{10}^H + \alpha_{01}^H + \alpha_{11}^H$$

However, this can be improved by considering how high quality firms take or avoid governance among VC-funded firms and non-VC-funded firms. Specifically, we can define a coefficient η as the ratio of the governance-non governance odds for VC and non-VC-funded firms as:

$$\eta = \frac{\left(\frac{\alpha_{00}^H}{\alpha_{10}^H}\right)}{\left(\frac{\alpha_{01}^H}{\alpha_{11}^H}\right)} = \left(\frac{\left(1-F\left(\frac{y^*}{\theta^H}\right)\right)}{\left(1-F\left(\frac{y^*}{\theta^H f(1,0)}\right)\right)}\right) \Bigg/ \left(\frac{\left(1-F\left(\frac{y^*}{\theta^H f(0,1)}\right)\right)}{\left(1-F\left(\frac{y^*}{\theta^H f(1,1)}\right)\right)}\right) \tag{15}$$

Where $\frac{\alpha_{00}^H}{\alpha_{10}^H}$ is the odds that a non VC growth firm takes governance, and $\frac{\alpha_{01}^H}{\alpha_{11}^H}$ is the odds that a VC firm takes it. Then, depending on the value of η, we can estimate α_{00}^H. For example, if $\eta = 1$ (i.e., the odds of taking governance or not are the same with and without VC), then $\hat{\alpha}_{00}^H = \alpha_{10}^H \alpha_{11}^H / \alpha_{01}^H$. However, if η is less than 1, then this value of $\hat{\alpha}_{00}^H$ is only an upper bound of the estimate.

Proposition 4 (Upper Bound Estimates for ζ)

$$\eta < 1 \implies \alpha_{00}^H < \alpha_{10}^H \alpha_{11}^H / \alpha_{01}^H$$

Proof: In Appendix

The insight in Proposition 4 is that because the ratio η is less than 1, then we are able to recover an upper value for the share of companies without governance or VC by looking at the relative incidence of the other three groups. Intuitively, while the benefit of governance is independent by assumption, the jump in probability space of the likelihood of achieving growth from α_{00}^H to α_{01}^H is higher than the jump from α_{10}^H to α_{11}^H due to decreasing marginal returns in the coverage of the probability. Therefore, we can estimate an upper bound as:

$$\text{Upper Bound: } \bar{\hat{\zeta}} = \left(\frac{\alpha_{10}^H \alpha_{11}^H}{\alpha_{01}^H}\right) + \alpha_{10}^H + \alpha_{01}^H + \alpha_{11}^H$$

Estimating the Returns to VC Financing. Last, we build on this setup to interpret estimates of the returns to VC financing using the VC likelihood, μ, to control for selection on observables. We first note that since the true VC likelihood is a sufficient statistic for the observable part of selection into VC, we can interpret estimates of the returns to VC as a propensity score model, equivalent to Rosenbaum and Rubin (1983) and subsequent literature. In this setup, the returns to VC, τ, are identified as the difference between two potential outcomes, conditional on this selectivity.

$$\tau = E[Y(1) - Y(0)|\mu]$$

Of course, we do not have a measure of μ, but instead only an empirical estimate, $\hat{\mu}$. We follow Imbens and Rubin (2016) by doing matching on $\hat{\mu}$ to improve balance in our sample. Yet, because the ability to control for selection is imperfect (we cannot control for unobservable quality or growth potential), $\hat{\mu}$ does not perfectly capture μ, leading to positive bias.

Proposition 5 (Upper Bound Estimates of the Returns to VC)

$$\hat{E}[\hat{\tau}|\hat{\mu}] \geq E[\tau|\mu]$$

Proof. In appendix.

That is, any estimate of the returns to venture capital controlling for founding choices is simply an upper bound of the true effect.

Putting all of this together, we can make a number of predictions that summarize the key propositions of our model. First, estimated quality from founding choices predicts venture capital financing. Second, the VC likelihood of a firm predicts performance also among non-venture-backed firms. Third, the quality distribution of venture-backed firms that grow is higher than the quality of firms that grow without VC (first order stochastic dominance). Fourth, our approach allows us to estimate an upper bound and a lower bound for the incidence of active growth startups in the economy. And fifth, the method can be used to control for selection into venture capital through founding choices and estimate an upper bound to the returns to VC.

III. Empirical Framework[6]

We build on our earlier work (Guzman and Stern, 2015, 2016, 2017) to propose an empirical approach that uses founding choices, growth outcomes, and VC financing events to create measures of entrepreneurial quality and VC likelihood. While our prior work only focused on estimating quality, in this paper we extend this to consider venture capital financing and the interaction between the two.

Our approach combines three interrelated insights. First, as the challenges to reach a growth outcome or receive venture capital financing as a sole proprietorship are extremely high, a practical requirement for any entrepreneur to achieve growth is business registration (as a corporation, partnership, or limited liability company). This practical requirement allows us to form a population sample of entrepreneurs "at risk" of growth at a similar (and foundational) stage of the entrepreneurial process. Second, we are able to observe additional founding choices *at or close to the time of registration.* For example, we can measure whether the firm is organized in order to facilitate equity financing (e.g., registering as a corporation or in Delaware), or whether the firm seeks intellectual property protection (e.g., a patent or trademark). Third, though rare, we observe venture capital financing events and growth outcomes (such as achieving an IPO or high-value acquisition). Combining these insights, we estimate entrepreneurial quality by estimating the relationship between observed growth outcomes and startup characteristics, and estimate VC likelihood as the relationship between venture capital and the same startup characteristics. For a firm i born at time t, with at-birth startup characteristics $K_{i,t}$, we observe growth outcome $g_{i,t+s}$ s years after founding and estimate:

$$\theta_{i,t} = P(g_{i,t+s}|K_{i,t}) = f(\alpha_\theta + \beta_\theta K_{i,t}) \tag{16}$$

And use the *predicted* value, $\hat{\theta}$, of this regression as our measure of entrepreneurial quality.

Similarly, for a firm i born at time t, with at-birth startup characteristics $K_{i,t}$, we observe venture capital event $V_{i,t+s'}$, s' years after founding and estimate:

$$\mu_{i,t} = P(V_{i,t+s'}|K_{i,t}) = f(\alpha_\mu + \beta_\mu K_{i,t}) \tag{17}$$

And use the *predicted* value of this regression as our measure of VC likelihood.

As long as the process by which startup characteristics map to growth outcomes remains stable over time (an assumption which is itself testable), this mapping allows us to calculate an

---

[6] As appropriate, some portions of Section III and Section IV borrow (often verbatim) from the descriptions of the same approach and the data in Guzman and Stern (2016).

estimate of entrepreneurial quality at founding and VC likelihood for any business registrant within our sample (even for firms that never seek venture-capital financing). In our analysis, we will use empirical measures of quality and VC likelihood ($\hat{\theta}$ and $\hat{\mu}$) and their relationships to g and V to test our underlying model of firm growth.

IV. DATA

Our analysis uses business registration records, venture capital financing events, and equity growth outcomes. By linking these three datasets together, we are able to create profiles of new firms and their early choices at the time of founding, and connect these to the subsequent receipt of venture capital (or not), and to equity growth outcomes. We perform this matching for all new companies registered between 1995 and 2005 in 34 U.S. states,[7] accounting for 95% of all venture capital financing. The resulting dataset, containing over 10 million firms, allows us to characterize the growth process of startups from founding to the receipt of VC, and equity growth (IPO or valuable acquisition). We now describe each of the components of the dataset and how it relates to our theoretical framework.

Business Registration Records.[8] Business registration records are public records created when an individual registers a new business as a corporation, LLC or partnership (Guzman and Stern, 2015; 2016; 2017). We use all registrations from 1995 to 2005 in 34 US states, representing 95% of the United States venture capital market in 2014 (SSTI, 2015, see Figure A1). 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 (or Commonwealth) 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

---

[7] Arizona, California, Colorado, Florida, Iowa, Idaho, Illinois, Kentucky, Massachusetts, Maine, Michigan, Minnesota, North Carolina, North Dakota, New Jersey, New Mexico, New York, Ohio, Oklahoma, Oregon, Rhode Island, South Carolina, Tennessee, Texas, Utah, Virginia, Washington.
[8] This section draws heavily from our prior work, where we introduce business registration records and some of the measures used in this paper.

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 10,451,896 observations. The founding documents contain the name of the firm, its president, CEO, or manager,[9] its address, date of registration, and jurisdiction, amongst others. For each observation we construct measures of founding observables based on choices that are observable in the registration record itself, or in USPTO filings that are contemporaneous to the registration of the company.

A key founding choice, the choice of jurisdiction — specifically whether to register the firm in Delaware or under the local jurisdiction — exhibits the characteristics of the quality-related choices described in our model: while its value increases with the quality of the firm, the costs only partially increase with firm quality. In the process of choosing a jurisdiction, founders benefit from registering in Delaware for several reasons, including: (1) that the Delaware General Corporate Law provides a long cannon of decisions that are useful in assessing the predictability of complex contracts; (2) that the state of Delaware has an advanced institutional setup to deal with corporate arbitration including its highly reputed Court of Chancery; and (3) that the decisions and legal framework of Delaware is generally regarded as pro-business. These benefits are more useful for startups that have growth ambitions, and for those that interact with investors, including VCs.[10] Registering in Delaware also implies the need to have one more registration in addition to the state of operation, which introduces fees that a business that expects to stay small is likely to deem unnecessary.[11] As such, we expect Delaware registration to be a useful variable for identifying proactive growers, and a good predictor for both the receipt of venture capital and equity growth.

We also document several other firm observables at founding which do not necessarily exhibit the characteristics discussed in the model. *Corporation*, is an indicator equal to 1 if a startup is registered as a corporation, rather than an LLC or partnership, and 0 otherwise. A corporation is the only legal form that allows the sale of shares to venture capitalists for investment. Our next

---

[9] We have several states in which the president name is not present. More generally, our models will include state fixed-effects to account for this and other systematic differences in the process of registration across states.
[10] In fact, venture capitalists often require that portfolio companies are in Delaware because their contracts are specifically written for Delaware corporate law.
[11] We estimate these at about $4,000 dollars.

two measures relate to how the founders choose to name the company. *Eponymous* is an indicator equal to 1 if the firm is registered after the name of the founder (e.g. 'Mike's Pastries'). Belenzon et al (*forthcoming*) build on Guzman and Stern (2015) to show that while eponymy is predictive of higher productivity (Belenzon et al, 2017), it is inversely predictive of growth due to financing channels. The second indicator is based on our qualitative observation that successful startups in the studied time period usually have short distinctive names (such as 'Spotify' or 'Google'), rather than longer, industry specific names. *Short Name* is a measure equal to 1 if the startup name is 3 words or less, and zero otherwise.

We also create several measures based on how the firm name reflects the industry or sector within which the firm is operating, taking advantage of the industry categorization of the US Cluster Mapping Project ("US CMP") (Delgado, Porter, and Stern, 2016) and a text analysis approach. We develop 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.

Intellectual Property Measures. We construct two measures related to quality based on whether the firm has a patent or a trademark as reported by the U.S. Patent and Trademark Office. The value of intellectual property is increasing in the market potential of the underlying ideas (i.e. the quality of the firm), but the costs of filing and maintaining this intellectual property do not increase at the same rate (at least as a first approximation). *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). Hsu and Ziedonis (2013) highlight the role of patents as 'dual advantages' in the performance of companies. They provide an intellectual property protection benefit to startups, but also facilitate transactions with venture capitalists (also Hellman and Puri (2000) and Gans et al (2008)). As such, we expect patents to be predictive of both performance and the receipt of venture capital. Our second measure, *Trademark*, is equal to 1 if a firm applies for trademark protection within a year from registration. Firms with an active

trademark could use the sales of products and services to bootstrap their growth without relying on external funding.

Venture Capital Financing. We collect information on Series-A venture capital financing for all startups in our sample from AngelList, CapitalIQ, Preqin, and Thompson Reuters VentureXpert. Our main variable, *Venture Capital*, is an indicator variable equal to 1 if the company receives venture capital in any of these datasets, and zero otherwise.

A central challenge in measuring venture capital events is the fact that existing VC databases are incomplete (Kaplan and Lerner, 2017). We try to limit the impact of this issue in three ways. First, we highlight that our VC variable is dichotomous and represents only the existence of VC financing. While there is often concern that VC databases do not adequately measure some of the details of individual VC deals, the existence of any financing (or its absence) is much more reliable. Second, we use multiple databases to both get a comprehensive set of VC events and compare the observability of these events across databases. We find almost all of the VC-backed firms we identify in VentureXpert are also in our other three databases and vice-versa. Finally, most of the data issues with VentureXpert are related to more recent years, and the data from our period is typically considered to be of much higher quality (Kaplan and Lerner, 2017). However, we may miss some venture capital events either because of our matching procedure or because of it not being observed in the data we use. The analysis can therefore include some Type-I errors on the presence of VC funding, but we believe it is unlikely to contain Type-II errors (firms that are classified as VC-funded, but that do not receive any). Our econometric approach bounding the returns to VC is consistent with this type of measurement error.

Growth Outcomes. Last, we consider growth outcomes that capture the same outcomes that venture capitalists care about when making their investment decisions. *Growth* is a binary measure equal to 1 if the firm has an initial public offering (IPO) or is acquired at a meaningful positive valuation within 10 years of registration as reported in the Thomson Reuters SDC database.[12] Between 1995 and 2005, we identify 13,292 firms that achieve growth, representing 0.13% of the

---

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

total sample of firms. We also experiment with different thresholds for defining what a successful acquisition represents.

Descriptive Statistics. Table 1 reports summary statistics. There are 10,451,896 firms on 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)[13]. 0.2% of firms have a patent (within 1 year from birth), and 0.09% have a trademark. 59% of firms are corporations, 47% have a short name, 7.9% are eponymous and 3.6% 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 0.13% of non-VC-funded firms achieve growth, this share is 27.6% for VC-funded firms. If we only focus on IPOs, whereas 1 out of 28 VC-funded firms achieves this milestone, among the remaining firms it is only 1 every 10,000 firms. Similarly, while approximately 1 out of 4 VC-funded firms is successfully acquired, only 1 out of 769 non-VC-funded firms does so.

V. EMPIRICAL RESULTS

ENTREPRENEURIAL QUALITY AND THE RECEIPT OF VENTURE CAPITAL. We begin our empirical analysis by studying Proposition 1, whether entrepreneurial quality — as estimated from choices at birth — predicts venture capital. To do so, we first estimate quality by implementing our predictive analytics approach discussed in Section III. We train a predictive logit model that maps the three measures of quality-oriented founding choices (*Delaware, Patent,* and *Trademark*) to equity growth events using a 50% random sub-sample (a training sample). The predicted value of this model is our empirical estimate of *Entrepreneurial Quality—$\hat{\theta}$* in the theoretical framework. We use this estimate to evaluate the role of underlying quality (as reflected in founding choices) in the receipt of venture capital in the remaining 50% of the data (the test sample).

---

[13] 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.

Table 3 is our predictive logit model.[14] (3-1) considers each effect entering the regression independently. (3-2), in contrast, includes interactions of Delaware and Patent. Our preferred specification is (3-3), which considers each possible combination of these parameters independently, effectively creating a non-parametric estimator. The relationship between our observables and equity growth is sizable, and increases when we introduce interactions suggesting complementarity between these choices. A startup registered in Delaware, for example, is 9.7 times more likely to grow than a startup registered in a local jurisdiction. However, this effect increases to 176 times when considering startups that both register in Delaware and have a patent, and to 119 for those that register in Delaware and have a trademark around founding. The largest effect is for startups that exhibit all of the three choices: a Delaware startup, with a patent and a trademark, is 279 times more likely to grow than a startup without any of these characteristics. Together, these results suggest a highly predictive value of founding governance choices and intellectual property on follow-on performance. As is clear in the theoretical framework, these relationships are not causal, but instead reflect choices made by the entrepreneurs at founding based on their own ambitions and assessment of the startup's potential. Specification (3-4) is a robustness test where we include additional founding choices. Though there are some small changes in the coefficients, the magnitudes remain similar.

The predicted value of (3-3) is our measure of *Entrepreneurial Quality*. It has a mean of 0.0013 and standard deviation of 0.0064. It is also highly skewed, with a skewness coefficient of 38. To consider its relationship to venture capital financing, we begin by performing an out of sample ten-fold cross validation procedure. In Figure 1, we plot the average out of sample share of all VC-backed firms located in each of twenty 5-percent bins in the distribution of quality. This graph provides a useful visual representation of the incidence of venture capital across our estimated entrepreneurial quality distribution, and constitutes a descriptive analysis of Proposition 1. The graph shows that VC funding is mostly concentrated among startups of extremely high entrepreneurial quality: 65% of all venture-backed firms are in the top 5% of the quality distribution, and another 13% in the next 5%.

To sharpen these results, we perform a regression analysis in Table 4. We estimate linear probability and logit models on the 50% test sample using *Venture Capital* as our dependent

---

[14] It is equivalent to the one used in Guzman and Stern (2017) to estimate entrepreneurial quality, which also use the same sample of US firms.

variable, and estimated entrepreneurial quality as the independent variable. In (4-1), we use the estimated quality without any controls. The coefficient is positive and significant, with a value of 0.82. This suggests that a 10 percentage-point increase in predicted quality at founding is associated with a 8 percentage-point increase in the probability of growth. To mirror the structure of our model (where each of the terms interacts multiplicatively, with log-logistic error), we estimate a log-linear relationship in (4-2). The mean of this variable is -8.9 and the standard deviation 1.35. We also include state and year fixed-effects to control for the most obvious sources of heterogeneity. The main coefficient is 0.0067 and significant. An increase of one standard-deviation in log quality is associated with a 0.9 percentage-point increase in the probability of raising venture capital. With a mean outcome of 0.08%, this difference is substantial. Given two firms that have a difference in estimated quality at founding of 1 standard deviation, the one of higher quality has an 11X higher probability of receiving venture capital. (4-3) performs the same analysis but in a logit model. The reported coefficient is an IRR, with a value of 3.3. In this model, an increase of 1 standard deviation in log quality is associated with a 4.3X increase in the probability of growth.

The skewness and high selectivity of VCs is best observed in (4-4). In this table, we include indicators for being at different ranges of predicted quality, and report the IRR of each one in a logit regression using *Venture Capital* as the outcome. The base category is being in the bottom 50% of the data by estimated quality. The differences increase quickly and are very pronounced at the upper levels of the quality distribution. Firms in the 50th to 90th percentile are 4.2X more likely to receive VC than those in the base category. Firms in the 90th to 95th percentile are 19X more likely to raise VC, and firms in the 95th to 99th percentile are 63X more likely to raise VC. At the very top, the effects are particularly salient. Being between the 99th and the 99.5th percentile is associated with a 165X increase in the likelihood of raising VC (relative to the baseline), and being in the top 0.5% with a 645X difference. We interpret these results as supporting Proposition 1: there exists a positive relationship between a firm's entrepreneurial quality (as measured by its choices at birth), and the likelihood of raising venture capital financing.

VC LIKELIHOOD AND GROWTH WITHOUT VENTURE CAPITAL. We continue our analysis by evaluating Proposition 2 — whether a firm's VC likelihood can also predict growth among firms that do not raise venture capital. This relationship can validate our assumptions on how founding

choices reflect fundamental drivers of growth, and can provide insight as to whether growth without VC is driven by a passive learning model (e.g. Jovanovic, 1981), or one of proactive investment and ex-ante differences in founding quality (Pakes and Ericcson, 1995).

To do so, in Table 5, we estimate a predictive model that relates *Venture Capital* to observables at founding. This follows the same logic of Table 3, but with a different dependent variable. We again divide our sample into a 50% training and a 50% test sample, and present our coefficients as incidence-rate ratios for ease of interpretation. (5-1) includes each of the three founding choices we observe for firms independently. (5-2) also interacts *Delaware* and *Patent*. Finally, (5-3) incorporates all the three-way interactions, effectively creating a non-parametric model. The relationship between our founding observables and the receipt of venture capital is particularly salient. As in Table 3, we find the odds of VC increase in the interactions, suggesting these choices are complements. Looking at specification (5-3), we see companies that only register in Delaware are 17X more likely to achieve a growth outcome, and firms that only have a patent are associated with an 87X increase in the odds of growth. However, firms with both Delaware and a patent (but no trademark) are 342X more likely to grow. Specification (5-4) is once again a robustness test where we include more controls. Though the coefficients change, the magnitudes remain meaningful, suggesting that our observables at founding are good predictors of venture capital financing.

The predicted value from (5-3) is our empirical estimate of the likelihood of raising venture capital for each firm (μ in the model), i.e. the *VC Likelihood* is estimated as the predicted probability from this regression. It is important to highlight that we estimate this measure for all firms in our data, and not only for venture-backed firms (all firms have a non-zero probability of raising venture-capital).

To test Proposition 2, we check if the *VC Likelihood* estimate predicts growth also among the firms that do not raise venture capital. In Figure 2, we look at this relationship by repeating the out of sample cross validation procedure used in Figure 1. The estimates inform where in the distribution of *VC Likelihood* non-VC-funded growth firms are likely to be found. The relationship between *VC Likelihood* and the growth outcomes of non-VC-funded firms is stark, and consistent with the presence of proactive growers also among these types of firms: 48% of non-VC-funded growth firms are in the top 5% of VC likelihood, and 11% are in the next 5%. In Table 6, we study this relationship in a regression using the same format of Table 4. We estimate linear probability

and logit models on the 50% test sample, excluding all venture-backed firms, using *Equity Growth* as our dependent variable and the firm's VC likelihood as an independent variable. In (6-1), we use the estimated value of VC likelihood without any controls. The coefficient is positive and significant, with a value of 0.69. The mean for the outcome variable is 0.001, implying a large difference in growth probability. As can be seen in (6-2), the effect of a one log-point increase in the log value of VC likelihood is associated with a 0.004 percentage points increase in the probability of equity growth. Given the mean value, this implies a 4X effect in the odds of growth. Model (6-3) is a logit model. The coefficient is an odds ratio of 2.3. That is, an increase of one log point of VC likelihood is associated with a 2.3X increase in the odds of growth. The skewness and difference of being at the upper tail of the VC likelihood is most visible in (6-4). In this column, we include individual indicators for being within different ranges of the VC likelihood distribution, and report the IRR of each one in a logit regression. The base category is being in the bottom 50% of the data by estimated VC likelihood. The parallels to Table 4 are significant, though all coefficients are somewhat smaller in magnitude. Differences are quite marked at the upper levels of the VC likelihood distribution. Firms in the 50th to 90th percentile are 1.7 times more likely to grow, firms in the 90th to 95th percentile are 9 times more likely, and firms in the 95th to 99th percentile are 19.4 times more likely to grow. Being between the 99th and the 99.5^{th} percentile is associated with a 45X increase (relative to the baseline), and being in the top 0.5% with a 162X difference. We interpret this evidence as supporting proposition 2, i.e. the estimated VC likelihood at founding predicts performance, even among the non-venture-backed sample.

THE QUALITY DISTRIBUTION OF VC AND GROWTH WITHOUT VC. We now consider Proposition 3, which predicts that the quality of growers with VC is higher than the quality of growers without VC (first-order stochastic dominance). This is a prediction that depends on the specific functional form of our model, and therefore we view it as a validation of the structure we propose. In Figures 3A and 3B, we investigate this by simply plotting the distribution of quality for growth with VC, and growth without VC. First-order stochastic dominance means that at any point of the distribution there is more mass to the right of the distribution for VC growers than for non-VC growers (simply said, the cumulative density distribution (CDF) for VC growers is always *below* the CDF for non-VC growers). Figure 3A shows the kernel density and CDF of the founding entrepreneurial quality of both types of growers. Consistent with our model, the CDF of quality of

firms that grow without VC is always above the quality of non-VC growers, and the PDF crosses only once (single crossing). That is, the quality of venture-backed growers first-order stochastically dominate non-venture-backed growers.

ESTIMATING THE RELEVANCE OF PASSIVE VS PROACTIVE GROWTH IN THE ECONOMY. Having empirically shown the importance of founding choices on selection into venture capital and growth, we move back to the theoretical description of Section II to evaluate the incidence of passive and proactive growers in overall equity growth events in the economy. Recall that in the model the role of proactive growth is defined by:

$$\zeta = 1 - \alpha^L = \alpha^H = \alpha_{00}^H + \alpha_{10}^H + \alpha_{01}^H + \alpha_{11}^H$$

Proactive growth is represented by the share of firms that grow because they are of high quality, but may or may not receive VC funding. This measure, however, is not observable because we cannot distinguish between proactive growers that grow without the predictive founding choices or VC (α_{00}^H), and passive growth (α^L). In Section II, we show that we can estimate a lower bound of ζ under the assumption that $\alpha_{00}^H = 0$, and an upper bound by considering the fact that $\alpha_{00}^H/\alpha_{10}^H$ is not lower than $\alpha_{10}^H/\alpha_{11}^H$. Here we use our three key indicators, *Delaware, Patent,* and *Trademark,* to represent the type of founding choices that reveal quality in the model (i.e. $K = 1$). We interpret as a high quality founding choice the presence of *any* of the three, and then split all growth firms into four groups: firms that growth without founding choices nor VC (composed of $\alpha^L + \alpha_{00}^H$), growth with founding choices but no VC (α_{10}^H), growth with VC but no founding choices (α_{01}^H), and growth with both founding choices and VC (α_{11}^H).

Table 7A documents the number of firms in each of these groups and the share they represent of all growers in the economy in different samples. (7A-1) includes all firms, and we identify 13,292 growth firms in this set, of which 52% grow with neither founding choices nor venture capital, while the remaining 48% grows with at least one of the two. Among the latter group, the majority is represented by firms with founding choices, but no venture capital (31%), with the remainder split between firms that have both (12%), or only have VC funding (5%). The fact that this last share is so small is consistent with the idea that early IP and stronger governance are a complements to venture capital funding (Hellman and Puri, 2002; Gans, Hsu, and Stern, 2008). Bringing these estimates together, we conclude that the lower bound for the share of

proactive growers is 48% of all firms in our sample, and the upper bound is 61%.[15] Conversely, between 39% and 52% of growth in the economy is passive.

In (7A-2) and (7A-3), we explore the same shares across different geographies. In particular, we compare MSAs with firms that are very attractive to venture capitalists and can be therefore considered "startup hubs" (mean $\hat{\mu}$ in the top 10% the distribution, e.g. Silicon Valley), to non-hubs (mean $\hat{\mu}$ in the bottom 25%, e.g. Florida). The differences are striking. Whereas the share of proactive growers in startup hubs is between 71% and 83%, this drops dramatically to 27% to 34% in non-hubs.[16] Interestingly, the difference between these two estimates appears to stem almost completely from the presence or absence of VC funding, since the share of proactive growth in the absence of VC is closer (34% vs 21%, respectively). This result is consistent with the existence of two paths to growth for proactive growers, and with low substitutability between these two paths.

Last, in (7A-4) and (7A-5) we study two different economic phases, a boom (1995-1999) and a bust period (2002-2005). The shares of types between these two periods are similar. However, the share of non-VC growth, either through passive growth, or through proactive growth without VC, is *higher* during the boom period. This result stands in contrast with statements about an oversupply of venture capital in the late 1990's (Goldfarb et al, 2007), and is instead consistent with the idea that the supply of startups with growth potential was possibly even higher – particularly in regions with less VC presence.

In Table 7B, we replicate this exercise changing the definition of equity growth. While we still include IPOs, we now only classify as successful acquisitions those with a valuation greater than $100 million. The idea is to identify a set of extremely selective growth events that likely have disproportionate impact on the economy. When we apply this higher threshold for what we define as a successful growth event, we observe a sizable change in the relative role of passive versus proactive growth. Only 22% of growth now occurs without founding choices nor venture capital, and as much as 71% includes founding choices (6% includes VC but no founding choices).

---

[15] The upper bound is due to an effort to include also the pro-active growth that occurs among firms with no founding choices and no VC (α_{00}^H). As show in Proposition 4, the estimate of the upper bound depends on $\frac{\alpha_{00}^H}{\alpha_{10}^H} < \frac{\alpha_{01}^H}{\alpha_{11}^H}$ so that $\alpha_{00}^H = \frac{\alpha_{01}^H \alpha_{10}^H}{\alpha_{11}^H}$ is an upper bound for this group. Using the estimates in Table 7, this number is 13%, adding it to 48% leads to an upper bound estimate of 61%.

[16] The upper estimates of α_{00}^H in this case are 12% and 8%, respectively.

Bringing these estimates together, we conclude that when looking at the right tail of growth events, the lower bound for the share of proactive growers is 78% of all firms in our sample, and the upper bound is 94%.[17] Conversely, between 6% and 22% of growth is passive. This result is striking, and highlights the prevalence of proactive growth choices among those firms associated with the highest level of growth.

OTHER OUTCOMES. A potential concern with our analysis thus far is that it has focused on financing outcomes that reflect only one of many measures of firm growth. While the financing outcomes of IPO or a high value acquisition reasonably indicate wealth created for the entrepreneur (and its investors), the nature of firm growth on outcomes such as employment or productivity, which drive the critical process of economic reallocation and business dynamism (Decker et al, 2014; Decker et al, 2016), could be different. In Table 8, we study the relationship between our founding observables and growth employment and revenue outcomes, and the role of active versus passive growth in driving them. To do so, we match all our business registrations to the annual snapshot files of the business database Reference USA/Infogroup (a competitor of Dunn and Bradstreet). We then create indicators for all firms that have reached certain thresholds of economic activity within the first 10 years in that sample: whether the firm has more than 500 employees in any of the first 10 years, whether the firm has more than 1000 employees, and whether the firm has sales higher than $100 million dollars. While we recognize that the Reference USA data is noisy and imperfect, this noise is more likely to exist for smaller firms; we assume that, conditional on being identified by Reference USA as having more than 500 or 1000 employees (or sales over $100 M), a firm's employment (or sales) level is likely to be high and above this threshold.[18]

Our results in Table 8 indicate striking similarities between equity growth success and other high growth outcomes—particularly employment growth. Panel A replicates our nonparametric model of (3-3) for these outcomes. The coefficients of (8A-1), (8A-2), and (8A-3), are all highly positive and significant, and show the same pattern of increasing (and very high returns) to the presence of multiple indicators of founding growth intent. Panel B uses the employment outcome (employment over 500) to estimate the incidence of proactive growth using this outcome. Using

---

[17] The upper bound estimates of α_{00}^H is 16%.
[18] We study the relationship at the macroeconomic level between models predicted indexes using high employment in Reference USA and equity growth in Guzman and Stern (2016).

our lower-bound estimate, we find that at least 44% of all firms are proactive growers, close to the 48% estimated in equity growth.

Bringing these together, we conclude that our analysis on the nature of firm growth indicates a systematic regularity about the role of both passive and proactive growth in the US economy (during our time period), and is not sensitive to the definition of growth itself.

AN UPPER BOUND OF THE RETURNS TO VENTURE CAPITAL FINANCING. In this last section, we compare proactive growers with versus without VC financing to estimate an upper bound to the returns to VC (Proposition 4). In particular, Table 9 runs a logit model with *Equity Growth* as the dependent variable and introduces our founding choices and additional variables to control for selection (on observables) into VC. 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 subsample we did not use to develop the predictive approach).

Column 1 of Table 9 reports the correlation between the presence of VC and equity growth without any controls. Firms that raise VC are 452 times more likely to grow relative to those that do not. Columns 2 through 4 introduce our measures of founding choices, first independently, then as interactions. The table also includes fixed-effects for geography (state fixed-effects), and time-periods (year fixed-effects). The estimated coefficient drops to 36X in Column 2, and then slightly in (8-3). In Column 5 we instead introduce the log odds of *VC Likelihood* with the estimate dropping further to 32X. It is interesting to note that this estimate is an order of magnitude smaller than the initial naïve correlation, but continues to be quite large. In Column 6 we extend the approach by performing an exact matching procedure on the empirical likelihood of VC.[19] 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 the remaining firms being matched at the MSA and state level. After matching, we estimate the differences in the odds of achieving an equity growth outcome between firms that received VC funding and 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 32 to 11. This estimate is significant: conditional on the *VC Likelihood*, firms that raise VC are still

---

[19] Our rationale for using matching comes from Imbens and Rubin (2015), who recommend using matching after estimating a propensity score (such as our measure) to further improve balance on estimating treatment effects.

11 times more likely to achieve an equity growth outcome than non-VC-funded firms. However, while significant, the coefficient is also two orders of magnitude lower than the original estimate from Column 1. While the gap is meaningful, 99% of the difference in outcomes between VC-backed and non-VC-backed firms is accounted for by simple choices that are observable at founding.

Table 10 extends the previous table by introducing a series of additional fixed effects to control for regional and micro-geographic 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: in Columns 2 through 4, VC-funded firms continue to be approximately 10 to 11 times more likely to grow than their counterparts. Column 5 also includes the additional observables at founding beyond Delaware, patent, and trademark, and the coefficient drops again to 6.9, suggesting these observables do account for some potential variation in founding differences even after controlling for fine-grained geographic and time controls. 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.

Taken together, results from Tables 9 and 10 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 naïve estimation VC-funded firms are 452 times more likely to grow than other firms, this estimate is reduced to only 6.9 times using our matching approach and founding observables. This is consistent with our descriptive results on selection presented in Figures 2 and 3, and confirms that VCs select firms that are already of very high quality based on observables.[20] Our exercise places an upper bound on how much value, on average, VCs may be adding to the firms they invest in.

In Table 11 we divide the sample by startup hubs versus not (Columns 2 to 4), and over economic cycles (Columns 5 to 7). Estimates are higher outside of hubs, where VC-funded firms

---

[20] 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.

are 7 times more likely to grow than their counterparts, and when follow-on capital is more likely to be available (as in the .com boom). 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.

Last, In Table 12, we estimate the same relationship for the right tail of the observable *VC Likelihood* distribution (firms in the top 5%, 1%, 0.1% and 0.05%). As we move up the distribution, the contribution of VCs to growth is drastically reduced. For firms that exhibit extremely high, observable quality at incorporation (Column 4 and 5), VC-funded firms are only 3 times more likely to grow than similar firms that do not receive VC funding. This group represents a sizable share of all VC funded firms (34%).

VI. CONCLUSION

Though venture capitalists are often considered central to the process of firm growth, a significant portion of the firms that IPO or are acquired do not rely on venture capital financing. In this paper, we benchmark two competing models of firm growth, passive growth and learning (Gibrat, 1931; Jovanovic, 1982; Gibrat, 1931; Iriji and Simon, 1977; Cohen and Klepper, 1992; Luttmer, 2007; Gabaix, 2009) versus proactive growth (Aghion and Howitt, 1992; Ericson and Pakes, 1995; Klette and Kortum, 2004), to understand their relative contribution to growth in the economy, and to assess if multiple paths to growth exist for proactive growers, independent of how they fund their scaling. In particular, we use a parsimonious model to show how entrepreneurs and VCs optimizing for equity growth make choices at (or near) the time of founding that are informative of underlying quality, and thus predictive of both selection into VC and performance. These founding choices allow us to separate firms of high potential and with growth ambitions from the rest of firms that might grow through a passive model or ex-post market interactions.

The model offers four interrelated predictions on the relationship between founding observables, VC financing, and firm growth. First, entrepreneurial quality estimated from firm choices at birth is a good predictor of VC financing. Second, the same founding choices can also be used to predict VC financing, and the resulting "VC-likelihood" in turn predicts performance even among non-VC-backed firms. Third, the distribution of quality for firms that grow with VC is higher (first-order stochastic dominance) than the quality of firms that grow without VC. This,

in turn, allows us to recover a lower bound of the incidence of passive growth in the economy (and also estimate an upper bound under certain assumptions about the distribution of payoffs). Last, using the VC-likelihood to control for selection into VC (on observables) we are able to recover an upper bound on the effect of VCs on proactive growers.

We test these predictions in the context of the US venture capital industry, from 1995-2005. The data allows us to observe all companies at founding independent of financing, and follow three key paths to growth: passive growth, proactive growth with venture capital financing, and proactive growth in the absence of venture capital. Consistent with the model, founding choices are good predictors of both VC funding and equity growth, and the estimated VC-likelihood also predicts growth within the non-VC sample, which suggests that proactive growers share commonalities irrespective of how they fund their growth. Once VC-funded firms are benchmarked against a more comparable set of other proactive growers that did not raise VC (instead of the full population of firms), we are able to estimate an upper-bound to the returns to VC, which corresponds to a 6X increase in the chances of overall growth, and a substantially smaller 3X increase for proactive growers in the top 0.05% of the entrepreneurial quality distribution.

Our results also highlight that an astonishing 78% to 94% of high impact equity growth events in the economy — IPOs or acquisitions above $100 million — comes from a proactive growth process in which founders and investors make deliberate choices and investments based on a firm's potential, with the finding being even more pronounced within innovative MSAs. Moreover, across regions and time periods, approximately 50% of proactive growth events comes from firms that never raise venture capital, suggesting that while this source of capital plays a major role in accelerating a sizable fraction of proactive growers in the economy, it is not the only path to growth for them.

REFERENCES

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*.

Akcigit, Ufuk, and William R. Kerr. 2018. "Growth through heterogeneous innovations." *Journal of Political Economy* 126.4: 1374-1443

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. *forthcoming*. "Choosing Between Growth and Glory". *Management Science.*

Catalini, Christian, and Xiang Hui. 2017. "Online Syndicates and Startup Investment". *SSRN Working Paper #2997710*

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.

Cohen, W.M. and Klepper, S., 1992. "The anatomy of industry R&D intensity distributions." *The American Economic Review*, pp.773-799.

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

Davis, Steven and John Haltiwanger. 1992. "Gross Job Creation, Gross Job Destruction, and Employment Reallocation". *The Quarterly Journal of Economics*. 107(3): 819-863

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The role of entrepreneurship in US job creation and economic dynamism." *Journal of Economic Perspectives* 28 (3): 3-24.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2016. "Declining business dynamism: What we know and the way forward." *American Economic Review* 106, no. 5: 203-07.

Ericson, Richard, and Ariel Pakes. 1995. "Markov-perfect industry dynamics: A framework for empirical work." *The Review of Economic Studies* 62 (1): 53-82.

Gabaix, Xavier. 2009. "Power laws in economics and finance." *Annual Review of Economics.* 1 (1):255-294.

Gans, Joshua S, David H Hsu, and Scott Stern. 2008. "The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence from Patent Grant Delays." *Management Science.* 54 (5): 982-997.

Gibrat, Robert. 1931. *Les inégalités économiques; applications: aux inégalités des richesses, à la concentration des entreprises, aux populations desvilles, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel*. Paris: Librairie du Recueil Sirey.

Goldfarb, *Brent*, David Kirsch, and David A. Miller. 2007. "Was there too little entry during the Dot Com Era?." *Journal of Financial Economics* 86.1: 100-144.

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. 2016. "The State of American Entrepreneurship: New Estimates of The Quantity and Quality of Entrepreneurship for 32 US States, 1988-2014." *NBER Working Paper.*

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.

Hellmann, Thomas, and Manju Puri. 2002. "Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence." *The Journal of Finance.* 57 (1): 169-197.

Hsu, David H., and Rosemarie H. Ziedonis. 2013. "Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents." *Strategic Management Journal* 34.7: 761-781.

Hurst, Erik and Benjamin W. Pugsley. 2011. "What do small business do?" *Brookings Papers on Economic Activity*. 43 (2): 73-142.

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

Iriji, Yuji and Herbert Simon. 1977. *Skew Distributions and the Sizes of Business Firms.* North-Holland Pub. Co.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica: Journal of the Econometric Society*: 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 Josh Lerner. 2017. "Venture Capital Data." *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges* 75: 413.

Klette, Tor Jakob, and Samuel Kortum. 2004. "Innovating firms and aggregate innovation." *Journal of Political Economy* 112 (5): 986-1018.

Kerr, William, and Shihe Fu. 2008. "The Survey of Industrial R&D—patent Database Link Project." *The Journal of Technology Transfer* 33(2): 173–86.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R.W., 1998. "Law and finance." *Journal of Political Economy*, *106*(6), pp.1113-1155.

Lerner, Josh. 1994. "Venture Capitalists and the Decision to Go Public." *Journal of Financial Economics.* 35 (3): 293-316.

Lerner, Josh. 1995. "Venture Capitalists and the Oversight of Private Firms." *The Journal of Finance.* 51 (1): 301-318.

Lerner, Josh, Antoinette Schoar, Stanislav Sokolinski, and Karen Wilson. 2018. "The globalization of angel investments: Evidence across countries." *Journal of Financial Economics.* 127 (1): 1-20.

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

Luttmer, Erzo GJ. 2007. "Selection, growth, and the size distribution of firms." *The Quarterly Journal of Economics* 122.3: 1103-1144.

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.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70.1 : 41-55.

Stone, Brad. 2013. *The Everything Store: Jeff Bezos and the Age of Amazon.* Little, Brown, and Company. 1st Edition. p. 384.

Sutton, John. 1997. "Gibrat's legacy." *Journal of Economic Literature* 35.1: 40.

TABLE 1
Summary Statistics

Measure	Source	Description	Mean	Std. Dev.
Firm Outcomes				
Venture Capital	Multiple	1 if the firm receives venture capital financing, 0 otherwise.	0.0008	0.028
Venture Capital (2 years)	Multiple	1 if the firm receives venture capital financing within 2 years of founding, 0 otherwise.	0.0005	0.023
Equity Growth (IPO or Acquisition)	SDC Platinum IPO and M&A.	1 if firm has an equity growth event in the first 10 years.	0.0013	0.036
Growth-Oriented Initial Choices				
Delaware	Business Reg.	1 if the firm's jurisdiction is Delaware	0.036	0.186
Trademark	USPTO	1 if the firm acquires for a trademark within 1 year of founding.	0.0009	0.031
Patent	USPTO	1 if the firm acquires for a patent application within 1 year of founding.	0.0020	0.045
Other Founding Choices				
Corporation	Business Reg.	1 if a firm is a corporation (not an LLC or partnership)	0.589	0.492
Short Name	Business Reg.	1 if the firm's name length is 3 words or less (including firm type (e.g. "inc."))	0.468	0.499
Eponymous	Business Reg.	1 if the firm's name includes the president or CEO first or last name.	0.079	0.270
Industry Dummies	Business Reg.	If firm name is associated to an industry group (see Appendix for details).		
Observations			10,451,896	

This table represents our full dataset, comprised of all registered firms registered within the years 1995 and 2005 in 34 US states. These states account for 83% 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 and our approach more generally, can be found in Guzman and Stern (2018). Growth IPOs include only 'true' startup IPOs; we exclude all financial IPOs, REITs, SPACs, reverse LBOs, re-listings, and blank check corporations.

TABLE 2
Venture Growth with and Without Venture Capital

	Equity Growth=1	Equity Growth=0	Fraction that Grow
Firms with Venture Capital	2,280	5,984	28%
Firms without Venture Capital	11,012	10,432,620	0.11%
Fraction of Firms with Venture Capital	17%	0.06%	

Firms within our sample that achieve equity growth by distribution of Venture Capital financing. IPOs are taken from SDC Capital and exclude all re-listings, reverse LBOs, SPACs, REITs, blank check companies, and financial IPOs.

TABLE 3
Predictive Analytics Model of Equity Growth
Dependent Variable: Equity Growth
Training Sample (50% Random Sub-Sample)
Logit model. Incidence Rate Ratios Reported

	(1)	(2)	(3)	(4)
Independent Effects				
Delaware	8.823***			
	(0.270)			
Patent	20.36***			
	(0.977)			
Trademark	6.793***	6.544***		
	(0.670)	(0.609)		
Delaware, Patent Interactions				
Delaware = 1, Patent = 0		9.848***		
		(0.302)		
Delaware = 0, Patent = 1		37.52***		
		(2.476)		
Delaware = 1, Patent = 1		146.3***		
		(7.945)		
Delaware, Patent, Trademark Interactions				
Delaware = 1, Patent = 0, Trademark = 0			9.683***	8.754***
			(0.304)	(0.284)
Delaware = 0, Patent = 1, Trademark = 0			42.83***	22.36***
			(2.740)	(1.522)
Delaware = 0, Patent = 0, Trademark = 1			21.34***	15.76***
			(2.620)	(1.963)
Delaware = 1, Patent = 1, Trademark = 0			176.4***	82.76***
			(9.095)	(4.686)
Delaware = 1, Patent = 0, Trademark = 1			118.8***	75.21***
			(11.72)	(7.876)
Delaware = 0, Patent = 1, Trademark = 1			78.85***	42.75***
			(19.65)	(11.29)
Delaware = 1, Patent = 1, Trademark = 1			279.4***	135.8***
			(41.24)	(21.99)
Other Controls	No	No	No	Yes
N	5225947	5225947	5225947	5225947
pseudo R-sq	0.158	0.159	0.162	0.203

Robust standard errors reported in parenthesis. * $p < .1$, ** $p < .05$, *** $p < .01$

TABLE 4

Does Entrepreneurial Quality Predict VC Financing?
Dependent Variable: Venture Capital
50% Test Sample

	(1) OLS	(2) OLS	(3) Logit	(4) Logit
Entrep. Quality ($\hat{\theta}$)	0.823***			
	(0.0246)			
Ln(Entrep. Quality) (ln($\hat{\theta}$))		0.00666***	3.268***	
		(0.000144)	(0.0299)	
Distribution of Quality				
Baseline: < 50%. ($\hat{\theta}$ < .0006)				
50% - 90%				4.169***
($\hat{\theta} \in$ [.0006 , .00018])				(0.405)
90% - 95%				18.65***
($\hat{\theta} \in$ [.0018, .0026])				(2.057)
95% - 99%				62.51***
($\hat{\theta} \in$ [.0026 , .014])				(5.872)
99% - 99.5%				164.6***
($\hat{\theta} \in$ [.014 , .022])				(17.86)
> 99.5%				645.0***
($\hat{\theta} \in$ [.022 , .45])				(65.49)
State F.E.	No	Yes	Yes	Yes
Year F.E.	No	Yes	Yes	Yes
Observations	5225949	5225949	5225949	5225949
R-squared	0.036	0.014		
Pseudo R-squared			0.262	0.244
Log-Likelihood			-24373.5	-24990.6

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 .0005 and a standard deviation of .0064. Standardized VC Likelihood changes this measure to have a standard deviation of 1. The mean value of the outcome variable is 0.0004. Robust standard errors reported. *** p < .001.

TABLE 5

Predictive Analytics Model of Venture Capital Financing
Dependent Variable: Venture Capital
Training Sample (50% Random Sub-Sample)
Logit model. Incidence Rate Ratios Reported

	(1)	(2)	(3)	(4)
Independent Effects				
Delaware	13.25***			
	(0.503)			
Patent	28.39***			
	(1.507)			
Trademark	2.071***	2.164***		
	(0.307)	(0.284)		
Delaware, Patent Interactions				
Delaware = 1, Patent = 0		16.64***		
		(0.629)		
Delaware = 0, Patent = 1		80.72***		
		(5.432)		
Delaware = 1, Patent = 1		297.4***		
		(17.24)		
Delaware, Patent, Trademark Interactions				
Delaware = 1, Patent = 0, Trademark = 0			16.52***	15.62***
			(0.633)	(0.632)
Delaware = 0, Patent = 1, Trademark = 0			86.87***	40.01***
			(5.792)	(2.886)
Delaware = 0, Patent = 0, Trademark = 1			19.29***	13.15***
			(3.457)	(2.398)
Delaware = 1, Patent = 1, Trademark = 0			342.4***	137.9***
			(19.11)	(8.595)
Delaware = 1, Patent = 0, Trademark = 1			100.4***	54.24***
			(13.76)	(8.054)
Delaware = 0, Patent = 1, Trademark = 1			52.27***	24.44***
			(20.52)	(10.32)
Delaware = 1, Patent = 1, Trademark = 1			164.5***	62.38***
			(34.90)	(14.98)
Other Controls	No	No	No	Yes
N	5225947	5225947	5225947	5225947
pseudo R-sq	0.248	0.252	0.256	0.332

Robust standard errors reported in parenthesis. * $p < .1$, ** $p < .05$, *** $p < .01$

TABLE 6

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 ($\hat{\mu}$)	0.689***			
	(0.0262)			
Log VC Likelihood ($\ln(\hat{\mu})$)		0.00407***	2.334***	
		(0.000101)	(0.0201)	
Distribution of VC Likelihood				
Baseline: < 50%. ($\hat{\mu} < .0002$)				
50% - 90%				2.708***
($\hat{\mu} \in [.0002 , .0011]$)				(0.197)
90% - 95%				10.16***
($\hat{\mu} \in [.0011 , .0015]$)				(0.899)
95% - 99%				20.44***
($\hat{\mu} \in [.0015 , .012]$)				(1.474)
99% - 99.5%				44.64***
($\hat{\mu} \in [.012 , .022]$)				(4.300)
> 99.5%				161.5***
($\hat{\mu} \in [.022 , .39]$)				(13.74)
State F.E.	No	Yes	Yes	Yes
Year F.E.	No	Yes	Yes	Yes
Observations	5221901	5221901	5221901	5221901
R-squared	0.015	0.006		
Pseudo R-squared			0.124	0.113
Log-Likelihood			-38046.5	-38550.4

VC Likelihood is the predicted probability that a firm gets venture capital given its characteristics at founding, predicted from (5-3). It has a mean of .0005 and a standard deviation of .0064. Standardized VC Likelihood changes this measure to have a standard deviation of 1. The mean value of the outcome variable is 0.0004. Robust standard errors reported. *** p < .001.

TABLE 7A

Estimates of the Incidence of the Passive and Proactive Firm Growth Models
Based on Founding Choices and the Receipt of Venture Capital
Number of Growth Firms

Category	Model Variable	All Firms (1)	Top 10% $\hat{\mu}$ MSAs (2)	Bottom 25% $\hat{\mu}$ MSAs (3)	Boom: 1995-1999 (4)	Bust: 2002-2005 (5)
All Firms	$\alpha^L + \alpha^H_{00} + \alpha^H_{10} + \alpha^H_{01} + \alpha^H_{11}$	13,292 (100%)	1,348 (100%)	1,574 (100%)	6,423 (100%)	4,352 (100%)
No Founding Choices, No VC	$\alpha^L + \alpha^H_{00}$	6,934 (52%)	396 (29%)	1,156 (73%)	3,501 (55%)	2,282 (52%)
Only Founding Choices (Delaware, Patent, or Trademark)	α^H_{10}	4,078 (31%)	453 (34%)	329 (21%)	1,800 (28%)	1,384 (32%)
Only VC	α^H_{01}	667 (5%)	136 (10%)	23 (1%)	460 (7%)	125 (3%)
Both Founding Choices and VC	α^H_{11}	1,613 (12%)	363 (27%)	66 (4%)	662 (10%)	561 (13%)
Share of Proactive Growers						
Lower Bound		0.48	0.71	0.27	0.45	0.48
Upper Bound		0.61	0.83	0.34	0.65	0.55

Represents all firm growth events (IPO or acquisition within 10 years of founding) across multiple different regimes of funding and early governance and IP choices. Founding choices are those firms that make choices at founding that are indicative of high growth intention. We define these choices are taking either a Delaware registration, or registering intellectual property near founding, either as a patent or a trademark. Top 10% MSAs are the 10% of MSAs with highest average VC likelihood. Bottom 25% MSAs are the bottom 25% by average VC likelihood. For all cases, we create 50 bootstrap samples of our complete dataset, and estimate the mean value for each column. Reported standard errors (in parenthesis) are taken as the standard deviation of this sample. In Panel B, the difference is the value of the column minus the value of the row.

TABLE 7B

Estimates of the Incidence of the Passive and Proactive Firm Growth Models
For High Value Growth Outcome (IPO or Acquisition with Equity Value > $100 Million)
Based on Founding Choices and the Receipt of Venture Capital
Number of Growth Firms

	Model Variable	All Firms (1)	Top 10% $\hat{\mu}$ MSAs (2)	Bottom 25% $\hat{\mu}$ MSAs (3)	Boom: 1995-1999 (4)	Bust: 2002-2005 (5)
Category						
All Firms	$\alpha^L + \alpha_{00}^H + \alpha_{10}^H + \alpha_{01}^H + \alpha_{11}^H$	1962 (100%)	260 (100%)	166 (100%)	1139 (100%)	513 (100%)
No Founding Choices, No VC	$\alpha^L + \alpha_{00}^H$	439 (22%)	25 (10%)	73 (44%)	273 (24%)	119 (23%)
Only Founding Choices (Delaware, Patent, or Trademark)	α_{10}^H	1002 (51%)	132 (50%)	68 (41%)	538 (47%)	281 (54%)
Only VC	α_{01}^H	126 (6%)	23 (9%)	5 (3%)	103 (9%)	16 (3%)
Both Founding Choices and VC	α_{11}^H	395 (20%)	82 (31%)	21 (13%)	230 (20%)	104 (20%)
Share of Proactive Growers						
Lower Bound		0.78	0.90	0.56	0.76	0.77
Upper Bound		0.94	1.05	0.66	0.97	0.85

Represents all firm growth events (IPO or acquisition within 10 years of founding) across multiple different regimes of funding and early governance and IP choices. Founding choices are those firms that make choices at founding that are indicative of high growth intention. We define these choices are taking either a Delaware registration, or registering intellectual property near founding, either as a patent or a trademark. Top 10% MSAs are the 10% of MSAs with highest average VC likelihood. Bottom 25% MSAs are the bottom 25% by average VC likelihood. For all cases, we create 50 bootstrap samples of our complete dataset, and estimate the mean value for each column. Reported standard errors (in parenthesis) are taken as the standard deviation of this sample. In Panel B, the difference is the value of the column minus the value of the row.

TABLE 8

PANEL A. Predictive Analytics Model of Growth: Other Outcomes
Training Sample (50% Random Sub-Sample)
Logit model. Incidence Rate Ratios Reported

Dependent Variable	(1) Employment > 500	(2) Employment > 1000	(3) Sales > $100M
Delaware = 1, Patent = 0, Trademark = 0	12.18***	13.83***	6.719***
	(0.948)	(1.791)	(0.378)
Delaware = 0, Patent = 1, Trademark = 0	51.61***	114.8***	17.63***
	(7.979)	(22.04)	(2.607)
Delaware = 0, Patent = 0, Trademark = 1	46.35***	12.24***	18.32***
	(10.11)	(8.708)	(3.722)
Delaware = 1, Patent = 1, Trademark = 0	174.4***	173.5***	61.40***
	(22.75)	(39.45)	(7.144)
Delaware = 1, Patent = 0, Trademark = 1	164.6***	146.7***	71.73***
	(33.38)	(53.73)	(12.84)
Delaware = 0, Patent = 1, Trademark = 1	218.4***	384.0***	135.4***
	(85.33)	(198.1)	(39.64)
Delaware = 1, Patent = 1, Trademark = 1	415.2***	621.7***	89.44***
	(110.2)	(232.9)	(28.25)
Other Controls	No	No	No
N	5106501	5024117	5115879
pseudo R-sq	0.115	0.118	0.067

PANEL B. Estimates of the Incidence of the Passive and Proactive Firm Growth Models
Based on Founding Choices and the Receipt of Venture Capital
Number of Growth Firms
GROWTH = EMPLOYMENT > 500

	Model Variable	All Firms (1)	Top 10% $\hat{\mu}$ MSAs (2)	Bottom 25% $\hat{\mu}$ MSAs (3)	Boom: 1995-1999 (4)	Bust: 2002-2005 (5)
Category						
All Firms	$\alpha^L + \alpha^H_{00} + \alpha^H_{10} + \alpha^H_{01} + \alpha^H_{11}$	2311 (100%)	78 (100%)	422 (100%)	1070 (100%)	838 (100%)
No Founding Choices, No VC	$\alpha^L + \alpha^H_{00}$	1295 (56%)	18 (23%)	287 (68%)	598 (56%)	490 (58%)
Only Founding Choices (Delaware, Patent, or Trademark)	α^H_{10}	960 (42%)	50 (64%)	129 (31%)	443 (41%)	335 (40%)
Only VC	α^H_{01}	17 (1%)	2 (0%)	2 (0%)	13 (1%)	4 (0%)
Both Founding Choices and VC	α^H_{11}	39 (2%)	8 (10%)	4 (1%)	16 (1%)	9 (1%)
Share of Proactive Growers						
Lower Bound		0.44	0.77	0.32	0.44	0.42
Upper Bound		0.62	0.93	0.47	0.78	0.59

TABLE 9
Venture Capital and Growth Outcomes Controlling for Observables and VC Likelihood
Dependent Variable: Equity Growth

	All Firms					Exactly Matched Sub-sample
	(1)	(2)	(3)	(4)	(5)	(6)
Venture Capital (2 Years)	451.5***	35.78***	34.56***	32.99**	31.69***	10.61***
	(47.06)	(4.687)	(4.449)	(4.242)	(2.363)	(1.103)
Independent Effects						
Delaware		7.364***				
		(0.722)				
Patent		11.06***				
		(0.994)				
Trademark		7.131***	7.144***			
		(0.768)	(0.739)			
Delaware, Patent Interactions						
Delaware = 1, Patent = 0			8.182***			
			(0.801)			
Delaware = 0, Patent = 1			22.68***			
			(2.081)			
Delaware = 1, Patent = 1			62.54***			
			(7.804)			
Delaware, Patent, Trademark						
Delaware = 1, Patent = 0,				8.322**		
				(0.838)		
Delaware = 0, Patent = 1,				26.09**		
				(2.443)		
Delaware = 0, Patent = 0,				21.11**		
				(3.550)		
Delaware = 1, Patent = 1,				71.66**		
				(9.256)		
Delaware = 1, Patent = 0,				64.38**		
				(9.743)		
Delaware = 0, Patent = 1,				43.89**		
				(13.01)		
Delaware = 1, Patent = 1,				257.1**		
				(43.69)		
Entrepreneurial Quality Controls						
Log-Odds of VC Likelihood					2.476***	
					(0.0255)	
State F. E.	No	Yes	Yes	Yes	Yes	Yes
Year F.E.	No	Yes	Yes	Yes	Yes	Yes
N	5225949	5225949	5225949	5225949	5225949	5248
pseudo R-sq	0.085	0.189	0.190	0.191	0.191	0.175

We estimate the relationship between entrepreneurial quality indicators and firm equity growth. All regressions are run on a 50% test sample drawn separately from the 50% training sample used to VC-Likelihood in (5-3).

TABLE 10
VC Financing and Equity Growth Outcomes
Dependent Variable: Equity Growth Outcome.
Logit Regression, Matched Firms. Incidence-Rate Ratios Reported.

	Baseline Model	Extra Controls			Preferred Model
	(1)	(2)	(3)	(4)	(5)
Venture Capital (2 Years)	10.61***	11.16***	11.35***	10.85***	6.86**
	(1.103)	(1.198)	(1.212)	(1.131)	(0.77)
Year F. E.	Yes		Yes	Yes	Yes
State F. E.	Yes				Yes
State X Year F. E.		Yes			
MSA F. E.			Yes	Yes	
Control for Average Neighbor Quality				Yes	
Other Controls					Yes
N	5248	4922	4848	5248	5248
Pseudo R-sq	0.175	0.184	0.189	0.182	0.204

Robust standard errors in parenthesis. 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. Robust standard errors in parenthesis. *** p < .01

TABLE 11
VC Financing and Equity Growth Outcomes.
Dependent Variable: Equity Growth.
Logit Regression, Matched Sample. Incidence-Rate Ratios Reported.

	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)
Venture Capital (2 Years)	6.862***	5.812***	6.521***	8.117***	11.42***	5.203***	5.679***
	(0.774)	(0.980)	(0.891)	(1.632)	(2.501)	(0.905)	(1.145)
State F.E.	Yes				Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes			
N	5248	1856	3120	2128	1372	2220	1435
Pseudo R-sq	0.204	0.194	0.198	0.193	0.241	0.167	0.193

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. A company is in a Startup Hub if it is in the CBSAs of Boston-Cambridge-Quincy, MA-NH, New York-Northern New Jersey-Long Island, NY-NJ-PA, San Francisco-Oakland-Fremont, CA, San Jose-Sunnyvale-Santa Clara, CA, Austin-Round Rock, TX, or Seattle-Tacoma-Bellevue, WA. Robust standard errors in parenthesis. *** p < .001

TABLE 12

VC Financing and Equity Growth Outcomes
Dependent Variable: Equity Growth Outcome.
Logit Regression, Matched Firms. Incidence-Rate Ratios Reported.

	All Firms	Within the Quality Distribution			
		Top 5%	Top 1%	Top 0.1%	Top 0.05%
	(1)	(2)	(3)	(4)	(5)
Venture Capital (2 Years)	6.862***	5.987***	4.868***	3.818***	3.527***
	(0.774)	(0.702)	(0.619)	(0.578)	(0.623)
N	5248	3849	2648	1269	850
Pseudo R-sq	0.204	0.196	0.161	0.131	0.112

Robust standard errors in parenthesis. 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. *** $p < .01$

FIGURE 1

The Incidence of Venture Capital Financing Events in the Quality Distribution
Out of Sample 10-Fold Validation

Percentile
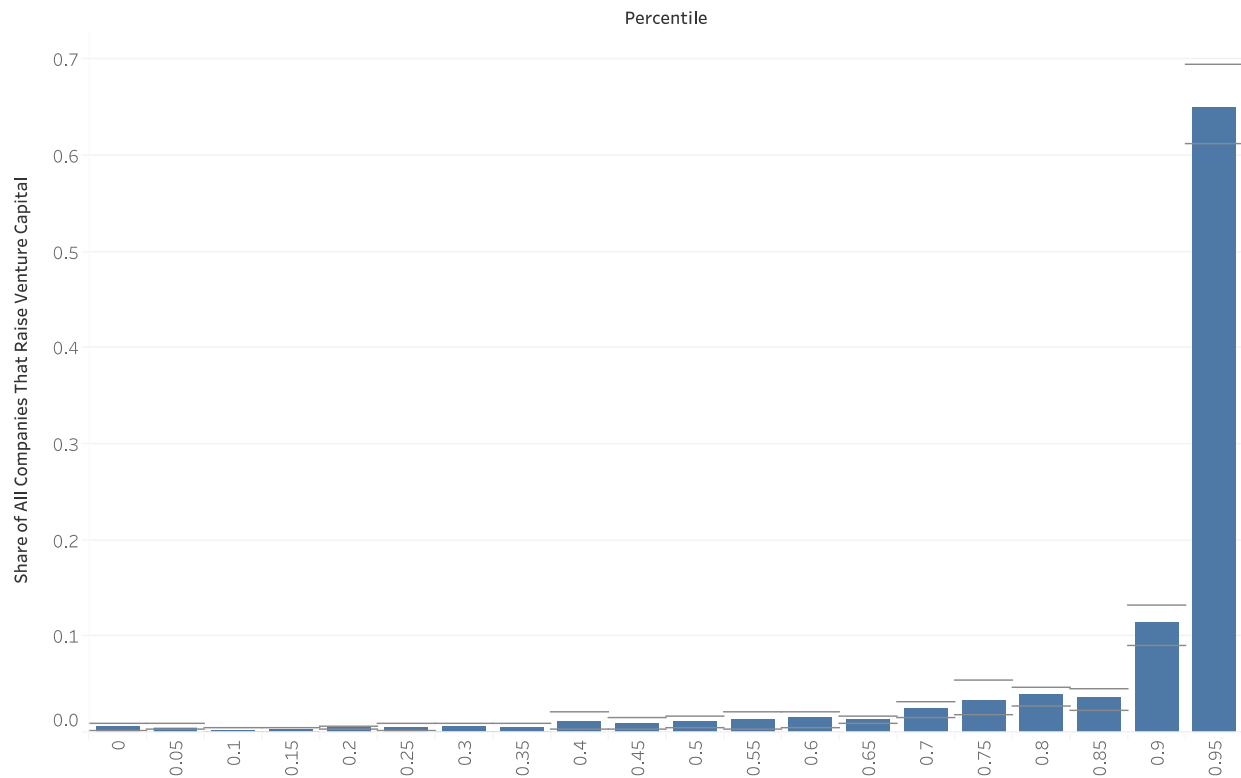
Notes: Figure 1 reports the out of sample incidence of venture capital events in the predicted quality distribution. Predicted quality is the predicted probability of equity growth, according to model (3-3). We estimate this model 10 times for 10 60% random samples of our data, then document the location of the VC events across the quality distribution in the remaining 40% of our data. The height of the bar is the mean estimated incidence of VC events, while the range displayed are the maximum and minimum of this estimate.

FIGURE 2

The Incidence of Growth Events in the Predicted Venture Capital Distribution
Out of Sample 10-Fold Validation
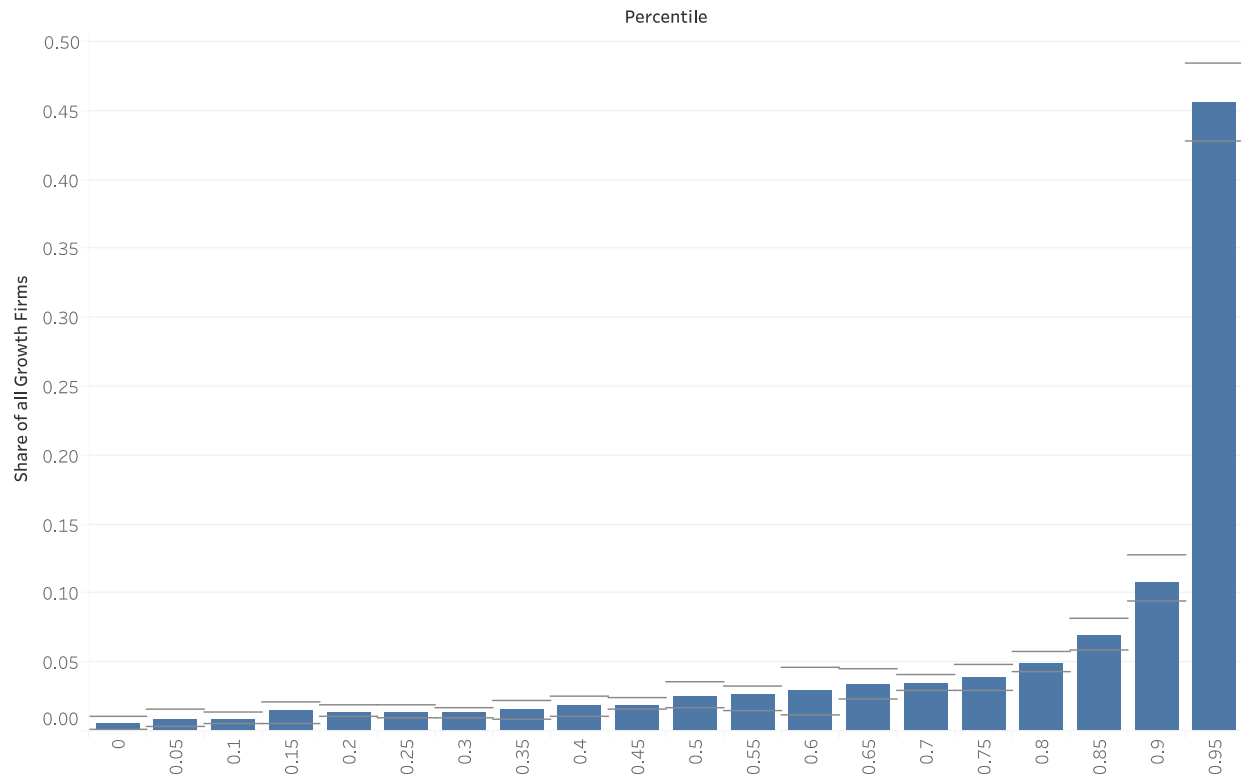
Notes: Figure 2 reports the out of sample incidence of equity growth events in the predicted venture capital distribution, for firms that never raise venture capital. Predicted venture capital is the predicted probability of raising VC, according to model (5-3). We estimate this model 10 times for ten 60% random samples of our data, then document the location of the growth events across the VC distribution in the remaining 40% of our data, after excluding VC funded events in this test sample. The height of the bar is the mean estimated incidence of growth, while the range displayed are the maximum and minimum of this estimate.

FIGURE 3

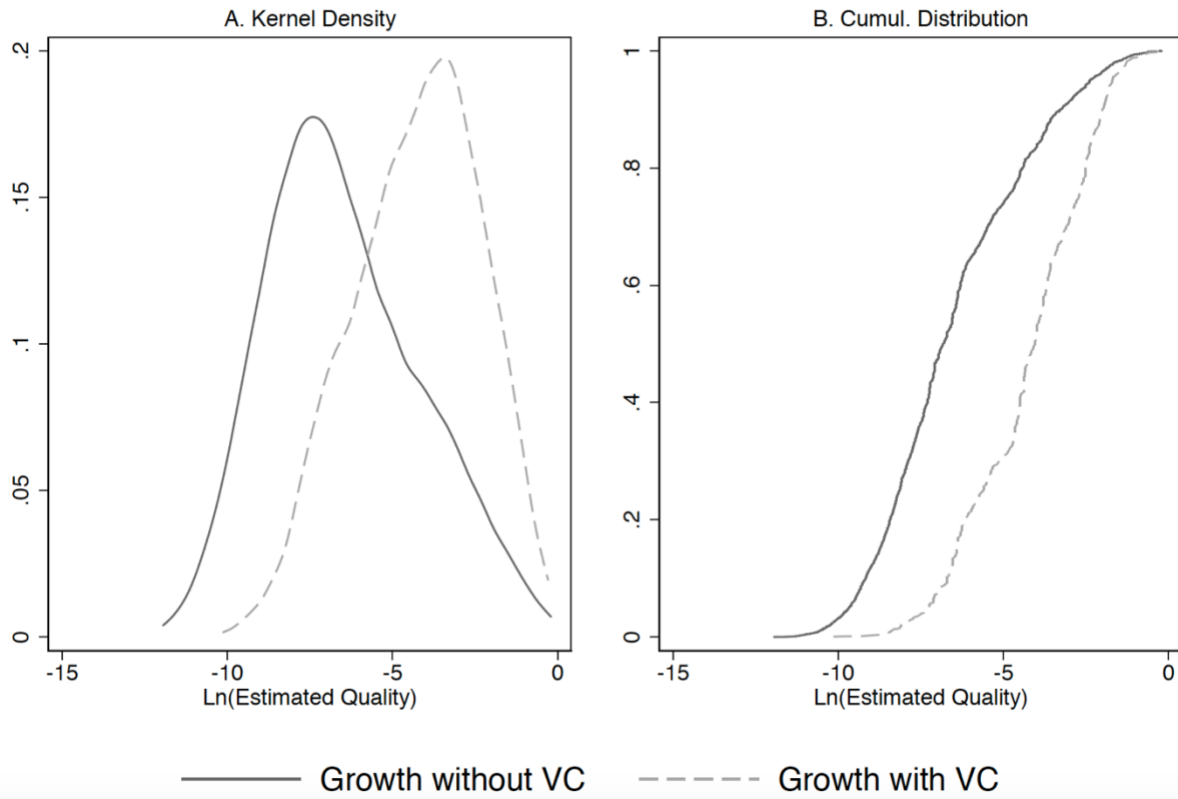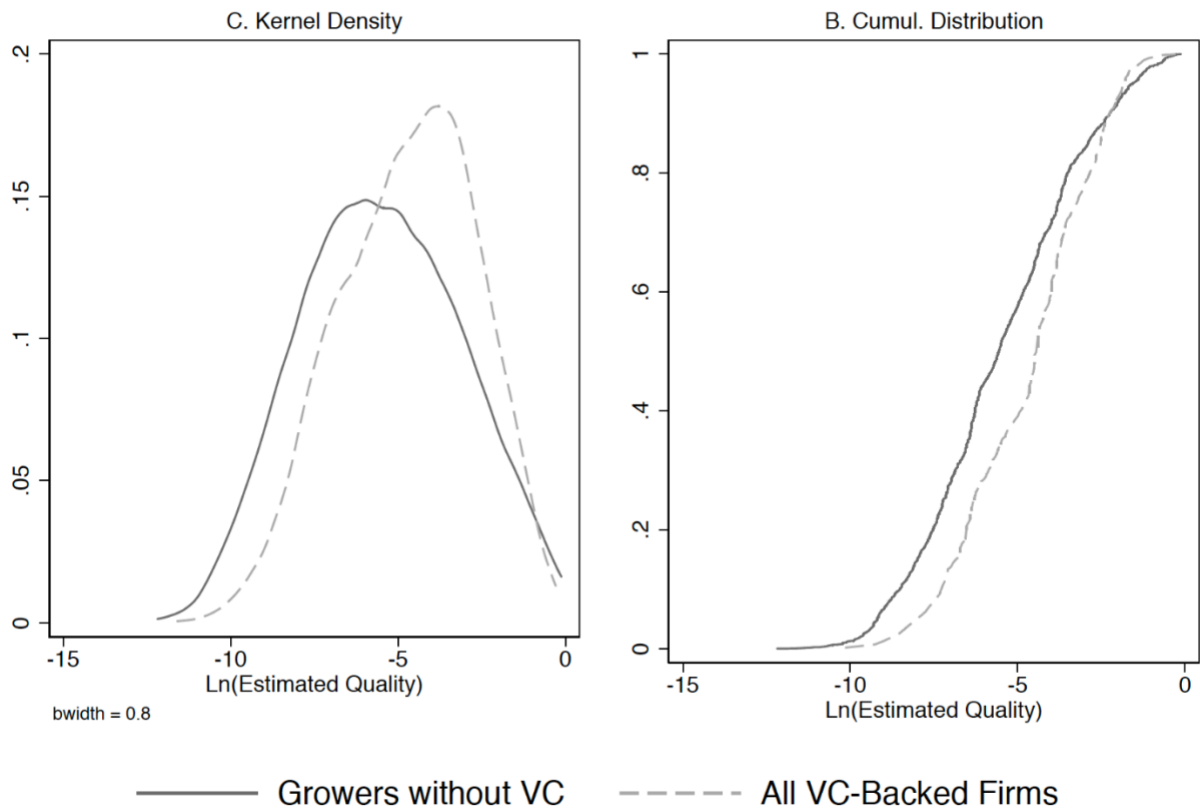## Distribution of Quality for Growth with VC and Growth Without VC

FIGURE 4

Distribution of Quality for all VC-Backed Firms and Growth Without VC

MODEL PROOFS

Lemma 1 (Startup Choices Inform Entrepreneurial Quality) There exists a function $h(\boldsymbol{K}) \rightarrow \hat{\theta}$ such that given a function $I_X(\theta)$, representing the information on a variable on θ,

$$I_{\hat{\theta}}(\theta) > I_{\bar{\theta}}(\theta), \hat{\theta} = h(\boldsymbol{K}), \bar{\theta} = \hat{E}[Y] \tag{A1}$$

Proof. Consider first the maximization function of the startup to choose $\boldsymbol{K}$, when there is only one choice ($\boldsymbol{K} = \{K_1\}$). The startup chooses $K_1 = 1$ if $\theta f(K_1, E[V|K_1, \theta, c_V]) > K_1 c_{K_1}$. Some firms choose $K_1 = 1$ and others do not (by the model's assumption). The question is if

$$E[\theta|K_1 = 1] > E[\theta|K_1 = 0] \tag{A2}$$

If this is true, then we know K_1 has information on θ. We proceed in two steps. We first assume that the probability of VC is a constant value P_V for all firms. In this case, the firm simply solves $K_1 = 1\{\theta f(K_1, P_V) > K_1 c_{K_1}\}$ and it is easy to see that there is a threshold value of $\frac{c_{K_1}}{f(1, P_V)}$ above which it chooses K_1. In this case, the presence of K_1 indicates being above a certain threshold, and, therefore, (A2) is true. Though this threshold is randomly allocated (due to a random c_{K_1}) but it nonetheless still separates any firm by quality at some value. Now, we can instead consider what happens when P_V depends on the quality of the firm $P_V(K, \theta) = E[V|K, \theta, c_V]$. Note that P_V increases with quality. The net effect of this is that the threshold is lowered for firms of high quality, which makes them even more likely to take K. That is, for high quality, the potential noise effect of c_K is muted when VC comes into the picture, making the use of K as a signal for quality even cleaner. Finally, to consider these for many choices, remember that f is positively increasing in K_j for all j. Therefore, if the above relationship holds with one K ($n = 1$), it would hold with many. ∎

Proposition 1 (Estimated Quality Positively Predicts Venture Capital)

$$\frac{\partial P(V = 1|\hat{\theta})}{\partial \hat{\theta}} > 0 \tag{A3}$$

Proof. This is straightforward from the VC's maximization choice. The probability that a VC invest given a true θ is $P(V = 1|\theta) = P(\theta f(K, V) > (1 + c_V)V)$. Since c_V is random and independent of θ, then it is straightforward that inequality is more likely to be true at higher values of θ, i.e. $\frac{\partial P(V = 1|\theta)}{\partial \theta} > 0$. The next question is how this relates to $\hat{\theta}$. For (A3) to be true we need $\hat{\theta}$ to be informative of θ. This is already established in Lemma 1, and, therefore, is it. ∎

Lemma 2 (Startup Choices Inform Venture Capital) Define $\mu = P(V = 1)$, and $I_X(\mu)$ as the information of some variable on μ. Then, there exists a function $h'(\boldsymbol{K}) \rightarrow \hat{\mu}$ such that

$$I_{\hat{\mu}}(\mu) > I_{\bar{\mu}}(\mu), \hat{\mu} = h'(\boldsymbol{K}), \bar{\mu} = \hat{E}[V] \tag{A4}$$

Proof. The logic begins in a similar way to Lemma 1. Consider first the maximization function of the VC to choose V with a single choice $K_1 \in \{0,1\}$ that has already been made. The VC chooses $V = 1\{\theta f(K, V) > (1 + c_V)V\}$.
The question that we would like to know is if

$$E[V|K = 1] > E[V|K = 0] \tag{A5}$$

If this is true, then we know K has information on V. As in Lemma 1, for any firm, there is a threshold value c_V above which investment profitable. This threshold is lowered when the quality of the firm is higher which, in turn, means that firms with higher quality are more likely to raise VC. Since K is informative of quality, it is therefore also informative of V. ∎

Proposition 2 (VC Likelihood Predicts Growth in Firms Without VC)

$$\frac{\partial P(g = 1|\hat{\mu}, V = 0)}{\partial \hat{\mu}} > 0 \tag{A6}$$

Proof. Similar to Proposition 1. Since performance is $y = \theta f(K, V)\epsilon$, and growth is $g = 1\{y > y^*\}$. We can write the function of (A6) as $P(\theta f(K, V)\epsilon > y^*) = P(g = 1|\mu)$. Since $\mu = P(\theta[f(K, 1) - f(K, 0)] > (1 + c_V))$ (from (5)), then it is easy to see that μ increases when either quality or governance increases. Because these are also the main components of performance then firms with higher μ (i.e. better quality or governance) are also more likely to grow.

$$\frac{\partial P(g = 1|\mu)}{\partial \mu} > 0 \tag{A7}$$

Whether this also holds for the non-VC subsample ($V = 0$) requires one more condition. If all non-VC growth is through ϵ (i.e. $\alpha_{00}^H = 0$ and $\alpha_{10}^H = 0$) then it would not hold, but this scenario is trivial. Instead, if either of these is positive, then (A7) is true even in the non-VC sample. Finally, to consider if this holds with the empirical VC Likelihood, we simply need $\hat{\mu}$ to be informative, but this is already established in Lemma 2. ∎

Proposition 3 (Quality of growth with VC is higher (first order stochastically dominates) than quality of growth with VC)

$$P(\theta > x|g = 1, V = 1) > P(\theta > x|g = 1, V = 0) \ \forall \ x$$

Proof. The condition for this to be true is that, conditional on growth, the probability of $V = 1$ is increasing in θ. That is, $\frac{\partial P(V = 1|g = 1, \theta)}{\partial \theta} > 0$. We already know that $\frac{\partial P(V = 1|\theta)}{\partial \theta} > 0$, since this follows directly from the VC choice equation (eq. (4)). But we would like to know if this changes when we consider $\frac{\partial P(V = 1|\theta f(K, V)\epsilon > y^*, \theta)}{\partial \theta} > 0$. This can only happen if $\theta f(K, V)$ has a higher value for lower quality firms, which would require a heterogeneous $f(K, V)$ that is unrealistic and not an assumption of our model. ∎

Proposition 4 (Upper Bound Estimates for ζ)

$$\eta < 1 \ \Rightarrow \ \alpha_{00}^H < \alpha_{10}^H \, \alpha_{11}^H \, / \, \alpha_{01}^H \tag{A8}$$

Proof. Recall that $\eta = \dfrac{\left(\frac{\alpha_{00}^H}{\alpha_{10}^H}\right)}{\left(\frac{\alpha_{01}^H}{\alpha_{11}^H}\right)} = \left(\dfrac{\left(1 - F\left(\frac{y^*}{\theta^H}\right)\right)}{\left(1 - F\left(\frac{y^*}{\theta^H f(1,0)}\right)\right)}\right) \Bigg/ \left(\dfrac{\left(1 - F\left(\frac{y^*}{\theta^H f(0,1)}\right)\right)}{\left(1 - F\left(\frac{y^*}{\theta^H f(1,1)}\right)\right)}\right)$ and that F is distributed logistic and that it being considered under a growth threshold model where the value inside the function is the threshold. Also, assume that the value of $\frac{y^*}{\theta^H f(1,1)}$ (and consequently all others) is to the right of the mean (i.e. they are all relatively rare events). Define $y_{ij}^{*H} = \frac{y^*}{\theta^H f(i,j)}$ as the threshold

value of F. Because governance and VC are independent by assumption, the setup of the model naturally implies that the value-add of VC is constant for both regimes on performance

$$\frac{y_{00}^{*H}}{y_{10}^{*H}} = \frac{y_{01}^{*H}}{y_{11}^{*H}} = 1/\delta^*$$

Yet, this is not the case in probability space. Namely, for any constant $\delta > 1$,

$$\frac{F(\delta a)}{F(a)} > \frac{F(\delta b)}{F(b)}, \forall\, b > a \text{ and } \delta b < mean(F)$$

Translating this to the right-tail of the distribution and inverting the terms, it means that

$$\frac{1 - F(a)}{1 - F(\delta a)} < \frac{1 - F(b)}{1 - F(\delta b)}, \forall\, b < a \text{ and } b > mean(F)$$

And since $y_{01}^{*H} < y_{00}^{*H}$ this implies $\eta < 1$.∎

Proposition 5 (Upper Bound Estimates of Returns to VC)

$$\hat{E}[\hat{\tau}|\hat{\mu}] \geq E[\tau|\mu]$$

Proof. This is easiest to see as an omitted variable. Consider $\mu = a_1\hat{\mu} + a_2\dot{\mu}$, where both $\hat{\mu}$ and $\dot{\mu}$ are correlated with μ, and assume

$$g_i = \alpha + \tau V_i + \mu_i + e_i \hspace{4cm} (A9)$$

is correctly specified. If we instead estimate $g_i = \alpha + \tau V_i + \hat{\mu}_i + e_i$ then $E[\hat{\tau}] = \tau + a_2\dot{\mu}$ which is higher. ∎

ONLINE APPENDIX

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.	0.194	0.396
Traded Industry	Business Reg.	If firm name is associated to a traded industry.	0.535	0.499
Resource Intensive Industry	Business Reg.	If firm name is associated to a resource intensive industry.	0.130	0.337
IT	Business Reg.	If firm name is associated to the IT industry cluster.	0.025	0.156
Biotechnology	Business Reg.	If firm name is associated to the Biotechnology industry cluster.	0.002	0.044
E-Commerce	Business Reg.	If firm name is associated to the E-Commerce industry cluster.	0.052	0.222
Medical Devices	Business Reg.	If firm name is associated to the Medical Devices industry cluster.	0.030	0.172
Semiconductor	Business Reg.	If firm name is associated to the Semiconductor industry cluster.	0.0005	0.023
Observations			10,451,896	

This table represents our full dataset, comprised of all registered firms registered within the years 1995 and 2005 in 34 US states. These states account for 83% 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	0.589	0.788	0.589	No VC Financing	0.079	0.017	0.079
VC Financing	0.901	0.954	0.909	VC Financing	0.009	0.005	0.009
All	0.589	0.824		All	0.079	0.015	
Delaware				*Patent*			
No VC Financing	0.035	0.478	0.036	No VC Financing	0.002	0.163	0.002
VC Financing	0.518	0.629	0.533	VC Financing	0.207	0.341	0.225
All	0.036	0.510		All	0.002	0.201	
Short Name				*Trademark*			
No VC Financing	0.468	0.742	0.468	No VC Financing	0.001	0.065	0.001
VC Financing	0.886	0.920	0.890	VC Financing	0.034	0.046	0.036
All	0.468	0.780		All	0.001	0.061	

TABLE A3

Share of firm in IPO and Acquisition Samples
that Raise Venture Capital

	IPO	Acquisition
Firms with VC Financing	288	1,091
(Share)	(33%)	(20%)
Firms without VC Financing	590	4,390
(Share)	(67%)	(80%)

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
Full Specification Including Other Controls for VC and Quality Models

Dependent Variable	(1) Equity Growth	(2) Venture Capital
Delaware, Patent, Trademark Interactions		
Delaware = 1, Patent = 0, Trademark = 0	8.754*** (0.284)	15.62*** (0.632)
Delaware = 0, Patent = 1, Trademark = 0	22.36*** (1.522)	40.01*** (2.886)
Delaware = 0, Patent = 0, Trademark = 1	15.76*** (1.963)	13.15*** (2.398)
Delaware = 1, Patent = 1, Trademark = 0	82.76*** (4.686)	137.9*** (8.595)
Delaware = 1, Patent = 0, Trademark = 1	75.21*** (7.876)	54.24*** (8.054)
Delaware = 0, Patent = 1, Trademark = 1	42.75*** (11.29)	24.44*** (10.32)
Delaware = 1, Patent = 1, Trademark = 1	135.8*** (21.99)	62.38*** (14.98)
Other Controls		
Corporation	2.151*** (0.0705)	5.493*** (0.304)
Short Name	3.158*** (0.0983)	5.966*** (0.310)
Eponymous	0.292*** (0.0251)	0.158*** (0.0263)
Industry Controls		
Local	0.415*** (0.0265)	0.324*** (0.0350)
Traded	1.091** (0.0317)	0.764*** (0.0286)
Resource Intensive	0.966 (0.0383)	0.655*** (0.0415)
IT	2.274*** (0.113)	3.219*** (0.189)
Biotechnology	2.628*** (0.272)	2.509*** (0.313)
e-commerce	1.569*** (0.0718)	1.642*** (0.105)
Medical Devices	1.329*** (0.0738)	1.225** (0.0831)
Semiconductor	1.725** (0.306)	2.397*** (0.433)
N	5225947	5225947
pseudo R-sq	0.203	0.332

TABLE A5
Does Entrepreneurial Quality Predict VC Financing?
Regressions Including Other Controls
Dependent Variable: Venture Capital
50% Test Sample

	(1) OLS	(2) OLS	(3) Logit	(4) Logit
Entrep. Quality ($\hat{\theta}$)	0.813***			
	(0.0246)			
Ln(Entrep. Quality) (ln($\hat{\theta}$))		0.00658***	2.769***	
		(0.000143)	(0.0267)	
Distribution of Quality				
Baseline: < 50%. ($\hat{\theta} < .0006$)				
50% - 90%				3.437***
($\hat{\theta} \in [.0006 , .00018]$)				(0.326)
90% - 95%				13.79***
($\hat{\theta} \in [.0018, .0026]$)				(1.533)
95% - 99%				49.84***
($\hat{\theta} \in [.0026 , .014]$)				(4.877)
99% - 99.5%				108.8***
($\hat{\theta} \in [.014 , .022]$)				(11.83)
> 99.5%				301.3***
($\hat{\theta} \in [.022 , .45]$)				(30.92)
State F.E.	No	Yes	Yes	Yes
State F.E.	No	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes
Observations	5225949	5225949	5225949	5225949
R-squared	0.037	0.015		
Pseudo R-squared			0.330	0.325
Log-Likelihood			-24373.5	-22135.9

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 .0005 and a standard deviation of .0064. Standardized VC Likelihood changes this measure to have a standard deviation of 1. The mean value of the outcome variable is 0.0004. Robust standard errors reported. *** p < .001.

TABLE A6

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

	(1) OLS	(2) OLS	(3) Logit	(4) Logit
VC Likelihood ($\hat{\mu}$)	0.680***			
	(0.0262)			
Log VC Likelihood ($\ln(\hat{\mu})$)		0.00400***	2.127***	
		(0.000101)	(0.0190)	
Distribution of VC Likelihood Baseline: < 50%. ($\hat{\mu} < .0002$)				
50% - 90%				2.460***
($\hat{\mu} \in [.0002 , .0011]$)				(0.174)
90% - 95%				8.382***
($\hat{\mu} \in [.0011 , .0015]$)				(0.735)
95% - 99%				16.07***
($\hat{\mu} \in [.0015 , .012]$)				(1.164)
99% - 99.5%				33.98***
($\hat{\mu} \in [.012 , .022]$)				(3.222)
> 99.5%				95.65***
($\hat{\mu} \in [.022 , .39]$)				(8.034)
State F.E.	No	Yes	Yes	Yes
Year F.E.	No	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes
Observations	5221901	5221901	5221901	5221901
R-squared	0.016	0.007		
Pseudo R-squared			0.159	0.151
Log-Likelihood			-38046.5	-36544.2

VC Likelihood is the predicted probability that a firm gets venture capital given its characteristics at founding, predicted from (5-3). It has a mean of .0005 and a standard deviation of .0064. Standardized VC Likelihood changes this measure to have a standard deviation of 1. The mean value of the outcome variable is 0.0004. Robust standard errors reported. *** p < .001.

TABLE A7

Venture Capital and Growth Outcomes Controlling for Observables and VC Likelihood
Including Other Controls
Dependent Variable: Equity Growth

	All Firms					Exactly Matched Sub-sample
	(1)	(2)	(3)	(4)	(5)	(6)
Venture Capital (2 Years)	580.8***	13.65***	14.08***	12.81**	10.53***	6.86**
	(90.80)	(3.234)	(2.956)	(2.733)	(1.412)	(0.77)
Independent Effects						
Delaware		19.39***				
		(3.348)				
Patent		16.15***				
		(2.657)				
Trademark		5.735***	5.781***			
		(0.998)	(0.950)			
Delaware, Patent Interactions						
Delaware = 1, Patent = 0			26.20***			
			(3.565)			
Delaware = 0, Patent = 1			75.61***			
			(12.64)			
Delaware = 1, Patent = 1			289.2***			
			(58.93)			
Delaware, Patent, Trademark						
Delaware = 1, Patent = 0,				26.85**		
				(3.802)		
Delaware = 0, Patent = 1,				84.04**		
				(14.26)		
Delaware = 0, Patent = 0,				43.85**		
				(15.94)		
Delaware = 1, Patent = 1,				347.5**		
				(74.44)		
Delaware = 1, Patent = 0,				225.4**		
				(55.04)		
Delaware = 0, Patent = 1,				321.8**		
				(131.7)		
Delaware = 1, Patent = 1,				1072.4*		
				(290.7)		
Entrepreneurial Quality Controls						
Log-Odds of VC Likelihood					3.357***	
					(0.0769)	
State F. E.	No	Yes	Yes	Yes	Yes	Yes
Year F.E.	No	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	5225949	5225949	5225949	5225949	5225949	5250
pseudo R-sq	0.118	0.303	0.309	0.311	0.307	0.204

We estimate the relationship between entrepreneurial quality indicators and firm equity growth. All regressions are run on a 50% test sample drawn separately from the 50% training sample used to VC-Likelihood in (5-3).