Go West Young Firm: Agglomeration and Embeddedness in Startup Migrations to Silicon Valley

Jorge Guzman
Columbia University*

March 15, 2019

Abstract

Startup growth requires local resources. Economic geography has emphasized the role of agglomeration in accessing these resources. Sociology and urban planning have instead emphasized the role of social embeddedness. For many startups, this creates a trade-off between their home location, in which they are socially embedded, and a higher agglomeration location, where resources are more available. I present a parsimonious model that jointly explains the role of agglomeration and embeddedness on startup migration, location choice, and performance. Then, using all Delaware jurisdiction firms in 26 US states, I use machine learning and firm fixed-effects models to estimate selection into Silicon Valley (a high agglomeration location) and the impact of moving on movers. Consistent with the model, higher quality firms are more likely to move, and movers leave low agglomeration areas for higher agglomeration areas. Moving to Silicon Valley increases performance under a variety of dimensions. Movers are more likely to be acquired and IPO, raise more financing, patent more, introduce more products, and have higher sales. The benefits appear to be (at least in part) driven by the benefit of knowledge spillovers in Silicon Valley. Looking at movers from 1996 to 2005, the financial benefits are only present before 2001, but non-financial benefits persist after the dot-com bust. The results imply that agglomeration is more important than embeddedness (for these high growth movers), and that the benefits are driven by access to localized knowledge and its influence on both financial and non-financial outcomes.

*I am thankful to Scott Stern, Pierre Azoulay, Christian Catalini, and Olenka Kacperczyk for their advice and support on this project. I also thank Kevin Bryan, Bo Cowgill, JP Eggers, Cathy Fazio, Joshua Gans, Deepak Hedge, David Hsu, Bill Kerr, Daniel Keum, Ramana Nanda, Rob Seamans, Dan Wang, and Sam Zyontz. This project was partially done with support of the Kauffman Foundation Dissertation Fellowship, the Jean Hammond (1986) and Michael Krasner (1974) Entrepreneurship Fund at MIT, and the Edward B. Roberts (1957) Entrepreneurship Fund at MIT. Some descriptive results in this paper were also contained in an earlier working paper titled Entrepreneurial Migration, which this paper has superseded. Email: jag2367@columbia.edu
1 Introduction

All startups start small, but some will grow to be quite large. These ‘high growth’ startups will utilize resources from their environment—such as financial capital, employees, customers, and ideas—to create a profitable product or service in which they hold a competitive advantage. For decades, social scientists have sought to understand the mechanisms through which this local resource accumulation fuels firm growth and productivity (Marshall, 1890; Weber, 1929; Porter, 1990; Krugman, 1991; Saxenian, 1994). At a high level, two dominant and conceptually orthogonal perspectives have emerged.

One perspective, long emphasized by economists (though not exclusively), is agglomeration. Agglomeration focuses on how distance from goods, people, and ideas, creates transportation costs that impact productivity. Using far-away resources is more expensive for a startup, making locations that are closer to resources more productive. But because good locations are competitive and migration costs are positive, firms do not all fall in one destination. Productivity and product differentiation are the result of optimal resource configurations being different in different places. Today, the most relevant of these distance costs seems to be the ability of firms to access knowledge. A firm can engage with local customers and employees by simply locating to a specific city; however, learning innovative ideas and tacit information usually requires being so geographically close that people have face-to-face interactions with relevant counterparts (Davis & Dingel, 2012). Accordingly, Arzaghi and Henderson (2008) and Kerr and Kominers (2014) show that knowledge can only travel a few blocks in Manhattan and about a 15 minute drive in the Bay Area (respectively). And Catalini (2017) shows that collaboration patterns even within the same university change when physical proximity is randomly altered.

A second perspective, more often emphasized by sociologists (though also not exclusively), is embeddedness. Embeddedness focuses on the fact that social (non-economic) relationships, such as family ties, ethnic bonds, or long-held friendships, also help firms access resources to be more productive (Granovetter, 2005; Uzzi, 1996).¹ Granovetter (1985)

¹Uzzi (1996) defines the three key characteristics of embedded economic relationships as ones based on
highlights how the majority of jobs are filled through personal networks rather than through public job postings, so that firms without these personal local relationships must pay a premium to find the right talent. Uzzi (1999) shows that social relationships to bankers allow firms to secure better loan terms, because they provide the banker both additional uncodified information on the creditworthiness of the business owner and a new avenue to punish the business owner in case of default (social costs). In a series of works, Sorenson has documented systematically the importance of embeddedness for startup performance (see Sorenson (2018) for review). The robust conclusion from the embeddedness literature is that startups that are not locally embedded, and that therefore can only transact through arms-length relationships, are at a meaningful performance disadvantage.

In a general sense, each of these perspectives must be partially true. Resources do have transportation costs, and firms certainly benefit from the personal relationships of their managers and employees. Yet, whether one takes an agglomeration or an embeddedness perspective can lead to starkly different predictions for how startups are affected by location choices, and how regional ecosystems are accelerated through entrepreneurship-driven economic growth.

For example, consider an entrepreneur evaluating the optimal location for her startup. If an agglomeration perspective is taken, migration could be a profitable idea for the entrepreneur. A company should move where the best resources are located, as long as the benefit covers the costs of moving. If it is a VC oriented IT, hardware, or biotechnology, company, then this destination might be Silicon Valley. On the other hand, if an embeddedness perspective is taken, migration is much less appealing. While there could be better resources elsewhere, startups benefit from being (and staying) at their home, where founders have friends and other social relationships that would allow them to access local resources at lower cost through personal, rather than arms-length, transactions.

Similar contrasting predictions occur when one uses these theories to design regional trust, fine-grained information transfer, and joint-problem solving.

See also Wang (2015), and Almeida and Kogut (1999).

Some economics papers have focused on the role of networks in entrepreneurship, most significantly Chinitz (1961), Michelacci and Silva (2007), and Buchardi and Hassan (2014).
entrepreneurship policy. The agglomeration perspective suggests that simply increasing the supply of local resources at a location could be valuable in creating regional development. Some development approaches such as the ‘Big Push’ (Murphy et al, 1988), SBIR matching programs (Lanahan and Feldman, 2015), or directly subsidizing startup migration costs (e.g. Startup Chile), build on this idea. On the other hand, an embeddedness perspective instead favors a longer-term local relationship development that allows startups to create profitable networks of firms that efficiently mobilize resources through social relationships and their associated trust.

Given that there is strong theoretical emphasis on both agglomeration and embeddedness in determining startup performance, it is perhaps surprising that there is a paucity of empirical evidence evaluating them together. The closest work are the papers by Dahl and Sorenson (2012) and Michelacci and Silva (2007), who document that more embedded entrepreneurs seem to do better in the population of Danish and Italian firms (respectively); and the paper by Rosenthal and Strange (2012), who consider how differences in embeddedness influence the location choices and performance of female versus male entrepreneurs. While each paper offers an important contribution, these three studies do not focus on explaining and documenting the trade-off between agglomeration and embeddedness. And, perhaps as importantly for linking entrepreneurship to economic growth, they all focus on comprehensive samples of all firms in the economy—which mostly constitute small, non-growth, businesses (Hurst & Pugsley, 2011)—but offer little information on the role of agglomeration and embeddedness on the growth oriented startups that theoretically drive economic growth (Akcigit & Kerr, 2018).

This paper evaluates agglomeration and embeddedness together in growth oriented startups. To do so, I begin in section II with a parsimonious model that is able to bear out some of the key underlying relationships between agglomeration, embeddedness, and firm performance. A key difference between this model and prior models of location and entrepreneurship (e.g Arzaghi & Henderson, 2008) is that startups are not necessarily born in their ideal location, but instead simply in the location where the founders happen to be

---

4By growth entrepreneurship I mean the smaller share of companies that have a perceptible probability of creating high economic growth outcomes in financial, employment, innovation, etc.
and assemble. This leads to a location choice problem after being born that trades off the embeddedness at home with the agglomeration of another destination and the cost of migration. The model also includes a measure of the ‘quality’ of the firm,\(^5\) which becomes a central element of the determinants on which firms choose to move, and of the benefits of location conditional on moving. The model bears out some useful predictions and comparative statics of the patterns of location choice in high growth entrepreneurship.

I then study location choice empirically in a systematic large-scale sample of growth oriented entrepreneurs and their performance after moving (or not) to Silicon Valley. The sample includes all companies born in 26 US states that register under Delaware jurisdiction (an early signal of high growth intention), where I use the process of registration across US states to map the location changes of companies and the migration of their headquarters. I build on the entrepreneurial quality work that uses founding observables to estimate the founding potential of firms (Guzman & Stern, 2016), to develop a machine-learning estimate of firm founding quality. I then confirm the relationships predicted by the theoretical model on the selection into migration. Higher quality firms, and firms located in lower agglomeration regions, are more likely to move, and they tend to move to higher quality regions. To estimate the impact of moving on performance (i.e. the difference in performance between the firm’s founding home and its destination), I use cross-sectional machine learning models (random forest and double LASSO) to create meaningful counterfactuals of non movers, and panel-data models that include both firm fixed-effects and age fixed-effects. The results from both analyses show very similar and sizable benefits from migration to Silicon Valley across equity outcomes (IPO and acquisition), innovation (patenting), product introductions (measured as the filing of trademarks), and sales.

Finally, to get closer to the underlying mechanisms and understand how agglomeration helps the movers, I take advantage of heterogeneity in the destinations within Silicon Valley as well as heterogeneity in the business cycle across migration. First, I estimate whether similar migrants perform differently depending on the entrepreneurship in

\(^5\)This is also in contrast to Arghazi and Henderson, who assume homogeneity in quality.
the Silicon Valley ZIP Code they move to, controlling for migrant characteristics (through machine learning) and other common effects. The key idea is that most agglomeration economies are roughly accessible within a city, but knowledge is only accessible at a much shorter distance. Consistent with a meaningful role of knowledge, I find local entrepreneurship at the destination has a significant effect on the performance of movers. Second, I look at the benefits of moving before and after the dot-com bust to consider the effects of the business cycle on the gains of migrants. I find that while migration benefits on financial outcomes (venture capital, acquisition, and IPO) drop precipitously after the dot-com bust, benefits to non-financial outcomes (trademarks, patents, and sales) persist, suggesting agglomeration impacts both the financial and real economies.

The rest of this paper is organized as follows. Section 2 presents the economic framework, including the theoretical model, the machine learning approach to measure entrepreneurial quality, and the econometric approach. Section 3 introduces the data and explains how individual measures are built. Section 4 presents the empirical results. Finally, Section 5 concludes.

2 Agglomeration, Embeddedness, and Startup Migration

2.1 Economic Model

We begin by considering an individual firm that has been born after a team of founders have assembled around an idea in some specific location. While many entrepreneurial ideas are optimal to the industry structure of the founding region (Sorenson and Audia, 2001), this is not necessarily always the case. Opportunities can stem from connecting knowledge domains that are different and unique (Hayek, 1941; Jacobs, 1965), and it is often unclear to nascent entrepreneurs what is the ideal value-chain, resources, and markets in which a new innovation should be commercialized (Gans, Stern, and Wu, 2017). In general for companies, once created, the location in which a given company has been founded could or
could not be the best place for the company to grow.\textsuperscript{6}

The startup considered in this model maximizes a production function that includes two components. The first component is the startup’s quality, defined as $\theta$, which is a summary statistic of the economic potential of the company given its fundamentals (initial resources, founder human capital, and founder ambitions and aspirations). This value is constant across locations. The second component is local resources, represented by $\tau_j$. How well a startup is able to mobilize resources in turn depends on three firm or location parameters—how available the resources are, their price, and the startup’s level of embeddedness to find them. For simplicity, we assume there exist only two types of resources: knowledge (indexed by $A$) which becomes expensive quickly with distance, and labor (indexed by $L$) which distance influences at a more moderate pace. We also assume that the size of these resources is ‘discrete’ and fixed. That is, the startup will always require values of exactly $\tau_A$ and $\tau_L$ to operate,\textsuperscript{7} independent of the context or the cost of these resources.\textsuperscript{8} The production function of the startup is

$$Y = f(\theta, \tau_A, \tau_L)$$

(1)

Each unit of $\tau_j$ creates a dollar-equivalent benefit $\delta_j$, so that $\tau_j = \delta_j \tau_j$ is the economic value created by resource $j$. The costs of using the resource in region $r$ are $c_j(r) \geq 0$. Following the trade literature, these costs are represented as the share of the economic value that is lost in accessing the resource (i.e. as Samuelson ‘iceberg’ costs). With constant returns to scale for each factor, production is defined as

$$Y(r) = \theta \left( \frac{\tau_A}{1 + c_A(r)} \right) \left( \frac{\tau_L}{1 + c_L(r)} \right) \epsilon$$

(2)

\textsuperscript{6}Consider, for example, the moment in which Bill Gates and Paul Allen decided to found Microsoft. It was potentially their presence in Cambridge, MA that allowed them to be at the forefront of technological development. However, once this was identified, they decided early they needed to be close to their core customer in Albuquerque, NM.

\textsuperscript{7}Effectively, these values should be some specific number $\tau_A^*$ and $\tau_L^*$ but the star is excluded for ease of exposition.

\textsuperscript{8}The advantage of this approach is that will allow us to focus on the costs of distance without having to worry about differences in the price of these goods due to general equilibrium considerations on access.
where \( \epsilon \) is a random, mean one, error term capturing the processes of ‘random-growth’ in firm performance (Gibrat, 1931; Sutton, 1997).

For simplicity, I assume that these resources have no direct purchase cost, so that the only cost is access costs. Access to a resource \( j \) in region \( r \) is determined by the distance to the resource \( d_j(r) \) and some exponent \( \eta \), which represents the rate at which increases in distance increase the access cost. Agglomeration and embeddedness will influence performance in the model by affecting either \( d_j(r) \) or \( \eta \). Agglomeration implies a reduction in \( d_j(r) \). Locations with a better supply of resources simply have a lower distance to them.

\( \eta \) is defined by two parameters. First, there is a resource-specific gradient \( \beta_j \) reflecting how quickly costs increase with distance. The value of \( \beta_j \) is a result of the economic nature of each resource and the technology available for transporting it. It is constant across regions. Second, there is a level of embeddedness \( \gamma(r) \geq 0 \) for each startup in each region. This embeddedness acts as a direct countervailing force to the impact of distance on cost. Well embedded companies can access difficult-to-get-to resources, as if they were close to them. Together, these two parameters create an embeddedness-adjusted gradient for distance. Putting them simply as a ratio, the cost to access resource \( j \) is

\[
c_j(r) = d_j(r)^{\beta_j/(1+\gamma(r))}
\]

Finally, we define \( \bar{\Theta}_r \) as the average quality of local entrepreneurship in region \( r \) and \( \bar{\Theta}_r = \theta_r \times N_r \) as the total quality-adjusted quantity of entrepreneurship in region \( r \).

I now impose several assumptions on the relationship of these parameters which will allow us to be more specific on how they shape startup performance.

**Assumption 1** The local entrepreneurship of a region is positively correlated to its underlying agglomerations (i.e. negatively correlated with distance to resources)

\[
corr(\Theta_r, \bar{d}_r) \ll 0
\]

\(^9\)These measures are also defined as regional indexes in Guzman and Stern (2016, 2017, 2018) as the Entrepreneurial Quality Index (EQI), for \( \bar{\Theta}_r \), and the Regional Entrepreneurship Cohort Potential Index (RECPI) for \( \Theta_r \).
Assumption 1 suggests that places that have stronger agglomeration economies will also have either more startups or higher quality startups (or both) born in a region. This result is at the core of economic geography research since Marshall (1890), and recent empirical evidence continues to show a positive link between startup underlying quality and the underlying agglomeration of a region (e.g. Combes, Duranton, Gobillon, Puga, & Roux, 2012). The assumption will be quite useful in both the theoretical and empirical analysis. It will allow us to use an empirical measure of the local entrepreneurship in a region (of \( \Theta_r \)) as a proxy for the local agglomerations of that region without requiring the measurement of each potential local agglomeration per-se.

**Assumption 2** Knowledge is more costly to access at a distance than labor

\[ \beta_A > \beta_L \] (5)

Assumption 2 states the well established fact that knowledge spillovers decrease quickly with distance while labor markets are much easier to access at a distance (Arzaghi & Henderson, 2008; Kerr & Kominers, 2014). In particular, knowledge spillovers usually require face-to-face interactions and being in the same room (Davis & Dingel, 2012),\(^{10}\) while recruiting usually occurs within a large metropolitan area. The role of Assumption 2 is to create differences in the role of distance in the model across different resources. Towards the end of the model, this will allow me to consider within city heterogeneity, as a measure of access to knowledge.

**Assumption 3** The startup is embedded at its founding location \( r^0 \) and is not embedded in other destinations

\[ \gamma(r^0) > \gamma(r) = 0, \forall r \neq r^0 \] (6)

\(^{10}\)As an intellectual exercise, note how academics find it cost-worthy to travel to a conference and be in the same conference room for the exchange ideas, even though the ideas (and papers) themselves might seem easy to share at a distance. Considering the substantial physical and organizational resources expended by academia to run and participate in academic conferences, the returns to face-to-face interactions for the academic pursuit of knowledge must be quite large.
This assumption is central to the trade-off that is the focus of this paper. It means that when a company is founded, the founders have been living in the location in which they start the company for a minimum amount of time, which begets them social relationships (i.e. embeddedness). These social relationships would decrease transport costs to resources in the founding location $r^0$, but would not in other potential destinations. However, if some other location offers resources that are closer (i.e. higher agglomeration), this creates a trade-off between agglomeration and embeddedness for the firm.

With this setup, we now look at the startup location choice problem. While a complete study of this problem would focus on the systematic evaluation of migration choice across all potential destinations, perhaps even allowing for multiple migrations, and general equilibrium effects, this paper focuses only on an individual piece of this puzzle: the choice to move from the origin region ($r^0$) to a single destination that has been popularly considered the one richest in resources for high growth startups, Silicon Valley (indexed as $r^s$).

Once born in $r^0$, the startup chooses its location by evaluating the expected performance at the destination $E[Y(r)]$, the expected performance at the source $E[Y(r_0)]$, and a migration cost $M(r)$ which is positive for $r = r^s$ and zero for $r = r_0$. The startup’s problem is

$$\max_{r \in \{r_0, r^s\}} E[Y(r)] - E[Y(r_0)] - M(r)$$

(7)

If embeddedness $\gamma(r)$ is a positive value $\gamma$ at home and zero at the destination. An indicator of whether the firm moves, $S$, will be equal to 1 when

$$\theta \left( \frac{\tau_A}{1 + d_A(r^s)^{\beta_A}} \right) \left( \frac{\tau_L}{1 + d_L(r^s)^{\beta_L}} \right) - M > \theta \left( \frac{\tau_A}{1 + d_A(r_0)^{\beta_A}} \right) \left( \frac{\tau_L}{1 + d_L(r_0)^{\beta_L}} \right)$$

(8)

Using this structure, we can now develop some key propositions and claims on the economics of migration to Silicon Valley for high growth startups.
2.1.1 Selection into Migration

**Proposition 1** Higher quality startups are more likely to move

\[
\frac{\partial P(S = 1)}{\partial \theta} > 0 \tag{9}
\]

**Proof:** In Appendix.

The intuition of this proposition comes from the nature of agglomeration benefits and migration costs. While migration costs are constant, agglomeration benefits are complementary to quality, creating positive sorting. Migration is not too valuable for lower quality firms, but higher quality firms could find substantial benefits. For example, very high quality companies that move to Silicon Valley might see an increase in the amount of venture capital they are able to fundraise, but companies below a certain level might not see an increase in their financing or might not see any financing at all. This creates very large potential gains to migration for high quality firms, and lower gains for other firms.

**Lemma 1** Startups closer to resources at founding are less likely to move

\[
\frac{\partial P(S = 1)}{\partial d_j(r^0)} > 0, \forall j \tag{10}
\]

**Proof:** In Appendix.

This is straightforward to see from equation (8). Whether there are gains in the destination requires that, at least for one resource \(j\), distance to this resource in Silicon Valley \((d_j(r^s))\) is lower than distance at origin \((d_j(r^0))\) to compensate for both the loss of embeddedness and the costs of moving. Therefore, as the availability of resources at home increases \((d_j(r^0)\) lowers), the startup is less likely to move.

**Proposition 2** Startups in locations with higher entrepreneurship are less likely to move

\[
\frac{\partial P(S = 1)}{\partial \theta_v} > 0 \tag{11}
\]
Proof: Building from Assumption 1 and Lemma 1, it follows naturally that because local entrepreneurship is a useful proxy for the quality of local agglomerations, firms surrounded by better startups are less likely to move. This is a testable comparative static on the role of regional versus firm quality, and their correlation to the choice to move.

2.1.2 Agglomeration vs Embeddedness in Migration to Silicon Valley

Definition. The net-agglomeration benefit of Silicon Valley for a startup that chooses to move (treatment on the treated), accounting for loss of embeddedness and adjusted for quality, is

\[
\Delta = E \left[ \frac{Y(r^*) - Y(r^0) - M}{\theta} \bigg| S = 1 \right]
\] (12)

The goal of this paper would then be to estimate \( \hat{\Delta} \) to evaluate the role of agglomeration compared to embeddedness, and, in particular, understand the extent to which it improves (or not) performance for those companies that choose to move.

2.1.3 Heterogeneity within Silicon Valley: The Value of Knowledge Spillovers

We can then expand this setup to also consider the role of location within Silicon Valley, and how heterogeneity across time and place influences performance. To do so, I consider two possible micro-regions in Silicon Valley. I define a core region \( r_c \) and a non-core region \( r_n \). Both regions offer roughly the same access to the specialized labor market and to other generalized elements of Silicon Valley (e.g. visiting venture capitalists), that is \( c_L(r_c) = c_L(r_n) \). However, the core is closer to other startups and entrepreneurially minded people, so that access to knowledge is easier in the core than outside of it (i.e. \( c_A(r_c) < c_A(r_n) \)).

If the startup is not only choosing to move to Silicon Valley, but also choosing between moving to the core and moving to the non-core, Proposition 3 follows.

Proposition 3 If knowledge matters for performance, then performance is higher for
firms that locate to micro-regions with better local entrepreneurship

\[ \frac{\partial E[Y|S=1, \theta]}{\partial \Theta_r}, r \in \{r^c, r^n\} \] (13)

**Proof:** In Appendix.

This proposition creates a second comparative static. It directly suggests a regression that uses measures of local entrepreneurship at the micro-destinations, and controls for firm quality, should show a positive relationship of local entrepreneurship to performance.

### 2.1.4 Heterogeneity in Time: The Role of Business Cycle

Finally, to get deeper into the role of the business cycle, I consider the importance of differences in timing of the move to Silicon Valley by considering two different potential periods, a boom period \((t = 0)\) and a bust period \((t = 1)\). The key difference between these two periods is the availability of financial capital in each one. A boom period has all good things for entrepreneurs: a large amount of financial capital, talented individuals, and strong new ideas. A bust period is weaker on capital, but the scientists and engineers (and the underlying ideas that they develop) are still locally present in the short-run. The final theoretical relationship in this paper is presented as a claim, rather than as a proven proposition within the model.

**Claim.** The effects of the ‘real economy’ for Silicon Valley are positive during both boom and bust periods, but the effects of the ‘financial economy’ attenuate during the bust.

### 2.2 Estimating Entrepreneurial Quality

To be able to map this setup to an empirical analysis, we need an approach for estimating entrepreneurial quality \(\theta\), for each startup that reflects a firm’s underlying ‘potential’. This challenge is particularly difficult in entrepreneurship, where the performance of the
firm has, by definition, not been observed. In Guzman and Stern (2015), we offer a novel methodology using predictive analytics to estimate quality as the empirical likelihood that a company will succeed, at founding.

The approach combines three interrelated insights. First, as the challenges to reach a growth outcome as a sole proprietorship are formidable, a practical requirement for any entrepreneur to achieve growth is business registration (as a corporation, partnership, or limited liability company). This practical requirement allows me to form a quasi-population of entrepreneurs ‘at risk’ of growth at a similar (and foundational) stage of the entrepreneurial process. Second, it is possible to distinguish among business registrants by observing choices the founders make at or close to the time of registration informed by their own ambitions and expectations for the firm. Examples of these choices include whether the founders name the firm after themselves (eponymy), whether the firm is organized in order to facilitate equity financing (e.g., registering as a corporation or in Delaware), and whether the firm seeks intellectual property protection (e.g., a patent or trademark). Third, I leverage the fact that, though rare, it is possible to observe meaningful growth outcomes for some firms (e.g., those that achieve an IPO or high-value acquisition).

Combining these insights, consider a firm fully characterized by many (even infinite) founding observables $Z_i$. Then, entrepreneurial quality can be defined as simply the relationship between a specific growth outcomes $g_i$ and these founding startup characteristics. Specifically, for a firm $i$ born in region $r$, and a growth outcome $g_i$ observed $s$ years after founding, quality is

$$\theta_i = P(g_i|Z_i)$$  \hspace{1cm} (14)$$

Given a subset of observed founding characteristics $Z'_i \in Z_i$, an empirical estimate of quality can then be consistently estimated as the predicted out of sample probability of measured founding characteristics on performance—i.e. $\hat{\theta}_i = \hat{P}(g_i|Z'_i)$.

This paper uses machine learning on high-dimensional data and firm fixed-effects as two complementary approaches to account for as many observables as possible in $Z'_i$. In the empirical portion of this paper, the machine learning predictive model is also esti-
mated on a sub-sample of firms that do not migrate, allowing the prediction to be interpreted as the expected performance of a firm if it stayed at home.

2.2.1 Testing Entrepreneurial Quality Estimates through ROC scores

Of key importance in a model that predicts a startup performance as in (14) is to develop a way to test whether the model predicts growth well or not. Different schools of thought offer different approaches to test predictive model fit, without any specific statistic being unequivocally better. Instead, the testing static to use depends on the distribution of the outcome, the relative costs of false positives vs false negatives, and other specific research goals.\textsuperscript{11} In this paper, I use one standard approach to testing binary outcome models—the out of sample ROC score.\textsuperscript{12} Formally, the ROC score is the area-under-the-curve of a model’s true positive rate compared to the false positive rate at all possible probability thresholds (this creates the ROC plot). Conceptually it represents an answer to the following problem in our setting: if two random startups, one which achieved growth and one which did not, are fed to the predictive model from (14), what is the probability that it will score the growth startup higher than the non-growth startup? A fully uninformative classifier will have an ROC score of 0.5, while a perfect classifier will have an ROC score of 1. I estimate the predictive performance of the model ($Fit$) as simply the share of the distribution between 0.5 and 1 that is covered by the ROC score. I interpret this number as the share of variation in outcomes accounted for by the predictive model.

$$Fit = (1 - ROC)/.5$$

(15)

2.3 Econometric Estimation

To estimate the impact of moving on the performance of the movers (treatment on the treated) using observational data, I propose two complementary econometric approaches

\textsuperscript{11}For example, a model predicting hospital infections would be overly cautious due to the high cost of the infection compared to the cost of preventive action.

\textsuperscript{12}ROC stands for Receiver-Operating-Characteristic, a name that is a remnant of the early application of this measure to radar signal processing during World War II.
that take advantage of different sources of variation. The first approach uses a cross-sectional machine learning estimator. Specifically, using the double LASSO approach of Belloni, Chernozhukov, and Hansen (2014), and comprehensive data on the characteristics of start-ups at founding (the comprehensiveness being evaluated through the *Fit* measure), I control for many observables at founding and estimate the performance of movers compared to very similar non-movers that stay in their region of birth. The second approach uses migrant fixed effects and takes advantage of differences in the timing of the move to estimate gains from moving for early migrants vs later-on migrants who have not moved yet. I review each one in turn.

### 2.3.1 Machine Learning Methods using High Dimensional Data for Index Models

The machine learning estimator can be best understood as an index model (Heckman, Ichimura, Smith, & Todd, 1998). Consider many firms, indexed by $i$, all born outside Silicon Valley. The firms are fully characterized by a high-dimensional (even infinite) number of observables $Z_i$. The firm’s performance $Y_i$ can be determined by two structural functions of these observables, $g_1$ for the performance in Silicon Valley and $g_0$ for the performance at home, and two additively separable error terms $U_{i1}$ and $U_{i0}$.

$$
Y_i = \begin{cases} 
Y_i(1) = g_1(Z_i) + U_{i1} & \text{if located in Silicon Valley} \\
Y_i(0) = g_0(Z_i) + U_{i0} & \text{if located outside Silicon Valley}
\end{cases}
$$

Our goal is to estimate

$$
\Delta = E\left[ \frac{Y_i(1) - Y_i(0)}{\theta} \right| S_i = 1] \tag{16}
$$

The econometric challenge is that we do not observe $Y_i(0)$ (nor $g_0$) for those who move, and therefore cannot estimate $\Delta$ directly. Traditional index models typically develop an estimate of $g_0$ through a low-dimensional non-parametric model such as local-linear and kernel regression (e.g. Ichimura, 1993; Heckman et al., 1998). The goal of index models is
to use some set of observables $X_i \in Z_i$ to estimate a function $\hat{g}_0(X_i)$ such that $\hat{g}_0(X_i) \approx g_0(Z_i)$.

$\hat{\Delta}$ is defined then by

$$\hat{\Delta} = E[Y_{i1} - \hat{g}_0(X_i)|X_i, S_i = 1] + B(X_i) + E[U_{i1} - U_{i0}|X_i, S_i = 1] \tag{17}$$

where $E[U_{i1} - U_{i0}|X_i, S_i = 1] = 0$ and $B(X_i)$ represent two sources of the bias.

Heckman et al. (1998) show how well-developed index models that include careful measurement and local information can make the bias $B(X_i)$ quite small. If the errors terms can be assumed to be mean-zero (i.e. $E[U_{i1} - U_{i0}|D_i, X_i] = 0$) then $\Delta$ is identified. Critically, this depends on how good the observables $X_i$ are. The recent boom in IT infrastructure generally called ‘Big Data’ has created a substantial number of measures that can be included in $X_i$, which computer scientists and econometricians have embraced with the hope that can enable us to estimate a better $\hat{g}_0$ and reduce this bias. The new methods developed (generally falling under the label ‘machine learning’) take advantage of these measures.\footnote{One particularly important that becomes more salient in machine learning is the risk of lack of common support. Though easily defendible in low-dimensional regressions, common support (also called covariate overlap) can often fail when high dimensional, an issue most recently highlighted by D’Amour, Ding, Feller, Lei, and Sekhon (2017). To achieve common support, a machine learning approach must rely on a variable regularization technique to guarantee overlap. I use the ‘double LASSO’ approach developed by Belloni et al. (2014) for variable regularization in this paper.}

At the center of the validity of this model is the assumption that the predicted value of performance from observables $\hat{g}_0(X_i)$ is close to the true value of underlying firm expected performance at home. This assumption is itself testable.

An initial sense of whether $\hat{g}_0(X_i)$ is a good measure can be estimated from the Fit measure in equation (15). However, this is not a measure of how far the model is from ideal performance, but instead only a lower-bound. The reason is that gaps in the model’s performance can be driven by either incomplete mismeasurement or random growth (stemming from $\epsilon$ in the structural model, and from the $U_i$ in this section), but we are only concerned here with the former (mismeasurement) and not the latter. A model with a Fit
value of I would require both that ex-ante potential is perfectly accounted for, and that there is no ex-post random noise influencing the process of firm growth. Given the long literature in economics arguing for the relevance of random growth in firm performance (Gibrat, 1931; see Gabaix, 2017, for a recent summary), it is only reasonable to assume it plays some role in the dynamics of our firms.

A better approach to evaluate model performance is to look at the stability of the coefficients as follow-on information is added. Oster (2017) offers a novel approach that considers whether adding additional observables \( x'_i \) to \( X_i \) increases the information captured in the regression (i.e. increases the \( R^2 \)), and changes the coefficient of interest. Under some parameters (defined by Oster), and under the assumption that unobservables are as correlated with the outcome as the observables added, it then allows us to ask how big would the additional unobservables have to be (relative to existing observables) to make the effect go away. In this paper, I compare the performance of a regression with region-year of birth pair fixed effects and machine learning controls, to one with machine learning controls (and without region-year pair fixed effects). This is a useful comparison because the location of birth of a startup can be reasonably assumed to influence migration choice and to be correlated with unobservable performance measures. This relationship between region and selection is also at the core of the propositions in the theoretical model, and their existence are also tested in the first part of the empirics of the paper. In short, if we assume there that is risk of upward bias for estimates of the migration treatment effect, but the coefficient does not statistically change between the regression with only machine learning, and the regression with machine learning plus region-year pair fixed effects, then the conclusion is that the bulk of the potential endogeneity was already accounted for in the regression with only machine learning.

2.3.2 Panel Data Regressions

The second econometric approach takes advantage of the panel-data nature of the data. Considering the fact that not all migrants move at the same time, I set up a structure that includes age and firm fixed-effects and compares the benefits of performance for early
movers compared to other movers that have not yet moved. That is, for each migrant firm $i$, of age $t$, I run OLS regressions of the form

$$Y_{i,t} = \lambda_i + \gamma_t + \beta' \tau M_{i,\tau} + \epsilon_{i,t}$$

where $M_{i,\tau}$ is a vector of individual indicators for each value of $\tau$, defined as the difference between the age at migration and $t$. $\lambda_i$ is a firm fixed-effect, $\gamma_t$ is an age fixed-effect, $Y_{i,t}$ is an outcome measure, and $\epsilon_{i,t}$ is white noise.\textsuperscript{14} The coefficients of interest are the vector $\beta_\tau$, which represent the differences in the performance of migrants after migration (or before if $\tau < 0$), once fixed firm differences ($\lambda_i$) and mean age differences ($\gamma_t$) are accounted for.

\section{Data}

The dataset is built using business registration records from 26 U.S. states, representing 70\% of U.S. GDP, from 1988 to 2014.\textsuperscript{15} Business registration records are public records created endogenously when a firm is registered as a corporation, partnership, or limited liability company, with the Secretary of State (or Secretary of the Commonwealth) of any U.S. state (or commonwealth). Several other studies have used the same kind of business registration records used in this paper to study the founding and location of startups.\textsuperscript{16}

To focus this study on startups with high growth intention, I select on the subset of startups that register under Delaware jurisdiction. In the process of choosing a jurisdiction for their company, founders benefit from registering the firm 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.

\textsuperscript{14}This approach has commonly been used in urban economics to study the effect of location on personal productivity (e.g. Glaeser & Maré, 2001).

\textsuperscript{15}A map of all states in the data is included in Appendix Figure A1.

\textsuperscript{16}See, Guzman and Stern (2016) and Guzman and Stern (2017) for detailed appendices on this dataset.

Electronic copy available at: https://ssrn.com/abstract=3175328
including its highly reputed Court of the Chancery, and (3) that the decisions and legal framework of Delaware are generally regarded as pro-business. These benefits are more useful for startups that will be large or for startups interacting with investors, including venture capitalists. However, being in the Delaware jurisdiction also holds extra costs and requires two registrations (one in Delaware and one in the state of operation), imposing fees that a business that expects to be small is likely to deem unnecessary. This creates a natural separating equilibrium, with mostly growth-oriented companies choosing to register in Delaware. Accordingly, while Delaware companies represent only about 4% of all firms, they account for 50% of all publicly listed firms, and over 60% of all VC financing (see Catalini, Guzman, & Stern, 2017). Delaware firms are also 50 times more likely to achieve an IPO or be acquired than non-Delaware firms (Guzman & Stern, 2017).

The resulting dataset contains the registration of 488,960 new firms. I enhance business registration data by using a name-matching algorithm to merge business registrations with five other datasets: (i) three types of intellectual property filings from the U.S. Patent and Trademark Office (trademark applications, patent applications, and patent assignments), (ii) all new IPOs in the U.S. from the SDC New Issues database, (iii) all U.S. M&A activity reported in the SDC Mergers and Acquisitions database, (iv) venture capital activity from ThompsonReuters VentureXpert, and (v) annual estimates of sales from Infogroup USA annual files.

I develop a series of measures documenting a startup’s characteristics at founding and its subsequent performance.

**Firm Observables at Founding.** From business registration I create 2 binary measures indicating whether a firm is a corporation (rather than an LLC or partnership), and whether it is an LLC. I create 12 measures of firm name length, including a continuous measure of the number of words in the firm name, the square of the number of words, and 10 binary indicators for whether the name is exactly 1 through 10 words long. I create industry mea-

---

17In fact, venture capitalists most often require that portfolio companies are in Delaware because their contracts are specifically written for Delaware corporate law.

asures by using the approach of Guzman and Stern (2015), which uses a large sample of firms with NAICS codes to create a name-based algorithm that allows categorizing firms into different economic clusters from the U.S. Cluster Mapping Project (Delgado, Porter, & Stern, 2014). I create 13 binary measures following this approach, one for each of the following groups: Agriculture and Food, Automotive, Chemicals, Clothing, Consumer Apparel, Distribution and Shipping, Energy, High Technology, Local Industries, Mining, Paper and Plastic, Publishing, and Services. I also create 5 more measures for names associated with specific high tech industries that have accounted for a meaningful share of high growth entrepreneurship in this time period: IT, Biotechnology, E-Commerce, Medical Devices, and Semiconductors. Finally, I create six measures from intellectual property filings. Three indicate whether the firm applies for a patent in their first year, has a patent assigned (from a prior inventor) in their first year, or files for a trademark in their first year. The other three indicate whether the firm applies for more than one patent, is assigned more than one patent, or applies for more than one trademark in their first year.

This leads to a total of 38 measures observable at the time of firm founding, which can be combined in 703 ways in two-way interactions, for a total of 741 observable measures at founding.

*Measures of Regional Entrepreneurial Quality.* To measure the quality of the entrepreneurial ecosystem, I use the public datasets provided through the Startup Cartography Project (see Andrews et al (2017)), where we use business registrations and a predictive approach to characterize the quality of all companies registered in each ZIP Code and year for all states and time-periods in my sample.

*Measures of Firm Performance.* I develop six outcome measures based on the firms observed performance six years after founding. *IPO* and *Acquisition* are two binary variables equal to 1 if the firm has an IPO or an acquisition within six years. The key outcome of interest, *Equity Growth*, is simply the union of these two variables. It is equal to 1 if a firm achieves an IPO or an acquisition and zero otherwise. Though rare, equity growth represents a highly desirable outcome for entrepreneurs with high growth intention (the sale of
their company), and closely matches the anecdotal incentives sought by many high growth founders. I also include four more outcomes. Follow-on Patent is equal to 1 if the firm files or acquires (is assigned) one or more patents, excluding the first year window used for at-founding observables. Follow-on Trademark is equal to 1 if the firm files one or more trademarks, excluding the first year window. Venture Capital, is equal to 1 if the firm receives venture capital. And High Sales is equal to 1 if the firm achieves $1 million or more in sales (as reported by Reference USA).

Measuring Migration. To track migration, I use the fact that companies need to register not only in the state in which they are founded, but also in every state in which they operate—i.e. rent real-estate, hire people, or set up a local bank account. These registrations include at least the name of the firm, the date of registration, the address of the firm’s local office within the state, and the address of the principal office (headquarters). Because Delaware corporate law specifically requires companies to name themselves sufficiently different from one another, and because (except for rare cases) companies must register with their true name in each state, the matching across registries is very simple and allows high confidence that two registrations of a Delaware jurisdiction firm under the same name, in two different states, represent the same firm.

I operationalize migration through three conditions: (1) The first state in which the firm registers is assumed to be the birth state, and its registration date the date of founding; (2) if a firm then registers in a second state with its principal office in this new state, this is considered a migration as long as the firm lived in the birth state for at least 3 months; (3) the date of registration in the destination state is the migration date. This limits the analysis to migration of registered firms across states, but with an ability to see the specific destination address (and hence also the destination MSA). Focusing on cross-state migrations allows me to generally abstract away from migrations within the same economic region.

I develop two measures documenting migration. Migrant (Anywhere) is equal to 1 if the firm moves to any destination within the first two years after founding, and Migrant to Silicon Valley is equal to 1 if a firm moves to Silicon Valley in the first two years after
founding. Table 1 shows summary statistics for some of the key variables, and for all migrant firms who migrate in the first two years.

### 3.1 The Incidence of Migration

We are finally able to move to the empirical results. Table 2 shows the average migration rate from ages 0 to 15 years old, for all firms founded on or before 2001 in the data (to avoid censoring). Two results stand out. First, migration is not rare: the total migration rate is 10%. This number is suggestive of an important economic phenomena. Second, migration is more common in the early stages of a firm. 6.7% occurs within the first 5 years, and 4.2% within the first two years. For the purposes of this paper, an ‘entrepreneurial migration’ is defined (rather arbitrarily) as migration done by a firm in the first two years after founding. Figure A2 reports in more detail migration patterns by quarter of age and different quality thresholds.

Figure 1 reports the incidence of entrepreneurial migration across different outcomes. Panel A is all startups. Relative to a baseline incidence of migration of 4.2%, 10.1% of the firms that achieve equity growth are migrants, 6.5% of the firms that raise VC financing are migrants, and 7.7% of the firms that patent are migrants. Panel B focuses only on firms founded outside California and their migrations to Silicon Valley. Only 0.25% of the firms move to Silicon Valley. Yet, these firms represents 2.9% of all equity growth outcomes, 2.5% of all VC financing, and 1.8% of all patenting.

The large differences between the incidence of migration and growth outcomes open the possibility that there are significant performance benefits of migration, particularly to Silicon Valley, but this assessment requires an understanding of how firms would perform if they stay at home.
4 Empirical Results

4.1 Estimating Entrepreneurial Quality

Building on the model of Section 2, I estimate entrepreneurial quality through a random forest (Breiman, 2001) (a type of predictive algorithm) and predict the likelihood of Equity Growth from firm observables at founding. Given the large number of observable measures available at firm founding (741 in total) it is necessary to do variable regularization to guarantee that the estimator has common support (as explained in Section 2). I use the double-LASSO approach of Belloni et al. (2014) to do so. This approach also allows the selected controls to be included directly in an OLS regression, providing a second methods (besides using predicted quality) to control for firm differences. LASSO Controls is defined as the set of 91 variables selected by this procedure.

The random forest model uses Equity Growth as the outcome variable and LASSO Controls as the independent variables on a 50% random sub-sample of non-migrants. Because the sample is limited to non-migrants, the predicted estimates can be interpreted as the probability of success when the firm stays at home.

Table A1 reports two simple logit models to highlight some interesting relationships. Column 2 reports striking relationships of the main variables on the odds of achieving equity growth outcomes. Firms with a short name are 42% more likely to grow, and corporations are 7.6 times more likely to grow. Firms with a patent are almost 5 times more likely to grow, and firms with a trademark 2.8 times. Firms whose name is associated with the traded economy are 18% more likely to grow. In terms of targeted innovative sectors, firms with a name associated with biotechnology are 130% more likely to grow, firms in e-commerce 31%, firms in IT 53%, and firms in semiconductors 56% more likely to grow. Interestingly, firms in medical devices are 19% less likely to grow. Bringing some of these estimates together, a biotechnology corporation with a patent and a trademark is 426 times more likely to achieve equity growth than a Delaware incorporated LLC without these indicators.
The random forest model has an ROC score of 0.86 in the training sample, and an ROC score of 0.85 in the hold-out (non training) sample. The out of sample ROC of 0.85 implies a *Fit* of 0.7, that is, 70% of the variation in outcomes is accounted for by the model. Figure 2A reports the ROC curve for all firms in the hold-out sample.

Figure 2B studies the predictive performance of this model through an out of sample 10-fold cross validation procedure. It plots the distribution of firms that achieve growth out of sample across 5% groups of the predicted quality distribution. The top 5% of the out of sample quality distribution accounts for 43% of all growth firms, and the top 10% for 59%.

Together, these results suggest a high level of out of sample predictive performance for the machine learning model. All follow-on analyses in this paper are performed in the hold-out sample to avoid any issues with overfitting of the predictions.

### 4.2 Selection into Migration

We now have enough structure to study the selection into migration and its relationship to local entrepreneurship levels at the source and destination regions. Figure 3 presents a series of scatterplots of estimated quality and their relationship to migration rates. Panel A is a binned scatterplot of firm quality in 5 percent bins and the average rate of migration for each bin. The relationship between migration rates and quality is positive, suggesting that better firms are more likely to migrate. Panel B is the binned scatterplot of birth state quality and the average rate of migration. The fit is negative, suggesting firms are more likely to leave low quality regions. Finally, Panel C plots, within a year, the share of all movers that move to a destination state and the quality of that state in that year. The fit is once again positive. Conditional on migration, firms are more likely to move to higher quality states.

These relationships are sharpened in Table 3, which presents a linear probability model with *Migrant (Anywhere)* as the dependent variable, controlling for year, state, and state-year pair fixed-effects. My preferred specification is Column 3. The results suggest that
increasing firm quality by one log-point (about two thirds of a standard deviation) leads to an increase in the likelihood of migration of 1.2%. Given a mean migration rate of 3.3% (in the full sample), this is a 36% change from the baseline.

These results suggest that there is a migration pattern of migrants seeking the agglomeration benefits of high quality ecosystems, with high quality firms moving from regions of low agglomeration to regions of high agglomeration.

4.3 Migration and Startup Performance: Machine Learning Estimates

I now proceed to the centerpiece of the analysis, estimating the impact of migration on startup performance.

Figure 4A reports graphically the change in odds of Equity Growth when comparing migrants and non-migrants in several statistical models. The odds are estimated by running a linear probability model of Equity Growth with migration as the independent variable, and then dividing the coefficient by the expected performance had the firm not moved (its entrepreneurial quality). The regression tables are available in Appendix Table A2. Standard errors are clustered by founding state.

The top portion of Figure 4A reports the increase in odds from migration to Silicon Valley. The first estimate is the naïve estimate: the coefficient of a regression without any controls, divided by the sample mean of the outcome. The naïve estimate is 8.3. The three models below the naïve model perform a series of improvements that reflect common changes that we would expect to see in a ‘classic’ selection on observables approach—adding state fixed-effects, or state-year pair fixed effects, and controlling for firms that have intellectual property or venture capital before moving. The estimated effect is very similar to the naïve estimate. The last three rows move away from the classic selection on observables and instead use high dimensional methods as presented in Section 2. The differences are significant and the odds of equity growth drops by half (though it is still meaningful). The preferred estimate is a model controlling for entrepreneurial quality and state-year
fixed effects (in green). The result suggests migration to Silicon Valley increases the odds of *Equity Growth* by 4.1X.

The two bottom portions of Figure 4A report differences in performance for migrations to other destinations (excluding Silicon Valley) and for all migrants. The effects are lower but still positive and economically meaningful. Moving to other destinations suggests an increase of 2.3X the odds of success. In both cases, the naïve estimate substantially overestimates the benefit of migration.

Figures 4B and 4C report a series of heterogeneity analyses. Figure 4B reports the estimates of the role of migration across a series of sub-samples—only firms with a patent, only firms with a trademark, only corporations, only firms with venture capital investment, and only firms with a name associated with high tech. There is variation in both the coefficients and the quality of the average migrant within each sample, and the samples are in some cases too small to get a 95 percent confidence interval that excludes zero. However, the point estimate of the change in odds is always positive, with values between 2 and 4, and never statistically different from the main effect.

Figure 4C shows the effect of migration on the performance of startups across the quality distribution using a kernel regression with an Epanechnikov kernel and a bandwidth of 0.05. The results suggest increasing returns of migration for firm quality, with the bottom 40% having no perceivable benefit from migration and the benefit increasing with the distribution.

Together, these results suggest a persistent positive benefit of migration on the performance of startups, which is particularly high for movers to Silicon Valley, and highlights at least a 2X overestimate of the effect of migration when using common controls, compared to the machine learning method used here.

*Other Outcomes.* I expand the analysis to other outcomes in Figure 5, by studying the effect of migration to Silicon Valley on the odds of patenting, trademarks (a proxy for commercialization), venture capital financing, and sales. The econometric models are equivalent to those in Figure 4A, though a different random forest is run for each outcome.
variable to estimate the expected performance for migrants if they had stayed at home.
The ROC scores of these random forest models range from 0.70 (sales) to 0.95 (patenting).
The increase in odds from migration implied by the naïve model is much higher than the
preferred model in all cases, in some cases being three times as much. The benefits of mi-
gration are all positive but do vary by outcome, and they seem to be higher for increases
in venture capital financing, and sales, than for patenting and trademarks.

The breadth of these results suggest that this benefit of location is not simply due to
access to financial agents to achieve equity outcomes, but also reflects important produc-
tivity benefits due to locational agglomeration at the destination.

Tests of Unobservables. A natural concern is that some other aspect unaccounted for
in the machine learning approach could be driving the estimate. This section looks at two
contems: whether the machine learning approach does a sufficiently good job of making
potential unobservables at founding small, and whether there are any follow-on productiv-
ity shocks to firms after founding that correlate with both migration and improved perfor-
ance, causing a misattribution of the effect.

I begin with the first concern, whether the machine learning approach accounts for
enough variation at founding. That the out-of-sample ROC score of the random forest
model is 0.85, suggesting it accounts for 70% of the variation in outcomes, already gives us
a sense that it performs reasonably well. Still, it is unclear what the effect of the remain-
ing 30% might be in biasing the estimates. In particular, economic literature has long em-
phasized the role of ‘random growth’ in firm performance (Gibrat, 1931) so that a portion
of the growth process is necessarily not accounted for in the predictive model of founding
characteristics. Instead, the main concern is that besides the random growth component,
there also exist meaningful unobservables that sistematically bias the estimated outcomes.
A common approach to test for the role of unobservables is to look at coefficient stability
across specifications, where a stable coefficient increases confidence that the estimate is
correct. As explained in Section 2, Oster (2017) provides a formalization of this approach:
she develops a method whereby the increase in $R^2$ and change in coefficient between a first
and a second regression can be used to assess the necessary size of follow-on unobservables
to make the effect go away, under the assumption that the unobservables are as related to performance as the added observables in the second regression. In Table 4, I report estimates using this approach comparing a model without state-year fixed effects to a model with state-year fixed effects, and report two scenarios: the size of unobservables needed for the coefficient to become zero, and for the 95% confidence interval to include zero (even if the coefficient is positive).

Unobservables need to be 10 times larger than observables for the effect of all migrants to include zero in the 95% confidence interval, and 25 times larger than observables for migrants to Silicon Valley. This is much higher than the 30% of variation available according to the ROC score. The ample difference between these two suggests that there is not enough unobservable variation left to make the results zero or even close to it.

Now I move to the second concern, the possibility of productivity shocks after founding. If firms are matched well at founding, but diverge after founding due to unobservable shocks that correlate with both migration and performance, then the likelihood of receiving such a shock should increase with firm age. We would then expect a positive bias to ‘creep in’ for firms that move at an older age—i.e., the benefit of migration would be correctly estimated for firms who move close to birth, but estimates for later movers would include both the benefit of migration and some upward bias from unobserved shocks. In Figure 6 I report separate coefficients for migrants moving at different semester ages, from semesters 1 through 4. The effect does not have a positive trend across semesters, providing evidence consistent with productivity shocks not being a main concern.

### 4.4 Migration and Startup Performance: Fixed-Effects Regressions

The second analysis on the effect of migration on firm performance does not rely on machine learning, but instead on using the timing of migration in a panel format. Building on the setup of section 2.3.2, I study the impact of migration on the performance of startups under a panel structure, and use firm and age fixed-effects to consider differences in
the performance of migrant startups of the same age, but who move at different times. To run this model, I re-design my dataset to have a panel format by measuring individual outcomes for each period $t$ up to six years of age. I also perform two other changes. (1) Because the estimation of the age-specific treatment effect ($\beta_\tau$) with firm fixed-effects requires that some migrants move in each period (each $\tau$), I include migrants who move up to six years after founding in the data, rather than only within the first two years. This allows comparability with the estimates in the prior section. (2) To obtain a more granular view of the evolution of startups before and after migration, I also change observations to semesters (rather than years), allowing 11 pre and 11 post periods to be reported.

Figure 7 reports the analysis by plotting the semester-level coefficients for migration for four outcomes: the number of patent applications, trademark applications, venture capital dollars raised (in millions), and equity growth, from $\tau=-11$ to $\tau=+11$. The figure also reports a ‘baseline’ measure for each outcome, which is the mean value of the outcome variable for a matched set of firms of the same quality, born in the same location and year, that do not move.

Together, the graphs mostly show no differences in firms before moving. No coefficient is statistically different from zero before the firm moves and most have tight confidence intervals. The differences in the effects are also slowly increasing (rather than a quick change at migration), reflecting the slow and cumulative benefits of Silicon Valley on the migrants. I review each sub-graph of Figure 7 in turn.

The number of patent applications does not show any differences before migration. There are meaningful differences after migration, that build through time. The large confidence intervals suggest the data is skewed, a common phenomenon in patent-based outcomes. 6 years after migration, migrants register about 1.7 more patent applications than non-migrants. Given a baseline of 0.3 patent applications for matched non-migrants, this suggests an increase of 4.7X in the patenting activity of migrants to Silicon Valley. The lack of pre-trends in patent applications also suggests that concerns over unobserved productivity shocks are not of first order in this setting.

The amount of venture capital raised also does not show any differences before mi-
migration. Because of the critical role venture capitalists are reported to play in many Silicon Valley-based startups, one key question is whether the venture capitalists themselves (rather than some generalized agglomeration benefits) are pushing the observed productivity differences by both moving firms and giving them more resources. This evidence suggests this is not the case; instead, access to financing is a localized benefit that accumulates through time. While the amount of venture capital dollars raised is no different right at migration, startups that move raise on average $4.6 millions more after 6 years than non-movers. Given a baseline of $1.03 million, this suggests a 3.5X increase in the amount of financing.

The graph for trademark applications is less clean in helping understand the dynamics. The coefficient at t=0 is not statistically different from zero, but seems to have a point estimate that is positive compared to the overall magnitude of the effect. The effects are rather small when compared to patenting (6 years after migration, migrants apply for 0.2 more trademarks than non-migrants). Though the coefficients in this graph are statistically consistent with a zero before migration, the data overall seems more noisy and it is hard to use this result either in favor of or against the hypothesis of locational effects on migrant performance.

Finally, the lower left graph estimates the impact of migration on the key outcome of interest in this paper—changes in the likelihood of achieving an equity growth event. The differences are once again quite flat before migration, but quite meaningful after migration. 6 years after migration, firms are 7.1% more likely to achieve the growth outcome than similar non-migrants. Given a baseline of 1.3%, this suggests a benefit of 4.5X in the likelihood of migration. This estimate is very close to the machine learning estimate of the prior section.

An important point to note on this graph is the appreciable increase in the likelihood of Equity Growth exactly in the semester of migration. Rather than differences accruing only some time after migration, firms that move are 2.4% more likely to be acquired at t=0 than non-movers. This could suggest some movers move with the option of already collaborating closely with firms at the destination, and perhaps with the short term goal
(and prospect) of vertically integrating with them. Understanding these dynamics better would be a question for future work.

4.5 Heterogeneity in Space

I then consider the micro-location choices of migrants. Since knowledge spillovers deteriorate quickly with distance,\textsuperscript{19} location within Silicon Valley matter for performance inasmuch as knowledge spillovers are the mechanism driving performance. Figure 9 shows the location of all Delaware startups born in Silicon Valley (left panel) and the destination location of the startups that moved to Silicon Valley (right panel). Table 5 studies how locating to a ZIP Code with a higher quality and quantity of local Delaware startups relates to the performance of firms, conditional on the characteristics of the migrant startup. As expected, the local entrepreneurship of a destination ZIP Code is positively related to the performance of the migrant firm and statistically significant, suggesting knowledge spillovers at least partially account for the effect of location on performance.

4.6 Heterogeneity in Time

Finally, I consider heterogeneity in the business cycle when a firm moves to Silicon Valley. To do so, in Figure 8, I report the effect of migrating in specific years during two eras that were quite distinct in Silicon Valley, the dot-com boom and the dot-com bust.

For all firms $i$ born in region $r$, I run OLS regressions of the form

$$Y_{i,r} = \alpha + \beta' M_{i,r} + \theta_i + \gamma_r + \epsilon_{i,r}$$

where $\theta_i$ represents the startup’s entrepreneurial quality, $Y_{i,r}$ is an outcome measure, $\lambda_r$ are region fixed-effects, and $\epsilon_{i,r}$ is random noise. $M_{i,r}$ is a vector of 11 elements each of

\textsuperscript{19}Kerr and Kominers (2014) use citation patterns in Silicon Valley to show innovations travel only a short distance within the region; (Arzaghi & Henderson, 2008) show that the knowledge-related value of location for advertising firms in New York City can drop quickly after a few blocks.
which takes a value of 1 if a firm moved to Silicon Valley in each of the 11 years between 1996 and 2006, and $\beta$ are individual coefficients of interest—the effect of moving in each year.

Figure 8 reports these coefficients (Panels B through F) for a series of outcomes. An important requirement to be able to draw comparisons is that the quality of migrants is consistent across years. Panel A shows this is roughly the case. Panel B studies the likelihood of equity growth for migrants at different years. The patterns before and after the dot-com bust are striking. While there was a large benefit of moving to Silicon Valley during the boom years, this benefit becomes negligible in the bust years. Panel C studies VC financing. The pattern here is different. The benefit raises up to 2001 (the year the bubble burst) and then recedes gradually. Both patterns match well anecdotal accounts of the time period and general intuition: while there was a sharp drop in IPOs and acquisitions after the bubble burst (and the markets collapsed), some VC financing continued after 2001 as VCs more slowly unloaded the capital already fundraised.

Panels D, E, and F, look at outcomes less related to the boom and bust process of capital markets: patents, trademarks, and sales. Interestingly, these outcomes show little movement across time, suggesting the effects of Silicon Valley on the ‘real economy’, such as knowledge spillovers or access to markets are less cyclical. This is also intuitive: though the IPO market crashed in 2001, the scientists and engineers in the region continued to work and (presumably) produce after 2001.

Together, these results provide further support for a role of agglomeration on the performance of migrants, but also highlight how these benefits are not homogeneous. Instead, these depend on the outcome in question and the condition of the local ecosystem at the time of migration.
5 Conclusion

This paper has focused on two competing theories of how location influences a startup’s competitive advantage and growth. Agglomeration emphasizes the importance of transportation costs (distance) in accessing resources, while embeddedness emphasizes the role of personal social relationships in transacting with tacit knowledge, rather than through arms-length interactions. These two theories lead to different predictions on the impact of startup migration on the performance of startups, and in turn have diverging recommendations for the design of regional entrepreneurship policy. I offered a simple model that includes agglomeration and embeddedness simultaneously, and provides some insight on how they relate to each other. Using all Delaware registrants across 26 US states, from 1988 to 2014, I verified that the key predictions of the model are true in the data, and used machine learning to estimate the performance improvements gained from moving for startups that move to Silicon Valley. The evidence suggests agglomeration is more important than embeddedness for these firms, and that it is partially due to access to localized knowledge at the destination. These benefits of location impact both the ‘financial economy’ (financing, and equity outcomes for startups), and the ‘real economy’ (sales, patenting, and product introductions).

These results are only an initial understanding of how migration shapes startup performance, and have considered only one specific destination. Future work should focus on understanding how this generalizes across different levels of regional fit, what migration says about the allocation of startup supply in general, and how much the results in this paper on Delaware startups are also representative of other types of migration (such as the location choices of would-be founders before starting a company).

Fundamentally, the role of location on the performance of firms depends not only on where they are born, but also whether and where they move to. Urban economists since Roback (1982) have often assumed people move, but companies do not. This paper shows there is migration in startups and that this migration has consequences for their performance; a fact that is likely important to understand the economy and the role of new
firms in it.
References


<table>
<thead>
<tr>
<th>Outcomes</th>
<th>All Firms</th>
<th>Migrants (2 years)</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equity Growth</td>
<td></td>
<td></td>
<td>0.008</td>
<td>(0.089)</td>
<td>0.049</td>
<td>(0.216)</td>
</tr>
<tr>
<td>(Source: SDC Platinum)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Files Trademark</td>
<td></td>
<td></td>
<td>0.045</td>
<td>(0.206)</td>
<td>0.087</td>
<td>(0.283)</td>
</tr>
<tr>
<td>(Source: USPTO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Files Patent</td>
<td></td>
<td></td>
<td>0.037</td>
<td>(0.188)</td>
<td>0.069</td>
<td>(0.254)</td>
</tr>
<tr>
<td>(Source: USPTO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gets VC</td>
<td></td>
<td></td>
<td>0.016</td>
<td>(0.126)</td>
<td>0.023</td>
<td>(0.148)</td>
</tr>
<tr>
<td>(Source: Thompson Reuters VentureXpert)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Sales</td>
<td></td>
<td></td>
<td>0.039</td>
<td>(0.194)</td>
<td>0.062</td>
<td>(0.240)</td>
</tr>
<tr>
<td>(Source: USPTO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Migration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migrant (Anywhere)</td>
<td></td>
<td></td>
<td>0.033</td>
<td>(0.179)</td>
<td>1.000</td>
<td>(0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migrant to Silicon Valley</td>
<td></td>
<td></td>
<td>0.002</td>
<td>(0.049)</td>
<td>0.988</td>
<td>(0.110)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Migration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migrant in 5 Years</td>
<td></td>
<td></td>
<td>0.052</td>
<td>(0.223)</td>
<td>1.000</td>
<td>(0.000)</td>
</tr>
<tr>
<td>(censored)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migrant Ever</td>
<td></td>
<td></td>
<td>0.073</td>
<td>(0.260)</td>
<td>1.000</td>
<td>(0.000)</td>
</tr>
<tr>
<td>(censored)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business Registration Observables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corporation</td>
<td></td>
<td></td>
<td>0.426</td>
<td>(0.494)</td>
<td>0.625</td>
<td>(0.484)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short Name</td>
<td></td>
<td></td>
<td>0.579</td>
<td>(0.786)</td>
<td>0.528</td>
<td>(0.705)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delaware Jurisdiction</td>
<td></td>
<td></td>
<td>1.000</td>
<td>(0.000)</td>
<td>1.000</td>
<td>(0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intellectual Property</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patent</td>
<td></td>
<td></td>
<td>0.025</td>
<td>(0.156)</td>
<td>0.025</td>
<td>(0.156)</td>
</tr>
<tr>
<td>(Source: SDC Platinum)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trademark</td>
<td></td>
<td></td>
<td>0.012</td>
<td>(0.108)</td>
<td>0.011</td>
<td>(0.103)</td>
</tr>
<tr>
<td>(Source: USPTO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>488960</td>
<td>16243</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

91 observables from business registration records and IP filings used on a double LASSO procedure to control for observables not included. Outcomes must be achieved by the firm within six years of founding. Migration is equal to 1 if the firm changes headquarters as evidenced in changes in its business registration across states. For migrant companies, only success in the destination region is counted (e.g., filing a patent where the assignee is in the destination region). Equity Growth is a binary measure equal to 1 if a firm achieves IPO or acquisition. IP is a binary measure indicating whether the firm files a trademark or a patent. VC outcomes are measured from Thompson Reuters VentureXpert and High Sales is a binary measure equal to 1 if the firm is reported as having over $1 Million USD in sales by year 6 in the Infogroup USA database.
TABLE 2
MIGRATION RATES FOR ALL FIRMS BORN BEFORE 2001*

<table>
<thead>
<tr>
<th>Does not migrate</th>
<th>89.9%</th>
</tr>
</thead>
</table>

*Entrepreneurial Migration*
- Migrates before 2 years of age: 4.2%
- Migrates before 5 years of age: 6.7%

*All Migration*
- Migrates from age 6 to 15: 10.1%

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

---

TABLE 3
QUALITY OF MIGRANTS AND NON-MIGRANTS LINEAR PROBABILITY MODELS

<table>
<thead>
<tr>
<th>DEPENDENT VARIABLE: MIGRANT (ANYWHERE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln(Firm Entrep. Quality)</td>
<td>0.0173**</td>
<td>0.0122**</td>
<td>0.0114**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00371)</td>
<td>(0.00253)</td>
<td>(0.00253)</td>
<td></td>
</tr>
<tr>
<td>Ln(State Entrep. Quality)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0356**</td>
<td>-0.0172</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0127)</td>
<td>(0.0116)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State F.E.</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Year F.E.</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>State Year Pair F.E.</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>263083</td>
<td>227758</td>
<td>227758</td>
<td>263083</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.006</td>
<td>0.014</td>
<td>0.069</td>
<td>0.084</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level. Dependent variable is a binary measure equal to 1 if the firm migrates in the first two years of life. Firm Entrepreneurial Quality is the predicted value of performance from a machine learning model (random forest) in a trained to predict Equity Growth from at-founding characteristics. The sample used for training is excluded from this analysis. State Entrepreneurial Quality is the average quality of local firms born in that state and year, including both Delaware and non-Delaware firms. *p < .1 **p < .05

Electronic copy available at: https://ssrn.com/abstract=3175328
### TABLE 4
OSTER TEST OF SIZE OF UNOBSERVABLES FOR EFFECT TO BE ZERO

<table>
<thead>
<tr>
<th></th>
<th>Size of Unobservables for effect to be zero</th>
<th>Size of Unobservables for effect to be not-significant at p = .05</th>
</tr>
</thead>
<tbody>
<tr>
<td>All migrants (Comparing Panel A (3) with Panel A (5))</td>
<td>18:1</td>
<td>10:1</td>
</tr>
<tr>
<td>Migrants to Silicon Valley (Comparing Panel B (3) with Panel B (5))</td>
<td>36:1</td>
<td>25:1</td>
</tr>
</tbody>
</table>

*Oster Test* refers to the methodology in Oster (2017). Under the assumptions that (a) the unobservables are not more correlated than the observables with the bias and (b) differences in the R² comparing experimental vs observational studies in general provide a good guidance for the upper bound of variation accounted for (Oster suggests setting this value R_{max} at 1.3 × R²) then it is possible to provide the expected size of unobservables to make the effect be of certain value. Estimates of size of unobservables to make the effect not significant also assume the same standard errors. The regressions and R² are available in the appendix.

### TABLE 5
SILICON VALLEY MICRO-LOCATION CHOICES AMONG MIGRANTS AND PERFORMANCE
LINEAR PROBABILITY MODELS
DEPENDENT VARIABLE: EQUITY GROWTH

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Destination ZIP Code</td>
<td>0.0151</td>
<td>0.0416*</td>
<td>0.0310*</td>
</tr>
<tr>
<td>Local Entrepreneurship</td>
<td>(0.89)</td>
<td>(2.47)</td>
<td>(2.07)</td>
</tr>
<tr>
<td>(Quality X Quantity)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln(Birth State Entrep. Quality)</td>
<td>0.00876</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASSO Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State F.E.</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Moved Year F.E</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>725</td>
<td>664</td>
<td>725</td>
</tr>
</tbody>
</table>

Dataset is only migrants to Silicon Valley. Robust standard errors in parenthesis. * p < .1  ** p < .05

Electronic copy available at: https://ssrn.com/abstract=3175328
FIGURE 1

A. All Regions
Migration vs Non-Migration

<table>
<thead>
<tr>
<th>Share of Movers</th>
<th>All Firms</th>
<th>Equity Growth (IPO or Acquisition)</th>
<th>VC Funded</th>
<th>Patenting</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>4%</td>
<td>5%</td>
<td>6%</td>
<td>7%</td>
</tr>
</tbody>
</table>

B. Silicon Valley
Migration vs Non-Migration

<table>
<thead>
<tr>
<th>Share of Movers</th>
<th>All Firms</th>
<th>Equity Growth (IPO or Acquisition)</th>
<th>VC Funded</th>
<th>Patenting</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>3%</td>
<td>5%</td>
<td>6%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Notes: Reports the share of moves represented in each of the outcome measures as well as migrants in general. Implied odds is simply the column divided by the value of All Firms and represents the higher likelihood of a migrant having an outcome than expected at random.
FIGURE 2

A. ROC Curve for predictive model in out of sample non-migrant firms.

![ROC Curve](https://ssrn.com/abstract=3175328)

**Notes:** Figure A represents the receiver-operating-characteristic (ROC) score for the random forest predictive of entrepreneurial quality when estimated in a 50% hold-out sample of non-migrants. The dotted line represents the performance of an alternative algorithm that assigns quality at random. The area under the curve of the solid line compared to the dotted-line equals 70% of the total area above the dotted line. Figure B is an out of sample 10-fold algorithm that compares the ordering of firms by predicted quality to their actual realized outcomes, across bins that are 5-percent wide. The bars indicate the mean value, while the gray lines indicate the min and max.

B. Ten Fold Cross Validation

![Cross Validation](https://ssrn.com/abstract=3175328)
Determinants of Migration

Notes: Figures A and B are the binned distribution of entrepreneurial quality and the percent of companies that move in each bin. Figure A is binned by quality of company, Figure B is binned by quality of location. Figure C is a scatter of all movers binned by individual destination-year, and reports the average local quality of the destination and the share of all movers in that year that move to this destination.
FIGURE 4A

Increase in Odds of Equity Growth

![Graph showing the increase in odds of equity growth.](image)

Each row represents an independent regression. Standard errors clustered at the state level (25 clusters). Sample excludes all California founded firms for first two groups. Moves to Silicon Valley estimates excludes all other migrants. Moves to other destinations excludes migrants to Silicon Valley. Increase in odds is the IRR of the effect minus 1. Standard errors are tighter for migrants outside S.V. due to a higher number of migrants in general. Green squares indicate preferred estimate. Appendix includes all regressions in tables and relevant statistics per regression.

FIGURE 4B

Increase in Odds of Equity Growth after Migration to Silicon Valley (sub-sample analysis)

![Graph showing the increase in odds of equity growth after migration to Silicon Valley.](image)

Each row represents a different regression. Standard errors clustered at the state level (25 clusters). Sample excludes all California founded firms and migrants to other destinations outside Silicon Valley. Increase in odds is the IRR of the effect minus 1. Standard errors adjusted to reflect errors in odds in right panel. All regressions include state-year pair fixed-effects and controls for firm quality. It is not possible to include the same LASSO controls in each regression due to them being too many for the smaller samples, and the possibility that differences would make the estimator sparse and inconsistent.
Marginal Effects.
Heterogeneous Gains Across Quality and Sorting Distributions

Delta in the Performance of Migrants vs Locals
Across Entrepreneurial Quality Distribution
Kernel Regressions with OLS model

Notes: The figure reports a kernel regression of the coefficient of migration across different points in the quality distribution of firms, using an epanechnikov kernel and a bandwidth of 0.5. All regressions include state-year fixed effects. Standard errors clustered at the state-year level.
Each row represents a different regression. Standard errors clustered at the state level (25 clusters). Sample excludes all California founded firms and migrants to other destinations outside Silicon Valley. Increase in odds is the IRR of the effect minus 1. Standard errors adjusted to reflect errors in odds in right panel. All regressions include state-year pair fixed-effects and controls for firm quality. It is not possible to include the same LASSO controls in each regression due to them being too many for the smaller samples, and the possibility that difference would make the estimator sparse and inconsistent.
FIGURE 6

Change in P(Equity Growth) after Migration to Silicon Valley

Notes: Represents the by-semester of migration coefficients of an equivalent regression of Table 8, Column 3—a linear probability model with equity growth as an outcome and LASSO controls and State-Year pair fixed effects. Standard errors are clustered at the state-year pair level.
Silicon Valley Migrants vs Non-Migrants

Notes: Model includes firm and age fixed effects, and reports the coefficient of time to migration by semester of age. Standard errors clustered by birth state (26 clusters). Dependent variables measured as cumulative values (stock) rather than per-period (flow).
Baseline is the mean value of $Y_t$ at $t=11$ for a matched sample of non-migrants of same quality.
Figure A is the mean quality and 95% percent confidence interval of mean quality for migrants that moved to Silicon Valley in each year. Figures B through F are individual year coefficients for the impact of moving to Silicon Valley across different outcomes, in a regression that also includes LASSO Controls and state-year pair fixed effects. The vertical line in each graph is on March, 31, of 2001, the date in which the NBER Business Cycle Dating Committee considers the post dot-com recession to have started.
FIGURE 9

Notes: This figure represents every address for which there is a Delaware firm in Silicon Valley. The color of the point is the average quality of firms, the size of the point is the number of firms founded in that address. The panel on the left represents all firms Delaware firms that are originally founded in Silicon Valley, and the panel on the right represents all migrants and their destination location. The scale of the points on the right is much larger than on the left to allow an easier comparability.
Proofs

**Proposition 1** Higher quality startups are more likely to move

\[
\frac{\partial P(S = 1)}{\partial \theta} > 0
\]

**Proof:** Consider that \( S = 1 \) when

\[
\theta \left( \frac{\tau_A}{1 + d_A(r^s)^{\beta_A}} \right) \left( \frac{\tau_L}{1 + d_L(r^s)^{\beta_L}} \right) - M > \theta \left( \frac{\tau_A}{1 + d_A(r^0)^{\beta_A}} \right) \left( \frac{\tau_L}{1 + d_L(r^0)^{\beta_L}} \right)
\]

Then we can simply re-arrange this formula to

\[
\frac{\left( \frac{\tau_A}{1 + d_A(r^s)^{\beta_A}} \right) \left( \frac{\tau_L}{1 + d_L(r^s)^{\beta_L}} \right)}{\left( \frac{\tau_A}{1 + d_A(r^0)^{\beta_A}} \right) \left( \frac{\tau_L}{1 + d_L(r^0)^{\beta_L}} \right)} > \frac{M}{\theta} \times \frac{1}{\left( \frac{\tau_A}{1 + d_A(r^0)^{\beta_A}} \right) \left( \frac{\tau_L}{1 + d_L(r^0)^{\beta_L}} \right)}
\]

So that the right hand side of this inequality decreases as \( \theta \) increases for all values of \( \theta \).

**Lemma 1** Startups closer to resources at founding are less likely to move

\[
\frac{\partial P(S = 1)}{\partial d_j(r^0)} > 0, \forall j
\]

**Proof:** We can see form the right hand side of the inequality in (1) that any increase in distance to \( r^0 \) increases the probability that the inequality is true.

**Proposition 2** Proof in text.

**Proposition 3** If knowledge matters for performance, then performance is higher for firms to locate to micro-regions with better local entrepreneurship

\[
\frac{\partial E[Y|S = 1, \theta]}{\partial \Theta_r}, r \in \{r^c, r^n\}
\]

**Proof:** This follows naturally from equation (2). Basically, because \( \Theta_r \) is positively correlated with agglomeration, then the value of this agglomeration (and therefore the likelihood to move to the destination) increases as \( \Theta_r \) increases.
FIGURE A1
States in Sample

Notes: All states in the data represented. The color represents the number of firms registered in Delaware in each state.
**FIGURE A2**

Notes: this figure shows the share of migrant firms that migrate within each of the quarters of firm life at different levels of firm quality. Firms are required to live at least a quarter in their location of birth to be considered migrants. It is easy to see a monotonic decline in migration rates, suggesting migration is mostly entrepreneurial.
**TABLE A1**

ENTERPRENEURIAL QUALITY MODEL
LOGIT REGRESSION
DEPENDENT VARIABLE: EQUITY GROWTH

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Business Registration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short Name</td>
<td>1.517**</td>
<td>1.419**</td>
</tr>
<tr>
<td></td>
<td>(0.0590)</td>
<td>(0.0517)</td>
</tr>
<tr>
<td>Corporation</td>
<td>10.44**</td>
<td>8.549**</td>
</tr>
<tr>
<td></td>
<td>(0.902)</td>
<td>(0.672)</td>
</tr>
<tr>
<td><strong>Intellectual Property</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patent</td>
<td>5.849**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.372)</td>
<td></td>
</tr>
<tr>
<td>Trademark</td>
<td>3.827**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.319)</td>
<td></td>
</tr>
<tr>
<td><strong>Name Based Industry</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local</td>
<td>0.817**</td>
<td>0.909</td>
</tr>
<tr>
<td></td>
<td>(0.0725)</td>
<td>(0.0797)</td>
</tr>
<tr>
<td>Traded</td>
<td>1.030</td>
<td>1.178**</td>
</tr>
<tr>
<td></td>
<td>(0.0607)</td>
<td>(0.0664)</td>
</tr>
<tr>
<td>Biotechnology</td>
<td>3.191**</td>
<td>2.252**</td>
</tr>
<tr>
<td></td>
<td>(0.293)</td>
<td>(0.220)</td>
</tr>
<tr>
<td>E-Commerce</td>
<td>1.418**</td>
<td>1.311**</td>
</tr>
<tr>
<td></td>
<td>(0.0842)</td>
<td>(0.0818)</td>
</tr>
<tr>
<td>IT</td>
<td>1.831**</td>
<td>1.534**</td>
</tr>
<tr>
<td></td>
<td>(0.153)</td>
<td>(0.118)</td>
</tr>
<tr>
<td>Semiconductors</td>
<td>2.385**</td>
<td>1.558*</td>
</tr>
<tr>
<td></td>
<td>(0.528)</td>
<td>(0.358)</td>
</tr>
<tr>
<td>Medical Devices</td>
<td>0.942</td>
<td>0.808**</td>
</tr>
<tr>
<td></td>
<td>(0.0691)</td>
<td>(0.0610)</td>
</tr>
<tr>
<td>State F.E.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year F.E.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>488960</td>
<td>488960</td>
</tr>
<tr>
<td>pseudo R-sq</td>
<td>0.097</td>
<td>0.141</td>
</tr>
</tbody>
</table>

Incidence rate ratios reported. Standard errors clustered at the state-year pair level. * p < .1 ** p < .05.
# TABLE A2
## PERFORMANCE OF MIGRANTS
### LINEAR PROBABILITY MODELS
#### DEPENDENT VARIABLE: EQUITY GROWTH

**PANEL A: MIGRATION ANYWHERE**

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Logit (IRR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Migrant (Anywhere)</td>
<td>0.0410**</td>
<td>0.0410**</td>
</tr>
<tr>
<td></td>
<td>(0.00794)</td>
<td>(0.00794)</td>
</tr>
<tr>
<td>Ln(Firm Entrep. Quality)</td>
<td>0.0112**</td>
<td>0.0106**</td>
</tr>
<tr>
<td></td>
<td>(0.00166)</td>
<td>(0.00146)</td>
</tr>
<tr>
<td>Ln(State Entrep. Quality)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASSO Controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State-Year F.E.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State F.E.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year F.E.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>232315</td>
<td>232315</td>
</tr>
<tr>
<td>R²/Pseudo R²</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>Mean Quality (All Firms)</td>
<td>0.00813</td>
<td></td>
</tr>
<tr>
<td>Mean Quality (Migrants)</td>
<td>0.0101</td>
<td>0.0101</td>
</tr>
</tbody>
</table>

**PANEL B: MIGRATION TO SILICON VALLEY**

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Logit (IRR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Migrant to Silicon Valley</td>
<td>0.0675**</td>
<td>0.0675**</td>
</tr>
<tr>
<td></td>
<td>(0.0102)</td>
<td>(0.0102)</td>
</tr>
<tr>
<td>Migrant Outside Silicon Valley</td>
<td>0.023**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00341)</td>
<td></td>
</tr>
<tr>
<td>Ln(Firm Entrep. Quality)</td>
<td>0.00915*</td>
<td>0.00878**</td>
</tr>
<tr>
<td></td>
<td>(0.00177)</td>
<td>(0.00169)</td>
</tr>
<tr>
<td>Ln(State Entrep. Quality)</td>
<td>-0.537**</td>
<td></td>
</tr>
<tr>
<td>LASSO Controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State-Year F.E.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State F.E.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year F.E.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>152298</td>
<td>152298</td>
</tr>
<tr>
<td>R²/Pseudo R²</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>Mean Quality (All Firms)</td>
<td>0.00813</td>
<td></td>
</tr>
<tr>
<td>Mean Quality (Migrants)</td>
<td>0.0141</td>
<td>0.0141</td>
</tr>
<tr>
<td><strong>Implied Increase in Odds</strong></td>
<td>8.301</td>
<td>4.802</td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the state level are reported. Firm Entrepreneurial Quality is the predicted value of performance from a machine learning model (random forest) in a trained to predict Equity Growth from at-founding characteristics. The sample used for training is excluded from this analysis. State Entrepreneurial Quality is the average quality of local firms born in that state and year, including both Delaware and non-Delaware firms. Mean quality is the average expected performance for all firms in the sample. Pseudo R2 is the McFadden (1974) R2 estimate. Estimates of size of unobservables to be non-significant assume the standard errors of the effect are the same. * p < .1 ** p < .05

Electronic copy available at: https://ssrn.com/abstract=3175328
**TABLE A3**
PERFORMANCE OF MIGRANTS VS NON MIGRANTS
OTHER OUTCOMES
LINEAR PROBABILITY MODELS

### PANEL A: MIGRANT TO SILICON VALLEY

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DEPENDENT VARIABLE:</strong></td>
<td>Patent</td>
<td>Trademark</td>
<td>Venture Capital</td>
<td>Sales</td>
<td>Latent Profits</td>
</tr>
<tr>
<td>Migrant to Silicon Valley</td>
<td>0.0944**</td>
<td>0.0739**</td>
<td>0.0728**</td>
<td>0.113**</td>
<td>0.854**</td>
</tr>
<tr>
<td></td>
<td>(0.0118)</td>
<td>(0.0139)</td>
<td>(0.0155)</td>
<td>(0.0171)</td>
<td>(0.0929)</td>
</tr>
<tr>
<td>LASSO Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State, Year F. E.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>152298</td>
<td>152298</td>
<td>152298</td>
<td>152298</td>
<td>152298</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.483</td>
<td>0.095</td>
<td>0.100</td>
<td>0.079</td>
<td>0.300</td>
</tr>
<tr>
<td>Mean of Outcome</td>
<td>0.0260</td>
<td>0.0379</td>
<td>0.0105</td>
<td>0.0375</td>
<td>-0.0612</td>
</tr>
</tbody>
</table>

### PANEL B: MIGRANT ANYWHERE

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DEPENDENT VARIABLE:</strong></td>
<td>Patent</td>
<td>Trademark</td>
<td>Venture Capital</td>
<td>Sales</td>
<td>Latent Profits</td>
</tr>
<tr>
<td>Migrant Anywhere</td>
<td>0.0451**</td>
<td>0.0766**</td>
<td>0.0245**</td>
<td>0.0509**</td>
<td>0.476**</td>
</tr>
<tr>
<td></td>
<td>(0.00644)</td>
<td>(0.0108)</td>
<td>(0.00601)</td>
<td>(0.00718)</td>
<td>(0.0473)</td>
</tr>
<tr>
<td>LASSO Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State, Year F. E.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>232315</td>
<td>232315</td>
<td>232315</td>
<td>232315</td>
<td>232315</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.520</td>
<td>0.107</td>
<td>0.121</td>
<td>0.079</td>
<td>0.343</td>
</tr>
<tr>
<td>Mean of Outcome</td>
<td>0.0347</td>
<td>0.0434</td>
<td>0.0170</td>
<td>0.0380</td>
<td>-0.0120</td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the state level are reported. Outcomes of patent and trademark do not include the first year of observations for the firm (which are incorporated in the quality model). Sales is an indicator variable on whether the firm reaches 1 million dollars. Latent profits is the projection of the first principal component of the outcomes IPO, Acquisition, Patent, Trademark, Venture Capital, and Sales. *p < .1 **p < .05