



# Getting schooled: Universities and VC-backed immigrant entrepreneurs<sup>☆</sup>

Natee Amornsiripanitch<sup>a</sup>, Paul A. Gompers<sup>b,\*</sup>, George Hu<sup>c</sup>, Kaushik Vasudevan<sup>d</sup>

<sup>a</sup> Federal Reserve Bank of Philadelphia, United States of America

<sup>b</sup> Harvard Business School and NBER, United States of America

<sup>c</sup> Harvard University, United States of America

<sup>d</sup> Yale University, United States of America

## ARTICLE INFO

### JEL classification:

I23

J61

L26

### Keywords:

Immigrant

University

Entrepreneurship

Startup

Venture capital

## ABSTRACT

This paper analyzes how US universities contribute to the quantity and quality of VC-backed immigrant entrepreneurship in the US. Using a novel data set that identifies immigration status and education history for the near-universe of VC-backed founders in the US, we document several interrelated facts. First, immigrants contribute disproportionately to US VC-backed entrepreneurship, accounting for approximately 20 % of VC-backed companies. More than 75 % of these immigrant entrepreneurs obtained post-secondary education in the US, which suggests that higher education represents a primary entry channel for foreign entrepreneurial talent into the country. Given these facts, we assess how universities shape both the geographic distribution and the quality of immigrant entrepreneurship. Close to 40 % of US-educated immigrants start a company in the state of their alma mater, suggesting that place of education substantially impacts immigrant entrepreneurs' startup location choice. Regarding firm quality, immigrant founders are also more likely to found financially successful and scientifically innovative startups than their US-born counterparts. Altogether, the results suggest that foreign students educated in US universities substantially contribute to local and national VC-backed entrepreneurship, thereby identifying higher education's global scope as a potential tool to attract entrepreneurial talent and encourage entrepreneurial growth.

## 1. Introduction

Immigrants play a vital role in innovation activities (Hunt and Gauthier-Loiselle, 2010; Bernstein et al., 2020) and entrepreneurship (Kerr and Kerr, 2016; Azoulay et al., 2020a; Azoulay et al., 2020b; Kerr and Kerr, 2020). Given the substantial contribution of immigrants in these areas, a set of natural questions arises: what pathways do immigrant entrepreneurs take to arrive in the United States, and how has the importance of these pathways changed over time? How do universities contribute to the quantity and quality of high-potential immigrant entrepreneurship? Do certain regions of the United States benefit disproportionately from high-skilled immigrant entrepreneurs, and if so, could the presence of local universities help explain the geographic

distribution of these benefits? The answers to these questions have important implications for designing immigration policy and regulation, which have become increasingly acrimonious topics in public discourse. They also have important implications for firms and universities that recruit talent from abroad, as well as for local communities that hope to promote vibrant entrepreneurial ecosystems.

To investigate these questions, we use a combination of unique datasets that allow us to (i) identify immigrant and native-born founders of venture capital (VC)-backed companies in the US and (ii) more closely study their educational backgrounds. Particularly, we combine a dataset from Infutor, which enables us to proxy for the immigration status of individuals in the United States, with VentureSource, which contains detailed information on the near-universe of VC-backed startups in the

<sup>☆</sup> Support for this research was provided by the Division of Research at the Harvard Business School. The authors thank Maryann Feldman and four anonymous referees for their insightful feedback and suggestions. Patrick Sweeney provided excellent research assistance. Paul Gompers has invested in and consulted for venture capital firms. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve Board, or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. No statements here should be treated as legal or investment advice.

\* Corresponding author.

E-mail addresses: [natee.amornsiripanitch@phil.frb.org](mailto:natee.amornsiripanitch@phil.frb.org) (N. Amornsiripanitch), [paul@hbs.edu](mailto:paul@hbs.edu) (P.A. Gompers), [george.hu@g.harvard.edu](mailto:george.hu@g.harvard.edu) (G. Hu), [kaushik.vasudevan@yale.edu](mailto:kaushik.vasudevan@yale.edu) (K. Vasudevan).

<https://doi.org/10.1016/j.respol.2023.104782>

Received 3 May 2022; Received in revised form 13 March 2023; Accepted 31 March 2023

Available online 29 May 2023

0048-7333/© 2023 Elsevier B.V. All rights reserved.

United States, including the identities of the startup's founder(s) and venture capital investors. We supplement these with manually collected data and resume data from Emsi, a labor analytics firm, which collectively allow us to observe details regarding the education and prior work experience of the entrepreneurs in our sample. Finally, the United States Patent and Trademark Office (USPTO)'s PatentsView dataset enables us to link US patent applications to VC-backed startups. These data together provide a rich and comprehensive source of background information on VC-backed entrepreneurs that we leverage to understand the educational experiences which high-skilled immigrants acquire before pursuing high-growth entrepreneurship.

Admittedly, our datasets do not allow us to study immigrant entrepreneurship on aggregate, since VC-backed startups represent only a small fraction of new firms created in the US at any given time.<sup>1</sup> Nonetheless, our focus on and comprehensive identification of VC-backed startups allows us to accurately examine immigrants' contributions to a particularly salient, relevant, and sought-after type of entrepreneurship. To elaborate, as prior literature (largely focused on the US economy) has shown, the VC ecosystem plays a crucial role in the US macroeconomy (Gornall and Strebulaev, 2015; Gompers and Lerner, 2000), and VC-backed firms contribute disproportionately to the right tail of the firm size and innovation distributions in the US economy (Akcigit et al., 2019). Venture-backed firms are also substantial job creators in the US economy—focusing on the economic contribution of immigrant founders of venture-backed firms highlights the job creating role that immigrants play in the economy, thereby bounding claims that immigrants primarily take jobs away from natives (Azoulay et al., 2020a, 2020b). Finally, various policymakers, representing localities ranging from Tel-Aviv to Bangalore, have endeavored to promote and foster high growth-potential entrepreneurship. Accordingly, understanding the contribution of immigrants to this important part of the economy is of interest per se. Additionally, while entrepreneurs who start venture-backed firms may be a selected sample, the detailed dataset we assemble on VC-backed immigrant entrepreneurs yields insights that are likely generalizable to high-skilled immigrants at large.

After assembling our novel dataset, we begin our analysis of the interplay between university education, immigrant talent, and VC-backed entrepreneurship by (i) identifying VC-backed entrepreneurs' immigration status and (ii) documenting two important facts about the immigrant entrepreneurs in our sample. First, to proxy for an individual entrepreneur's immigration status, we adapt Bernstein et al.'s (2020) approach. Specifically, we use the Infutor-documented age at which individuals in our dataset received their social security numbers (SSN) to identify their immigration status. Specifically, we define immigrants as those who received their SSN on or after their 18th birthday.<sup>2</sup> Mechanically, this approach excludes child immigrants, i.e., foreigners who arrived in the United States as children. However, as we discuss in more detail below, this limitation is unlikely to introduce significant biases into our empirical results. Using this definition of immigrant, we estimate that approximately 20 % of the VC-backed founders in our sample are immigrants, broadly in line with previous research that estimates a similar immigrant share in entrepreneurial and innovation activities (Kerr and Kerr, 2016). This initial fact suggests that immigrant founders are important contributors to the US VC-backed entrepreneurship ecosystem.

Next, we further utilize our unique data on VC-backed entrepreneurs' educational histories to better understand immigrants' educational pathways towards high-potential entrepreneurship. We categorize each

immigrant entrepreneur in our sample into one of three “pathway” categories: those who first came to the US for undergraduate study (Undergraduate/UG immigrants), those who first came for postgraduate education (Graduate/GR immigrants), and those who first came after receiving their education in a foreign country (Post Degree/PD immigrants). This classification allows us to more closely explore the extent to which multiple distinct institutions (i.e., undergraduate programs vs. graduate programs vs. corporations) contribute as entry points to the pool of foreign-born entrepreneurial talent in the United States. This education-based classification method allows us to document a second important fact: >75 % of immigrant entrepreneurs for whom we have education information received some form of education in the United States. Of this 75 %, roughly half received an undergraduate degree in the United States. The critical role played by higher education, and increasingly undergraduate education, as a primary entry point for high-potential immigrant entrepreneurs into the US provides new insights about the role of research universities in attracting high-skilled talent from abroad.

Motivated by universities' importance as a source of foreign entrepreneurial talent, we use the remainder of the paper to explore how universities add to the quantity and quality of VC-backed entrepreneurship in the United States. We begin this portion of the paper by exploring founders' propensity to start VC-backed companies in the state in which they received their final post-secondary education degree. We find that approximately 35 % of the founders in our sample found VC-backed firms in the state where they were educated. This fact is not simply driven by Berkeley- and Stanford-educated entrepreneurs founding companies in the Bay Area, or Harvard- and MIT-educated entrepreneurs founding companies in Massachusetts. Instead, it captures a more general phenomenon that also applies to graduates of universities in regions that are not hubs for VC investment.

To build upon this descriptive result concerning VC-backed founders' geographic stickiness, we also explore differences in would-be founders' propensity to stay in their state of education across the following groups: in-state natives, out-of-state natives, and immigrants.<sup>3</sup> We find that, relative to out-of-state native founders, immigrant founders are less likely to migrate before founding their startup when educated in venture capital hub states (defined in this paper to be California, Massachusetts, and New York), as opposed to other non-hub states. This disparity suggests that hub states disproportionately benefit from immigrant entrepreneurship, at the margin, partly because of post-education domestic migration. In contrast, non-hub states experience a relative brain drain of immigrant entrepreneurial talent, as would-be immigrant entrepreneurs are more likely to move to other states to start companies than their out-of-state native counterparts. Finally, to provide additional evidence of a distinct and clear connection between universities' enrollment of foreign talent and local entrepreneurial activities, we show that, holding demographic and local economic factors fixed, current student enrollment in local universities, both foreign and native, predicts future local VC-backed startup formation. There is little evidence of crowding out: enrolling more foreign students in local universities is not correlated with fewer future native-founded VC-backed startups.

This evidence suggests that the presence of research universities has likely been an important determinant of which areas have benefited most from the domestic migration of high-skilled founders, especially those of non-US/foreign origin. More generally, this result provides additional evidence for the role that universities play in local agglomeration economies. Universities are known to contribute to local economies in a variety of ways, for example by training a skilled labor force, or by knowledge diffusion from innovation activities (e.g., Hausman,

<sup>1</sup> A 2013 public press article (Entis, 2013) estimates that 0.05 % of startups in the US are funded by VCs.

<sup>2</sup> Our core conclusions are not sensitive to our choice on individuals' age. For example, changing the cut-off age to 16 or 22 does not materially affect the core results on immigrants' contribution to VC-backed entrepreneurship on the whole.

<sup>3</sup> In-state natives are natives who did not move out of state when completing their postsecondary education, whereas out-of-state natives are natives who did.

2020). Our results suggest that this agglomeration benefit extends to attracting skilled immigrants, some of whom end up starting high-growth potential firms.

In the last portion of the paper, we study the marginal differences in venture quality between companies started by immigrant and native entrepreneurs, and we examine the extent to which these differences apply to all three distinct immigrant pathway groups (Undergraduate, Graduate, and Post Degree). If one were to believe that selection into immigration or the treatment effects of the immigrant experience might materially influence the quality of VC-backed immigrant entrepreneurship, one might potentially be concerned that the quality of VC-backed companies in the United States may be diluted if immigrant entrepreneurs are less productive than their native counterparts. Using standard regression methods from the VC and innovation literature, we find that immigrants are more likely to start (i) companies that reach financial success (IPO and profitable acquisition) and (ii) companies that are more innovative, as encapsulated by several related measures of patent output. The results suggest that, on average, immigrants tend to start more productive companies than natives. The difference in productivity is most salient when comparing natives to immigrants who arrived in the United State for graduate school (Graduate immigrants) or work (Post Degree Immigrants). This nuance highlights the importance of immigrant entrepreneurs' educational pathways in informing and predicting their quality as VC-backed founders. The result is likely to be driven by a selection on quality effect (Borjas, 1989; Borjas, 1991; Borjas and Bratsberg, 1996; Rosenzweig et al., 2006) because we also find that Graduate and Post Degree immigrants tend to have more education than native founders and start companies in different industries than native founders. In other words, Graduate and Post Degree immigrant founders may be more adept at selecting into and exploiting opportunities in more productive sectors of the economy.<sup>4</sup>

From a policy perspective, our results emphasize the importance of immigrant entrepreneurs as founders of high-growth potential startups. While a substantial focus in the current public discourse revolves around work visas, such as the H-1B visa, our evidence suggests that student visas may deserve even more attention, given the role of universities in bringing talented, high potential foreign students into the country. Policy proposals in the past few years have sought to place restrictions on student visas for foreigners. Given the substantial contribution of immigrant entrepreneurs educated in the United States, our results suggest that such policies would likely carry significant costs for the country by restricting the supply of talented potential entrepreneurs. Our results also suggest that there is a substantial temporal lag between immigrant entrepreneurs' entry into the US and the founding of their firms. Accordingly, the effects of policies that increase or decrease the flow of immigrants may have very lagged but persistent effects on immigrant entrepreneurship. Such policies' effects would only show up decades after implementation and would be hard to reverse in a short period of time. Finally, although our data and results focus on the US economy, our empirical findings on the substantial nexus between university education, immigrant talent, and VC-backed entrepreneurship could inform non-US policymakers seeking to effectively support institutions and attract talent necessary to sustain vibrant domestic entrepreneurial hubs.

<sup>4</sup> Undergraduate immigrants are also likely to be positively selected on quality because, as a high school student in a foreign country, attending college in the United States requires non-trivial resources. Resources here refer to financial assets, the student's human capital, and the parents' human capital. Nonetheless, data on Undergraduate immigrants' educational backgrounds suggest that Undergraduate immigrants possess skills that are more similar to natives' skills, and these facts support the notion that Graduate and Post Degree immigrants are especially adept at entering and leveraging opportunities in the economy's most dynamic sectors.

### 1.1. Literature review and contributions

Our paper contributes to several segments of academic research and literature. The first main contribution is the collection of data that we use. We combine data from Infutor, VentureSource, Emsi, and Patents-View to study the how universities contribute to the quantity and quality of high-potential immigrant entrepreneurship in the United States. Although, the method that we use to identify immigrants introduces some measurement error, when compared to direct identification in the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) dataset (Kerr and Kerr, 2016), our data have several advantages. It contains information on entrepreneurs' education attainment, job titles, and work history, features not available in the LEHD, which allow us to study the immigration and educational pathways that would-be foreign entrepreneurs take to arrive in the United States. Additionally, given the cumbersome access procedure that is required for researchers to use the LEHD, our collection of data offers a robust alternative to study questions that are related to immigration, innovation, education, and entrepreneurship.

Second, our work addresses an unanswered question of significant policy importance: how do high-potential immigrant entrepreneurs come to the United States? We answer this question by assembling our detailed data that captures founders' work and educational history. Our data allow us to identify immigrant founders, summarize their educational background, and classify them according to their path of immigration. Therefore, we can paint a detailed picture of how immigrant entrepreneurs came to the United States and of the educational path that preceded their entry into high-growth entrepreneurship. Accordingly, this paper highlights how American universities represent a key source of foreign entrepreneurial talent for the country, a fact that possesses broad policy implications. In this light, our paper is related to the literature on education and immigration, which has been mostly focused on the relationship between foreign students' enrollment in American universities and their participation in the US labor force (Rosenzweig et al., 2006; Bound et al., 2015; Shumilova and Cai, 2016). Our study of would-be immigrant entrepreneurs' pathways to America is particularly related to Hunt's (2011) work that shows that the path and entry visa of immigration matters for the innovative and intellectual contributions that immigrants produce while residing in the United States.

As implied by its relation to Hunt (2011), our paper also adds the academic literature concerned with the contribution that immigrant entrepreneurs make to innovation and technological advances in the United States (Stephan and Levin, 2001; Kerr and Fu, 2008; Stephan, 2010; Balasubramanian and Sivadasan, 2011; Akcigit and Kerr, 2018). In this paper, we document that immigrant founders are more likely to hold STEM degrees, start information technology companies, and found startups that patent inventions. These results suggest that immigrant entrepreneurship is a channel through which American universities contribute to the commercialization of innovation and technology in the United States. Additionally, our paper contributes to the literature that studies the impact of immigrant entrepreneurship on local job growth and economic development (Kerr, 2010).

In addition, we show that, for both native-born and immigrant founders, education location is an important determinant of startup location. In other words, founders are more likely to start their companies in the state where they received their post-secondary education. This result contributes to the literature on the determinants of firm location (Carlton, 1983; Bartik, 1985; Reynolds et al., 1994; Sorenson and Audia, 2000; Stuart and Sorenson, 2003; Masumba et al., 2011). Our work particularly complements several papers' findings. Specifically, Fini et al. (2022) find that likelihood of self-employment is higher among those who study and stay in their home region. Eckhardt et al. (2022) find that the composition of university students' origin matter for the likelihood that those students will become entrepreneurs in their universities' local economy. Overall, our work suggests that establishing high quality universities to attract both talented native-born and foreign

students may be a viable strategy to promote local high-growth firm creation.

Finally, our paper sheds light on the role of universities in bringing immigrants into the United States' entrepreneurial ecosystem. Our work thus contributes to a broader literature that focuses on immigration, education, and entrepreneurship (Etzkowitz, 1998; Di Gregorio and Shane, 2003; Bramwell and Wolfe, 2008; Bound et al., 2021; Åstebro et al., 2012; Grogger and Hanson, 2015; Guerrero et al., 2015; Hanson and Slaughter, 2017; Lee and Eesley, 2018; Kerr, 2020; Uhlbach et al., 2022). Prior work has demonstrated that universities, through different levers (e.g., creating entrepreneurship-focused academic programs or making early-stage equity investments in affiliate-founded spin-offs) can contribute to local economies by producing and supporting individuals, both students and faculty, who start high-growth companies. Specifically, prior literature has been largely focused on the role that universities play as providers of technologies that could be commercialized and converted into high-potential startups (Audretsch et al., 2017; Tartari and Stern, 2021; Babina et al., forthcoming). We contribute to this line of work by showing that universities contribute to the *quantity and quality* of high-potential entrepreneurship in the United States at both the local and the aggregate level. In this light, our findings are most related to the work by Baptista et al. (2011), which finds a positive correlation between the establishment of universities and the entry/formation rate of knowledge-based firms. A key takeaway from our work is that the majority of VC-backed entrepreneurs are educated in the United States; in fact, many are educated at top US universities and choose to start firms in close proximity to their place of education. The role of universities in bringing high-skilled immigrants to the United States complements the results of the literature, and suggests that student visas, as well as immigration policies surrounding foreign students more broadly, are a critical area on which immigration and entrepreneurship policy should focus.

The remainder of the paper is structured as follows. Section 2 describes our various data sources and discusses the assembly of our dataset. Section 3 provides descriptive statistics of our dataset and introduces two important facts about the immigrant share of VC-backed entrepreneurship in the US. Section 4 presents more rigorous empirical analyses on how immigrant and native-founded VC-backed startups differ in their geographic distribution, financial success, and innovative activity. The section in particular focuses on analyzing the extent to which US university education might influence immigrant founders' startup activities along these dimensions. Finally, Section 5 concludes by discussing policy implications and potential avenues for future research.

## 2. Data

Our analysis of immigrant founders utilizes several data sources that help identify the immigration status, educational background, and work history for nearly all founders of VC-backed companies in the US. The first source is the Infutor database, which contains address history and information for US residents. The Infutor database is especially useful for our study because we use it to construct a reasonable proxy for the immigrant status of all individuals in our data. The second main data source we use is Dow Jones VentureSource (VS), which is one of the main databases used to study VC-backed firms. VentureSource provides information on VC-backed founders and their startups' ultimate financial outcomes. We also collect resume data for the vast majority of US-based founders in VS from Emsi, a labor analytics firm. These data contain work history and education information for the founders in our sample. We supplement these data with additional, hand-collected data on the education of founders for whom Emsi does not have a resume. Finally, we obtain information on firm patenting from PatentsView, which provides a publicly available database covering patents granted by the US Patent and Trademark Office (USPTO). Having collected data from these various sources, we proceed to describe how we merge their key components to obtain the final dataset that we use for our descriptive

and empirical analysis of VC-backed founders. We discuss these data and their uses within the final merged dataset for our analysis in more detail below.

### 2.1. Data sources

#### 2.1.1. Infutor

The Infutor database provides a variety of information for >260 million US residents. The data is aggregated from various sources including phone books, magazine subscriptions, and credit header files. For each individual, the database contains the individual's first and last name as well as complete address history with date and exact street address for each location. The dataset also contains demographic information for many individuals in the sample, including birth year, gender, and social security number (SSN). Diamond et al. (2019) and Bernstein et al. (2020) have also utilized the Infutor data and have found that the data appear to be largely representative of the overall US population. The address history appears to be reasonably comprehensive starting in 1990, though some individuals have address information going back to the 1980s. Other than the Census Bureau data that Kerr and Kerr (2020) use, Infutor data's coverage render it one of the most representative datasets with which to study the immigrant population in the US. More details on the Infutor dataset are provided in Appendix A.1.

Infutor data coverage likely skews towards property owners and individuals who are actively participating or involved in the formal financial system within the US given that the data are collected from credit header files. However, when considering the context of VC-backed entrepreneurship that underlies our empirical analysis, the overrepresentation of these individuals should not be of significant concern because property ownership and participation in the formal credit market represent de facto prerequisites that facilitate entry into high-potential entrepreneurship (Chaney et al., 2012; Bell et al., 2019). Both property ownership and activity in formal credit markets are likely highly correlated with participation in VC-backed entrepreneurship. Therefore, empirical results from the Infutor-VS merged data examining the demographic and educational backgrounds of VC-backed entrepreneurs (discussed in Sections 3 and 4) should not be biased by Infutor's potential omission of individuals who (i) do not own property or (ii) do not have credit files.

#### 2.1.2. Dow Jones VentureSource

The Dow Jones VentureSource (VS) dataset contains information on the near universe of venture capital fund investments in startups from around the world and is one of the two main datasets used in academic research on the venture capital industry.<sup>5</sup> VentureSource has several distinct data files. First, VentureSource tabulates investment data that contain information at the portfolio company (i.e., startup) level on investments dates, investment amounts, and identity of venture capital firms participating in each round. From these data, we can trace a startup's funding history. VentureSource also collects a variety of information on the portfolio company including company start date, company industry, business description, and office location. Additionally, VentureSource collects information about individuals associated with each startup, including venture capital investors, founders, board members, and senior employees. For these individuals, VentureSource provides first and last names as well as information on prior work experience, including names of past employers, past job titles, and the associated dates of employment. We focus our analyses on startups that are based in the United States. Following other work in the literature, we focus on firms that receive VC-backing funding from 1990 to 2019 because both Infutor and VentureSource data has substantially more comprehensive coverage after 1990 (Gompers et al., 2005; Gompers et al., 2010; Amornsiripanitch et al., 2019). Overall, VentureSource

<sup>5</sup> The other is Thompson VentureExpert.



provides information on 86,378 founders and 50,063 startups globally. Of these, 53,273 founders and 31,095 startups are based in the US. 98 % (over 30,000) of these US-based startups were founded during the 1990–2019 period (i.e., the timeframe of our empirical analysis).

### 2.1.3. Emsi

The third data source that we use is Emsi resume data, which is provided by Emsi Burning Glass Technologies, also called Lightcast.<sup>6</sup> The company specializes in providing labor market data such as resume and job posting data for individuals and companies working or operating in the United States. The Emsi dataset is the company's primary resume data product. The data are gathered from various internet sources. In practice, this gathering method likely implies that most of the Emsi data come from LinkedIn, especially when focusing on a sample of VC-backed entrepreneurs. The data are updated every month. For individuals who appear in the data, we have multiple pieces of “profile level” (i.e., individual level) information. In addition to first name, middle name, and last name, the data cover job title, company, start date, end date for all jobs reported in the individual's work history, as well as degree, education institution, start date, and end date for all degrees reported in the individual's educational history.

For this project, we sent a list of all founders who appear in the VentureSource (VS) to Emsi so that the company could match them to individuals within their entire universe of profiles. Emsi was able to uniquely match a subset of founders. In addition, we acquired approximately 5 million separate profiles from Emsi. These profiles contain the complete education and work history information for all undergraduate alumni from 50 leading universities in the US. Given these additional Emsi profiles and data, we use the procedure described in Appendix A.2 to match additional founders in VS to Emsi profiles. In total, we are able to match 61 % of the VS founders who started at least one VC-backed company in the US to Emsi profile data. We use this merged VS-Emsi founder-level data as well as hand-collected education information, which we describe in more detail in Section 2.1.5.

### 2.1.4. Founders' gender and race/ethnicity

Founders' genders are primarily determined based on their first names. In cases of unisex names, we determine gender by reading news articles and web pages mentioning or containing pictures of the individual founders. For race/ethnic background, we use the name-matching algorithm developed by Kerr and Lincoln (2010) to determine the most likely race/ethnicities of founders based on their last names. Individual founders are classified into the following racial/ethnic groups: East Asians, Indian, Jewish, Hispanic, and White. The name-matching algorithm does not allow us to identify all possible ethnicities. For example, we often cannot determine whether a founder is White or African American based on last name alone. Therefore, for all founders classified as White, we manually search for online pictures to determine if the founder is White or Black. As discussed in prior research, the percentage of Black founders in the VC-backed startup space is very small. More details on the gender and race/ethnicity identification procedures are provided in Appendices A.3 and A.4.

### 2.1.5. Founders' education and employment data

For each founder in the VentureSource data, we collect information on educational history and prior employment using both resume data from Emsi and hand-collected education and work experience information from LinkedIn, Bloomberg Businessweek, and company websites. We collect education data for 87 % of the founders in our final merged sample (as well as for 92 % of all founders in VentureSource). For founders with complete background information, we observe their undergraduate institution, undergraduate major, graduate institution(s), graduate degree(s), and year(s) of graduation, as well as prior work

history. We aggregate colleges, professional schools, and graduate schools up to the institution level. For example, Harvard College and Harvard Business School are coded as Harvard University. Using information on undergraduate and graduate majors, we classify degrees into three categories: STEM, business, and other. More details on the education data collection procedure are provided in Appendix A.5. We also collect information on the geographic location of universities using the Google Maps API.

### 2.1.6. Patent and other supplemental data

To obtain information on startup-level patenting activities, we link companies in VentureSource to patent assignees in the USPTO PatentsView data by matching company names within both datasets. We Derwent-standardize company names via methods based upon Hall et al. (2001). Other research that matches patents with venture-capital backed firms employs a similar matching procedure (e.g., Bernstein et al., 2016; Howell et al., 2020). Upon linking patenting VS startups to patent assignees in PatentsView, we collect information on (i) whether a startup has filed any successful patent applications (“patent indicator” or “patent rate”), (ii) the number of patents assigned to a startup (“patent count”), and (iii) the number of patents assigned to a startup, weighted by forward patent citations (“citation weighted patent count”). These collected variables on firm patenting activity also follow existing standards in the literature. Prior work, including Hunt (2011) and Brown et al. (2019), has constructed “patent indicator” variables to characterize the extensive margin of startup and individual-level patenting (i.e., whether a firm decides to enter into patenting activity), while patent count and citation-weighted patent count measures remain standard for evaluating the productive output and scientific value of firms' innovative patenting activities (Hall et al., 2001; Hall et al., 2005; Kogan et al., 2017).

Finally, we use several standard datasets to collect relevant state-level information for our empirical analysis examining the relationship between local university enrollment and local startup activity. Native and foreign student enrollment data are collected from two standard higher education datasets: College Scorecard and the Integrated Post-secondary Education Data System (IPEDS). For each state-year pair, these datasets provide the aggregate number of foreign student enrollees at the undergraduate and graduate levels, as well as their share of total undergraduate and graduate enrollees. From these variables that identify foreign and native student enrollment at the graduate and undergraduate level, we can study the correlation between university enrollment and local entrepreneurship activity. We collect these data for the years 2000 to 2018 (we are unfortunately unable to collect analogous data for earlier years). In addition, data used to construct local economic condition and demographic control variables at the state-year level are collected from the U.S. Census Bureau's American Community Survey, the Bureau of Labor Statistics's Local Area Unemployment Statistics, and the Bureau of Economic Analysis's Regional Economic Accounts. Specific variables include state-level income per capita, unemployment, labor force participation, US-born population share, and white population share during the 2000–2018 period. Appendix C.1 provides additional details on these collected variables.

## 2.2. Merging procedure and identification of immigrant founders

### 2.2.1. Infutor-VS merge and identification of immigrant founders

In this section, we describe how we identify founders' immigrant status. To determine immigrant status of VC-backed founders in VentureSource, we match individuals from VentureSource with Infutor. More specifically, to identify founders in the Infutor data, we use an iterative procedure that matches individual-level observations using name, location, and age information. Infutor contains a list of an individuals' residential address history and date of birth. We use this information to match individuals into VentureSource, which contains information on founders' names and the locations of founders' startup

<sup>6</sup> The data can be found at <https://www.economicmodeling.com/>.

firms. To further verify that our match is correct, we use supplementary information on individuals' education graduation years (added to the VS founder data) to infer an approximate range for a founder's birth year. We use this graduation information to eliminate potential individual matches in Infutor whose ages do not align with the ages of founders in the VentureSource data.

We are able to identify a unique match within Infutor for approximately half of the founders in VentureSource. Some additional founders in VentureSource have multiple potential matches in Infutor. In these instances, we classify founders as immigrants if >80 % of potential matches in Infutor are immigrants, native-born if <20 % of potential matches are immigrants, and do not assign an immigrant classification otherwise.<sup>7</sup> Overall, our merged dataset includes 70 % of the US-based founders in VentureSource. After further restricting the merged dataset to US-based founders in VentureSource with non-missing education information, the resulting dataset, which we use as the starting point for our analyses, accounts for 72 % of US-based founders in VentureSource with non-missing education information, as well as for 61 % of all US-based founders in VentureSource (with or without education information). In Appendix A.6, we discuss the procedure for matching our data in more detail. Appendix A.7 provides a comparison between the matched and unmatched observations in our sample.

We follow the approach of Bernstein et al. (2020) when we identify VC-backed founders' immigrant status using data from Infutor. This identification exploits the fact that from 1936 to 2011, social security numbers (SSNs) were assigned using a specific formula. The first three digits of the (nine-digit) SSN (the "area number") reflect the geographic region (state) in which the social security number was assigned. The next two digits corresponded with a "group number," while the last four digits are an individual-specific serial number. Group numbers were assigned sequentially within a geographic region over the given time period, i.e., for a given area number, the same group number was used for all SSNs until all possible serial numbers (the last four digits), ranging from 0001 to 9999, were exhausted. Accordingly, any combination of the first five digits of the SSN was only assigned during a certain year or couple of years within a certain state. The mapping from the first five digits of an SSN to state and year(s) in which the SSN was issued is publicly available.<sup>8</sup>

Using this mapping and the data from Infutor, we are able to estimate the age at which any individual received her SSN using (i) an Infutor-provided date of birth variable and (ii) the year associated with her SSN group number (detailed in the previous paragraph). We accordingly classify immigrants as individuals who received their SSNs on or after the age of 18. We classify all individuals who received their SSNs before the age of 18 as native-born. Our empirical results are not sensitive to the cutoff age we use to distinguish between immigrants and native-born Americans, although the choice of cutoff age does slightly influence the proportion of founders whom we identify as immigrants.<sup>9</sup>

Having ascertained the immigrant status of most VS founders, we arrive at the final sample for our empirical analysis by utilizing the VS founders with non-missing information about (i) their immigration status and (ii) their educational background. This final sample contains approximately 32,000 founders and 36,000 founder-company pairs. Overall, the sample covers 61 % of approximately 53,000 US-based founders in VentureSource, while it covers 72 % of the approximately

46,000 US-based founders in VentureSource with non-missing education information. Finally, this sample covers approximately 24,000 of the 31,000 US-based startups in VentureSource. A key innovation that our final merged dataset allows us to make is the ability to identify the immigration pathways/US entry points of VC-backed immigrant entrepreneurs in the US. To elaborate, the procedure described in the previous paragraphs first allows us to divide our founders into two groups: natives and immigrants. Furthermore, our data's education and work history information enable us to sort immigrant founders into three pathway-based groups. Immigrant founders who arrived in the US to receive an undergraduate degree are classified as Undergraduate (UG) immigrants. Immigrant founders who did not receive an undergraduate degree from a US institution but instead arrived in the US to receive a postgraduate degree are classified as Graduate (GR) immigrants.<sup>10</sup> Immigrant founders who did not receive a postsecondary degree from a US institution are classified as Post Degree (PD) immigrants. This group covers immigrant founders who initially came to the United States for work.<sup>11</sup> Throughout the paper, we focus our sample on founders who have sufficient immigration status, education, and work information such that we can categorize them as natives, UG, GR, and PD immigrants.

### 2.2.2. Limitations and caveats to our definition of immigrant

While the US Census Bureau's various datasets used in other immigrant-focused research directly identifies individuals' place of birth, entry visa, and subsequent visa history and allows researchers to accurately identify immigration status with near certainty,<sup>12</sup> Census data may not provide useful comprehensive information about individuals' complete within-US address and migration histories, which our data can ascertain. Furthermore, access to these data is restricted and difficult to obtain for most researchers. Additionally, we do find many prior papers that use such data to specifically study VC-backed entrepreneurship. Consequently, we use an indirect method to identify and classify immigrants. This indirect method may raise several potential concerns, which we discuss in detail below.

<sup>10</sup> For example, an immigrant who receives his or her SSN during a short non-degree-granting program and, later, returned to the US to complete a graduate degree would be classified as a Graduate immigrant. An interesting group of individuals are those who were born in the US, grew up abroad, and returned to the US for college. In principle, this group of individuals is more likely to behave like Undergraduate immigrants, but will be classified as natives, which may introduce some biases into the statistical analyses presented in Section 4. However, for the purpose of this paper, the bias is likely to be small because, by size, this group of individuals is very small. The Center for Immigration Studies estimates that approximately 36,000 women came to the US to give birth and leave. Compared to the total number of immigrants, approximately 45 million, this number is very small, as this article notes: <https://www.voanews.com/a/foreigners-seeking-american-citizenship-children-flout-law-endanger-babies/3626080.html>. See <https://www.pewresearch.org/fact-tank/2020/08/20/key-findings-about-u-s-immigrants/> for the number of immigrant population in the US.

<sup>11</sup> The reader may be concerned that we were unable to merge 30 % of VS founders to individuals in Infutor. Appendix Tables B.6–B.7 and Appendix Fig. D.1–D.2 present a set of robustness checks where we include unmerged founders who did not receive a postsecondary education degree from a US institution as Post Degree Immigrant founders. This method increased our Infutor match rate to 76 % for all US-based founders in VentureSource and 79 % for US-based founders in VentureSource with non-missing education information. In result, 68 % of all US-based founders in VentureSource possess non-missing immigration and education information under this alternate method. Furthermore, the resulting robustness analysis shows that our core results are largely unchanged.

<sup>12</sup> However, even the US Census data's visa history information would be unable to disentangle whether immigrants initially arrived in the US for undergraduate vs. graduate education. Our dataset's ability to disentangle these separate pathways represents one of its advantages vis-à-vis Census data.

<sup>7</sup> To minimize the risk of misclassifying founders' age, we also do not assign a specific birth year or startup founding age to founders in these "multiple match" instances.

<sup>8</sup> We use data from the website [www.ssn-verify.com](http://www.ssn-verify.com) to map from the first five digits of SSN to state and year, once again following Bernstein et al. (2020).

<sup>9</sup> While Bernstein et al. (2020) use a cutoff age of 20, we use a cutoff age of 18 to better capture individuals that may have arrived in the United States for college. The number of foreign-born college students has increased dramatically over the past thirty years.

**Table 1A**

Founder characteristics by immigration status.

The table presents summary statistics for native-born and immigrant founders' characteristics. Each observation is a founder. IPO Rate is the percentage of the founder's startups that had gone public by 2019. Success Rate is the percentage of the founder's startups that, by 2019, either reached the IPO stage or were acquired for more than the total amount of money invested, adjusted for inflation. Patent Rate is the percentage of the founder's startups that successfully filed at least one patent within two years of founding. Patent Count is the number of patents filed by the founder's startups within two years of founding. Citation-Weighted Patents is defined as the number of patents, weighted by forward patent citations, filed by the founder's startups within two years of founding. Number of Firms counts the number of VC-backed ventures that each founder had started throughout the sample. Founding Age is the average of the founder's age at the time that each of his startup was formed. Asterisks denote statistical significance level from *t*-tests on differences in sample means, \* is for 10 %, and \*\*\* is for 1 % level.

	(1)	(2)	(3)	(4)	(5)	(6)
	Natives		Immigrants		Natives – Immigrants	
	N	Mean	N	Mean	Difference	t-Statistic
Female	26,439	0.09	6199	0.10	−0.008	−1.81
Jewish	26,439	0.19	6199	0.13	0.062***	12.73
East Asian	26,439	0.05	6199	0.18	−0.132***	−26.24
Indian	26,439	0.04	6199	0.32	−0.283***	−46.93
Hispanic	26,439	0.04	6199	0.07	−0.031***	−9.05
White	26,439	0.70	6199	0.34	0.356***	53.45
# of firms	26,439	1.11	6199	1.14	−0.0303***	−4.44
Avg. team size	26,439	2.37	6199	2.41	−0.035*	−2.09
Founding Age	16,115	39.46	4104	43.81	−4.347***	−21.12
IPO Rate	26,439	0.04	6199	0.05	−0.002	−0.80
Success Rate	26,439	0.15	6199	0.17	−0.026***	−5.20
Patent Rate	26,439	0.14	6199	0.17	−0.032***	−6.39
Patent Count	26,439	0.47	6199	0.70	−0.221***	−4.46
Citation-Weighted Patents	26,439	14.33	6199	17.80	−3.464*	−2.06

From the perspective of the United States' population, the most conventional and accepted definition of the term “immigrant” is an individual born outside of the United States. Comparing our method's classification of immigrant status to this conventional definition implies that we will, in principle, misclassify several segments of the foreign-born population in the US. First, child immigrants, i.e., individuals who immigrated to the United States before the age of 18, would be classified as natives in our sample. This misclassification introduces the concern that any comparison that we make between our indirectly identified group of native founders and our three (indirectly identified) groups of immigrant founders may be biased. Hence, if child immigrants are not sufficiently similar to natives along certain dimensions that we study (e.g., VC-backed entrepreneurs who are child immigrants are more successful than native entrepreneurs), our analysis may mischaracterize the extent to which immigrants' contributions to VC-backed entrepreneurship might differ from natives'.

We believe that such statistical biases should be small. First, prior work has shown that, when considering the innovative and entrepreneurial outcomes that this paper examines, child immigrants are more similar to natives than to high-skilled immigrants who came to the US during adulthood for postsecondary education. Hunt (2011) finds that, in terms of innovation, commercialization, and the dissemination of knowledge through publication, child immigrants perform similarly to natives, while immigrants who came to the US for postsecondary education outperform both groups. Furthermore, Blume-Kohout (2016) finds that child immigrants' proclivity to start a STEM-oriented company is similar to natives', while the probability of starting a STEM-oriented company is higher among immigrants who moved to the US for college. These two studies suggest that it is reasonable to assume that, in the context of VC-backed entrepreneurship, child immigrants perform similarly to natives. Second, if any such statistical bias were to remain in our estimates of immigrant entrepreneurs' performance and contributions, the size of such bias is likely to be small because child immigrants make up only a very small proportion of the total immigrant population in the US. Indeed, US Census statistics from 2018 find that approximately 42 million (94.3 %) of the 44.8 million immigrants in the US are age 18 or older (Budiman et al., 2020). Finally, as we discuss in more detail below, where comparisons can be made, our descriptive and empirical results remain largely in line with those from studies that use US Census data to identify immigrants (Kerr and Kerr, 2016; Azoulay

et al., 2020a; Azoulay et al., 2020b). Such similarity between our data and US Census data suggests that the measurement error issues likely have at limited impact on our analysis. These facts help mitigate concerns that our classification of child immigrants as natives may bias our results.

Additionally, our study classifies second-generation immigrants as natives. We find this grouping standard and justified for several reasons. First, if child immigrants are sufficiently similar to natives along the relevant dimensions of this study, then it almost certainly follows that second-generation immigrants will be similar to natives along these dimensions. Second, per the natural definition of immigrants that we highlight above, second-generation immigrants would be classified as natives. Key papers in the literature on immigrant entrepreneurship also exclude second-generation immigrants from their analysis (Kerr and Kerr, 2016; Azoulay et al., 2020a; Azoulay et al., 2020b). Thus, while the entrepreneurial activities of second-generation immigrants may be an interesting research topic, it is well beyond the scope of this paper.

Finally, another group of immigrants that our identification method misclassifies is undocumented or unauthorized immigrants who do not have SSNs. Indeed, since we would not be able to identify such immigrants' immigration status in Infutor, any such individuals would be excluded from our merged Infutor-VS dataset and would therefore not be considered in our empirical analysis. This omission should not be a significant concern for our empirical analyses because it is extremely difficult for undocumented immigrants to participate in the formal economy due to institutional frictions. Hence, it is intuitive to conclude that unauthorized immigrants play no more than a negligible role as founders in the US VC-backed entrepreneurship ecosystem. Therefore, even if we were able to identify undocumented immigrants in the Infutor data, it is unlikely that the same individuals would show up in the merged Infutor-VS dataset used in our analysis.

### 3. Summary statistics

In this section we present descriptive statistics on the characteristics of native and immigrant founders in the sample used for our empirical analysis. This section has three goals. First, since the data that we have assembled are new to the literature on immigrant entrepreneurship, it is worthwhile to document stylized facts about VC-backed immigrant entrepreneurs in the United States which may be informative for future

research. Second, the section demonstrates that, when compared to prior work using US Census data to evaluate total immigrant entrepreneurship, the descriptive statistics derived from this new dataset are similar. In other words, the descriptive statistics help to verify that our method for identifying immigrant entrepreneurs accurately classifies entrepreneurs' immigrant status and yields reasonable qualitative results. Third, this section provides suggestive evidence that (i) VC-backed immigrant entrepreneurs differ from native entrepreneurs in important dimensions and (ii) universities appear to serve as an important source of this distinctive foreign entrepreneurial talent within the US. We thus use these descriptive findings to motivate and frame the more rigorous statistical analyses that we present later in the paper.

### 3.1. Founder and startup characteristics by immigration status

We first summarize the main variables in our final sample (i.e., the merged Infutor-VS data, restricted to founders with non-missing education information). Table 1A presents summary statistics on various founder characteristics (e.g., demographics, educational background, startup activities) by immigration status. Approximately 20 % of the founders in our sample are immigrants. This share is higher than the US population's aggregate immigrant share, which rose from under 10 % to 13.7 % percent during the 1990–2018 period which we study (Budiman et al., 2020). Nonetheless, as we discuss in Appendix A.7, due to data limitations, this number may slightly understate the proportion of immigrant founders in our sample. Kerr and Kerr (2016) find that 28 % of VC-backed startup founders are immigrants. Our results are broadly consistent with Kerr and Kerr (2016), even though they define immigrant founders by recorded country of birth. To reiterate, we in contrast define immigrants based on an individual's age of SSN attainment (detailed in Section 2.2.1). Given this difference in definitions, we would expect to see a slight difference in the proportion of founders reported as immigrants in the two samples, even if the true proportions are identical. Each of these two approaches to defining immigrants contains advantages and disadvantages. On the one hand, Kerr and Kerr (2016) identify those born abroad but brought to the US as children as immigrants. While this identification more closely follows the conventional definition of immigrant, policy reforms targeting this type of immigration may be difficult. On the other hand, we decide to only identify individuals who voluntarily chose to come to the US as immigrants. This definition of immigrant is certainly less conventional/intuitive, but it allows us to more directly examine how selection into immigration and institutional entry points might influence future immigrant entrepreneurial activity. This focus appears more likely to generate relevant insights that can directly guide various portions of entrepreneurship, innovation, education, immigration, and talent recruitment policy.

Table 1A highlights several differences between native-born and immigrant founders in our sample. First, immigrant founders have a slightly higher proportion of females than the native-born subsample (though not statistically significant). Second, immigrant founders are more racially/ethnically diverse. Most significantly, the proportions of Indian and East Asian founders are much higher among immigrant founders than among native-born founders. In fact, the proportion of Indian founders in the immigrant subsample is almost ten times higher than that of the native-born subsample. Similarly, the proportion of East Asian founders is more than three times higher. This difference is not surprising since a significant influx of immigrants from India and China arrived in the United States to pursue education and employment opportunities, particularly in high-tech sectors over the 1990–2019 period. This table also demonstrates a notable advantage to our approach of classifying immigrants using SSN relative to commonly employed/alternative classification schemes which rely on name-based algorithms to proxy for immigration status. A substantial proportion of immigrant founders (36 %) in our sample are classified as White. These are founders whom an ethnic classification algorithm à la Kerr and Lincoln (2010) may not identify as immigrants. The high proportion of

immigrant founders identified as White suggests a substantial presence and contribution of immigrant entrepreneurs from Canada, Western Europe, Eastern Europe, and Commonwealth countries with significant white populations within the United States' VC-backed entrepreneurial ecosystem.

In line with previous literature on immigrant founders' productivity, conditional on being a VC-backed founder, immigrants start more VC-backed companies than native founders on average.<sup>13</sup> Furthermore, the average success rate of immigrant-founded startups is 2 percentage points higher than that of native-founded startups. Throughout our analysis, a founder's startup is considered successful if (i) it achieves an IPO or (ii) is acquired at a value that is greater than the value of its total VC investment.<sup>14</sup> Though more rigorous analysis is needed to corroborate the full implications, this descriptive result offers preliminary evidence that immigrant founders might perform exceptionally well when compared to native-born counterparts. Furthermore, we find that immigrants tend to be older than native-born founders at the time that their firms obtain venture capital funding. The average founding age of immigrant founders is 43.8 years old, while the average age for native-born founders is 39.5 years old.<sup>15</sup> Finally, consistent with Azoulay et al. (2020a), we find that immigrant founders tend to start more innovative firms as measured by patenting activity. On the extensive margin, immigrant founders are 3 percentage points more likely to start a company that produces at least one patent within the first two years of existence. On the intensive margin, immigrant founders start firms that produce 0.2 more patents within the first two years of existence than firms started by native-born founders. This result preliminarily suggests that, even when abstracting away from the economic effects of their financial success, VC-backed firms started by immigrant founders may disproportionately benefit the US economy by producing more innovations than firms started by native-born founders.

We next consider the sectors in which immigrant founders establish their startups. In Table 1B, we present industry breakdowns for VC-backed startups founded by native-born and immigrant entrepreneurs. Immigrants are significantly more likely to start a company in the IT sector than natives and significantly less likely to start companies in the Business and Finance or the Consumer Services sectors. The proportion of immigrant-founded companies that are in the IT sector (48.5 %) is >30 % higher than the proportion of native-founded companies in IT (35.6 %). The differences in Business and Finance (18.2 % of immigrant-founded companies vs. 22.5 % of native-founded companies) and Consumer Services (12.3 % of immigrant-founded companies vs. 18.2 % of native-founded companies) are equally striking.<sup>16</sup> The result logically extends the implications of prior research results, which find that high-skilled immigrants tend to come to the United States to study in STEM fields and pursue STEM-related employment opportunities (Hanson and

<sup>13</sup> Kerr and Kerr (2020) find that immigrants tend to start more companies than natives. Azoulay et al. (2020a) find that, at every point of the firm size distribution, immigrants start larger companies than natives. Bernstein et al. (2020) and Hunt and Gauthier-Loiselle (2010) find that immigrant inventors produce more patents than native-born inventors.

<sup>14</sup> We consider M&A deals where the startup was valued at an inflation-adjusted dollar amount that is greater than the total amount of money that was raised by VCs as successful acquisitions.

<sup>15</sup> These numbers for founding ages are broadly in line with results found in other work. Azoulay et al. (2020b) find that the average founders' age of the fastest-growing companies in the United States is 45 years old. The number of observations that has non-missing founding age values is lower because we do not ascribe a birth year or founding age to VS founders who are not uniquely matched into Infutor. In other words, we do not take a position on which Infutor individual's birth year to use when a VS founder is matched to multiple Infutor individuals, and we therefore ascribe a missing founding age value to these VS founders.

<sup>16</sup> Wadhwa et al. (2010) present the proportions of immigrant-founded companies that specialize in technological and/or engineering endeavors.



**Table 1B**

Industry breakdown by immigration status.

The table compares number and proportion of startups in each industry across native-born and immigrant founders. Z-statistics from tests for differences across population proportions are presented in the final column. Each observation is a founder-startup pair. Asterisks denote statistical significance level from t-tests on differences in sample means, \* is for 10 %, \*\* is for 5 %, and \*\*\* is for 1 % level.

	(1)	(2)	(3)	(4)	(5)	(6)
	Natives		Immigrants		Natives – Immigrants	
	N	%	N	%	Difference	Z-statistic
Business and Financial Services	6586	22.49 %	1285	18.22 %	0.043***	8.21
Consumer Goods	946	3.23 %	133	1.89 %	0.013***	7.02
Consumer Services	5318	18.16 %	865	12.26 %	0.0591***	13.08
Energy	413	1.41 %	85	1.21 %	0.002	1.36
Health	4990	17.04 %	1117	15.84 %	0.012*	2.46
Industrials	600	2.05 %	146	2.07 %	−0.0002	−0.13
IT	10,413	35.56 %	3421	48.50 %	−0.129***	−19.69
Unassigned	18	0.06 %	1	0.01 %	0.0005	2.33
Total	29,284	100 %	7053	100 %	–	–

**Table 2**

Immigrant founder count and share by state.

The bottom panel presents the top and bottom ten states with the highest and lowest number of immigrant founder-startup pairs. The bottom panel presents the top and bottom ten states with the highest and lowest immigrant founder-startup pair shares. Shares are calculated as the proportion of immigrant founder-startup.

Top 10 states by count		Bottom 10 states by count	
State	Count	State	Count
CA	4007	IA	4
MA	648	LA	3
NY	544	ID	1
TX	225	MS	1
WA	195	MT	1
PA	159	VT	1
NJ	127	WY	1
IL	114	ND	0
FL	108	SD	0
VA	104	WV	0

Top 10 states by share		Bottom 10 states by share	
State	Share	State	Share
DE	29.4 %	ND	0.0 %
CA	26.2 %	SD	0.0 %
NJ	25.8 %	WV	0.0 %
MA	19.2 %	ID	2.5 %
OK	17.3 %	VT	2.5 %
FL	17.0 %	MT	4.2 %
MD	17.0 %	LA	5.6 %
CT	16.2 %	SC	5.7 %
NY	15.7 %	KS	6.7 %
WI	15.2 %	KY	7.1 %

Slaughter, 2017).

Table 2 shows the top ten and bottom ten states by number and proportion of immigrant founder-startup pairs.<sup>17</sup> To construct the top panel, we use data on each startup's headquarter office address and count the number of immigrant founder-startup pairs. The three states with the greatest number of VC-backed startups in the sample are

<sup>17</sup> Immigrant founder-startup pairs are observations at the founder level, i.e., if a company has two founders, then there will be two founder-startup pairs in these analyses. Similarly, if a founder starts two companies in our sample, they will be in the founder-startup pair analysis for both companies. Adjusting our analyses to be exclusively conducted at the founder or startup level does not materially change our key results.

California, Massachusetts, and New York, and these states are also the three states with the highest number of immigrant founder-startup pairs. In general, states that appear among the top ten in number of immigrant founder-startup pairs tend to be coastal states and states that feature prominently in the US VC ecosystem. On the other hand, states that appear among the bottom ten tend to be smaller states and states in the south or middle of the country.

The top panel of Table 2, which counts founder-startup pairs, is highly influenced by many factors including a state's population size, economic affluence, and level of venture capital activity. The bottom panel of Table 2 effectively controls for some of these factors by tabulating the top ten and bottom ten states with the highest immigrant-founded fraction of total VC-backed startups. Within the top ten, California, Massachusetts, and New York remain on the list, indicating that the largest VC hubs have not only a high number of immigrant-founded VC-backed companies but also a high share of immigrant-founded startups relative to the total number of startups.<sup>18</sup> Furthermore, similar to the results in the top panel, states with the lowest share of immigrant-founded startups are in the south and middle of the US.<sup>19</sup> These results suggest that venture capital hubs on the coasts, especially California, Massachusetts, and New York, appear to draw the largest benefits from high-growth immigrant-founded companies.

To further contextualize the statistics presented in Table 1B, Appendix Table B.1 presents each state's total immigrant population and immigrant population share in 2018. The states are sorted by immigrant population share, from largest to smallest. Comparing the state rankings that we present in Table 2 to those presented in Table B.1 suggests that the distribution of VC-backed immigrant entrepreneurs across states is highly influenced by each state's immigrant share of population. For example, six states appear in both top ten lists and five states appear in both bottom ten lists across the two tables. Also, states that appear in the top and bottom ten lists in Table 2 are generally near, if not in, respectively, the top and bottom ten lists in Table B.1.

### 3.2. Immigrant founder share time trends

In this section we consider how the aggregate national immigrant share of VC-backed startups and founders has evolved over time. Fig. 1 plots the share of immigrant founder-startup pairs by 5-year cohorts from 1990 to 2019. From the 1990–1994 cohort to the 2000–2004 cohort, the share of immigrant founder-startup pairs increased from

<sup>18</sup> Delaware likely appears as a top ten state in the bottom panel because the data contain a few company registration addresses as opposed to physical company headquarter addresses.

<sup>19</sup> Results are similar if we instead use founders as the unit of observation.

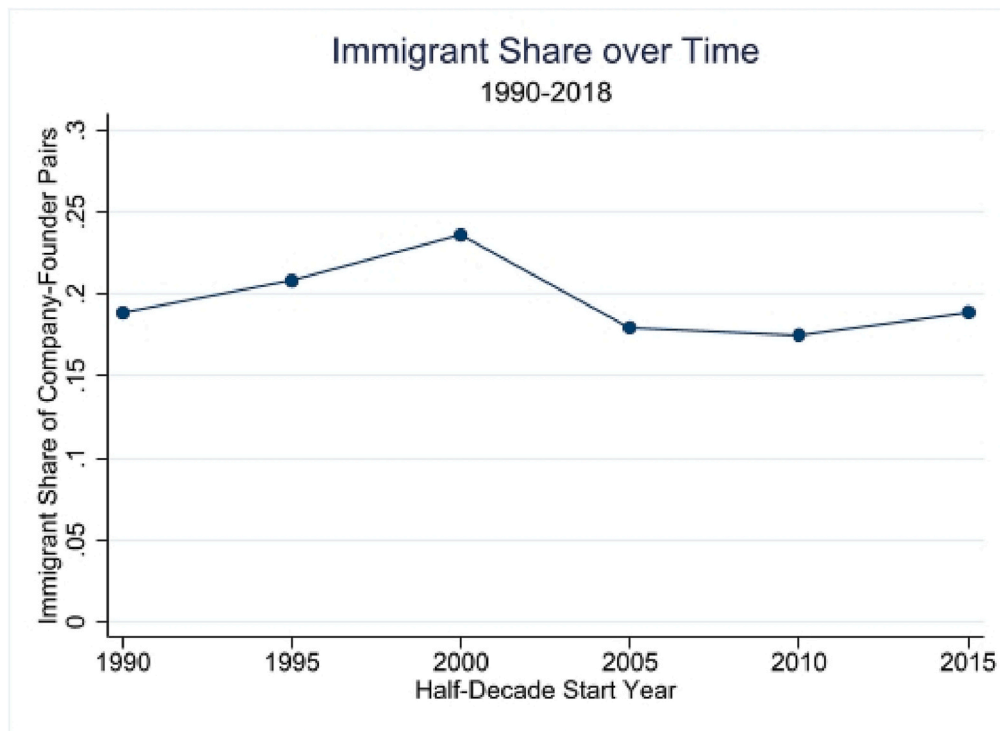


Fig. 1. Immigrant founder share over time.

The figure plots the share of immigrant founders over time. Shares are calculated from all founder-startup pairs in each 5-year cohort.

around 20 % to a peak of nearly 25 %. This pattern is broadly consistent with the findings of [Kerr and Kerr \(2016\)](#), who find that the proportion of immigrant founders among VC-backed startups rose from 1995 until reaching a peak during the tech bubble. After this period, we find that the share of immigrant founders dropped to around 17 % between 2010 and 2014, before starting to rise again in the most recent 2015–2019 cohort.

While much debate about the effects of H-1B Visa Reform Act of 2004 has received much media attention,<sup>20</sup> this reform is unlikely to have affected the trends we identify. More likely, factors such as (i) the impact of various late 1990s crises (e.g., the 1997 Asian financial crisis, the dot-com bubble) on immigrants' already restrictive financial constraints and (ii) the 2000s and 2010s growth in non-US VC-backed entrepreneurship more plausibly account for these trends. First, the average age at which immigrant founders received their social security number is close to 26 years old. This age is a reasonably reliable proxy for these immigrants' age at entry into the United States. The average age at founding of a VC-backed company for immigrants is >40 years old, indicating that the majority of founders, even in the latter part of the sample, likely immigrated to the United States before 2004. Therefore, it is unlikely that the H-1B Visa Reform Act of 2004 is the primary driver of the change in the immigrant founder share during our sample. However, as we discuss in [Section 4](#), we would expect the law's effects on high-potential entrepreneurship to more clearly manifest in the coming years.

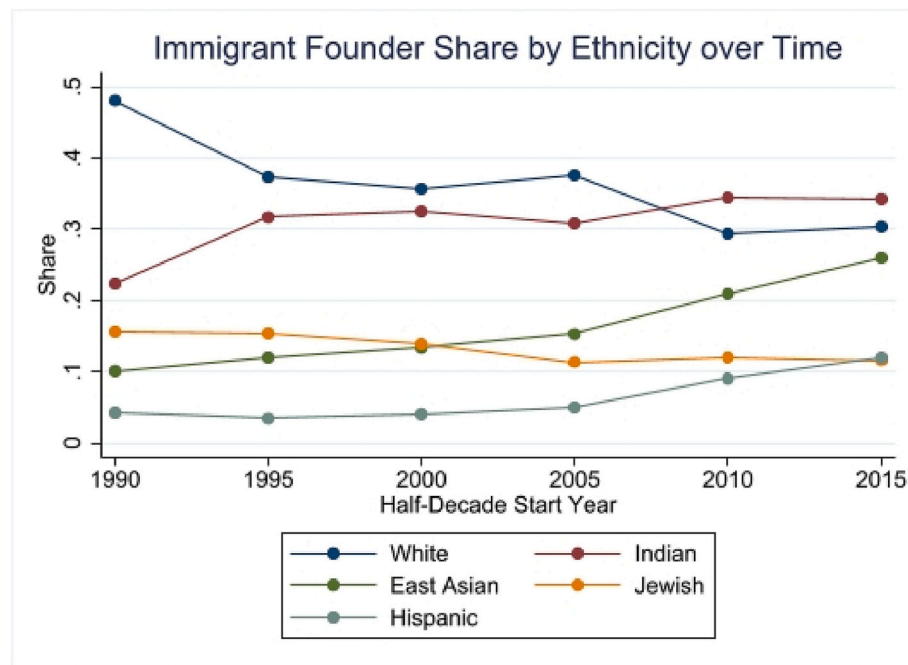
Next, decomposing the aggregate trends shown in [Fig. 1](#), [Fig. 2](#) plots the share of each ethnic group among immigrant founders by 5-year cohorts from 1990 to 2019. Ethnic composition of immigrant entrepreneurs has changed substantially over time. First, the share of white immigrant founders has decreased from nearly 50 % to roughly 30 %

over the 1990–2019 period. Likewise, the share of ethnically Jewish immigrant founders decreased from around 15 % to slightly >10 %. In contrast, the share of Indians, East Asians, and Hispanics among immigrant founders significantly expanded over the same timeframe. The share of Indian immigrant founders increased from 20 % to >30 %. The share of East Asian immigrant founders rose from around 10 % to >25 %. The share of Hispanic immigrant founders increased from <5 % to >10 %. The rise of Indian and East Asian entrepreneurs is in line with results from prior work on trends in immigrant entrepreneurship ([Wadhwa et al., 2007a, 2007b](#)). It also mirrors the trends in US college and graduate school admissions as the share of students from India, China, and South Korea enrolling in US universities rose substantially over this period ([Bound et al., 2021](#)). However, identification of the share of Hispanic immigrant entrepreneurs is a new result that, to the best of our knowledge, has not been directly examined in prior academic research. At first glance, the increasing share of Hispanic immigrant entrepreneurs that we find stands in contrast to aggregate trends and estimates of Hispanic immigration. The aggregate share of Hispanic immigrants in the US population has been trending downwards for a sustained period ([Budiman et al., 2020](#)).

### 3.3. Founder and startup characteristics by immigration path

Our primary contribution is to identify and impute immigrant entrepreneurs' primary immigration pathways into the US: college education, postgraduate education, and work. First, the immigrant founders in our sample have all attained at least an undergraduate degree. This level of educational attainment already differentiates them from the overall population of immigrants in the US, only 32 % of which possessed a bachelor's degree in 2018 ([Budiman et al., 2020](#)). Nonetheless, the remaining heterogeneity in immigrant founders' educational

<sup>20</sup> The H-1B Visa Reform Act of 2004 reduced the US's annual H-1B visa cap from 195,000 to 65,000 visas. H-1B visas represent the primary channel through which high-skilled immigrants are legally authorized to work in technically-oriented jobs in the US.



**Fig. 2.** Immigrant founder share by ethnicity over time.

The figure plots immigrant founders' ethnicity breakdown over time. Ethnic groups' shares of the immigrant founder subsample are calculated from all founder-startup pairs in each 5-year cohort.

**Table 3A**

Founder characteristics by immigration path.

The table presents summary statistics for immigrant founders' characteristics by immigration path. Each observation is a founder. Undergraduate (UG) immigrant founders are those who came to the United States for undergraduate studies. Graduate (GR) immigrant founders are those who came to the United States for graduate studies. Post Degree (PD) immigrant founders are those who came to the United States for work. IPO Rate is the percentage of the founder's startups that had gone public by 2019. Success Rate is the percentage of the founder's startups that, by 2019, either reached the IPO stage or were acquired for more than the total amount of money invested, adjusted for inflation. Number of Firms counts the number of VC-backed ventures that each founder had started throughout the sample. Patent Rate is the percentage of the founder's startups that successfully filed at least one patent within two years of founding. Patent Count is the average number of patents filed by the founder's startups within two years of founding. Founding Age is the average of the founder's age at the time that each of his startup was formed. Entry Age is the founder's age when he received his social security number.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Natives		UG immigrants		GR immigrants		PD immigrants	
	N	Mean	N	Mean	N	Mean	N	Mean
IPO Rate	26,439	0.04	2391	0.04	2346	0.05	1462	0.06
Entry Age	15,916	9.58	1225	22.23	1759	24.04	1094	28.96
Founding Age	16,115	39.46	1235	45.21	1771	41.78	1098	45.50
Female	26,439	0.09	2391	0.12	2346	0.09	1462	0.09
Jewish	26,439	0.19	2391	0.15	2346	0.10	1462	0.13
East Asian	26,439	0.05	2391	0.19	2346	0.20	1462	0.12
Indian	26,439	0.04	2391	0.20	2346	0.46	1462	0.28
Hispanic	26,439	0.04	2391	0.08	2346	0.05	1462	0.07
White	26,439	0.70	2391	0.42	2346	0.22	1462	0.43
Success Rate	26,439	0.15	2391	0.15	2346	0.19	1462	0.19
# of Firms	26,439	1.11	2391	1.12	2346	1.17	1462	1.13
Patent Rate	26,439	0.14	2391	0.13	2346	0.21	1462	0.17
Patent Count	26,439	0.47	2391	0.49	2346	0.94	1462	0.64
Cite-W Patent Count	26,439	14.33	2391	11.81	2346	23.98	1462	17.67

profiles (e.g., the extent of their education in the US) could directly identify just how critical a role US universities play in attracting high-quality, distinctly productive foreign-born entrepreneurial talent. Since the identification of immigration pathways is new to the literature, we present summary statistics on each pathway-based group of immigrant founders introduced in Section 2.2.1. Table 3A presents summary statistics of founder and startup characteristics by immigration pathway. The majority of immigrant founders came to the United States first to pursue some type of university degree. Thirty-nine percent of immigrant

founders came to the United States first for undergraduate studies (Undergraduate immigrants), and 38 % came first for postgraduate studies (Graduate immigrants). Only 24 % of immigrant founders received their education entirely abroad before coming to the US for employment (Post Degree immigrants). These statistics show that universities serve as the primary gatekeepers for foreign entrepreneurial talent relative to companies. In turn, the statistics suggest that student visa policies and universities' admission policies likely play a substantial role in determining the quality of entrepreneurial talent in the United

States, even if work visa policy currently draws more policy debate.

The ethnic composition of Undergraduate, Graduate, and Post Degree immigrant founders shows considerable heterogeneity. A relatively higher share of Graduate (46 %) and Post Degree (28 %) immigrant founders are Indian, while a lower share of Undergraduate founders (20 %) are Indian.<sup>21</sup> A higher share of Undergraduate (19 %) and Graduate (20 %) founders are East Asian, whereas a lower share of Post Degree founders (12 %) are.<sup>22</sup> These statistics are consistent with trends in ethnic composition of foreign undergraduate and graduate students in the United States (Bound et al., 2021). We also find that ethnically white immigrants make up a large proportion of Undergraduate (42 %) and Post Degree (43 %) founders, relative to their share of Graduate founders.<sup>23</sup> In other words, white immigrant founders are likely to first enter the US for undergraduate education or for work.

Using data on birth dates and the years in which founders received their SSNs, we calculate founders' ages when they received their SSNs. These ages should be close to the ages at which the founders immigrated. The average age at which immigrants received their SSN increases monotonically as we move from Undergraduate to Post Degree immigrants as would be expected given group definitions. *t*-Tests verify that differences in average entry age across all three groups are statistically significant in all pairwise permutations with respect to (i) the other immigrant groups and (ii) the native subsample. Additionally, there is a noticeable gap between average SSN age and average founding age across all three immigrant founder groups, indicating that immigrant founders generally stay and work in the United States for many years before starting their firms. This delay between immigrating and founding suggests that changes in work visa policies (e.g., H-1B caps in the US context) are unlikely to have an immediate effect on the rate of immigrant entrepreneurship or the number of immigrant-founded companies. Over time, however, changes in H-1B quotas may have strong lagged effects.

Productivity, as measured by average number of companies started and their success rate, appears especially high for Graduate and Post Degree immigrants, suggesting that immigration pathways may inform and predict entrepreneurial skill/quality. Specifically, Graduate and Post Degree immigrant founders enjoy statistically significant higher success rates when compared to native and Undergraduate immigrant founders, while Graduate immigrant founders also start more companies than native and Undergraduate immigrant founders to a statistically significant degree. Innovative startup activities, as measured by the patenting activity share ("patent rate") and patent count associated with founded companies, is highest (to a statistically significant degree) among Graduate immigrant founders. This finding suggests that universities' graduate programs indirectly contribute to the US's innovation capacity by attracting individuals who are very capable in starting high-potential companies that produce patents.

We also examine differences in educational backgrounds for immigrant founders by path of entry. Table 3B summarizes founders' education information across immigration status and path. A higher share of immigrant founders' majors in STEM fields, and a lower share majors in business-related fields when compared to native-born founders. The largest difference appears when we compare Graduate and Post Degree immigrant founders to native-born founders. We find that 87.8 % of Graduate immigrant founders and 77.0 % of Post Degree immigrant

founders have an undergraduate STEM degree compared to 64.4 % for native-born founders. Differences in STEM major share across all three mentioned groups are statistically significant. In addition, higher share of immigrant founders holds some type of graduate degree when compared to native-born founders. Additionally, a higher share of immigrant founders holds a technical post-graduate degree compared to native-born founders. However, Undergraduate immigrant founders (i.e., immigrant founders who come to the US first for their undergraduate education), appear similar to (i.e. different to a statistically insignificant degree from) native-born founders for the first measure (all other pairwise differences between groups are statistically significant). Approximately 80 % of Graduate immigrant founders have a STEM masters or Ph.D. degree. Similar shares of native-born, Undergraduate, and Graduate founders earn an MBA (approximately 20 %), while a lower share (statistically significant) of Post Degree immigrant founders (8.3 %) earn MBAs.

The bottom panel of Table 3B focuses on the likelihood that a founder received at least one degree (undergraduate or post-graduate) from a top school.<sup>24</sup> 32 % of immigrant founders hold a top school degree while 35 % of native-born founders hold a top school degree. This difference is statistically significant, indicating that on average immigrants are less likely to attend a top school. These averages, however, mask substantial heterogeneity across groups and for various types of degrees. For top undergraduate colleges, a higher share of Undergraduate immigrant founders, to a statistically and economically significant extent, hold a degree from a top college when compared to native-born founders (30.9 % vs. 21.9 %). For top graduate schools, higher shares of Undergraduate and Graduate immigrant founders, to a statistically and economically significant extent, hold a degree from a top university (26.4 % and 37.9 %) compared to native-born founders (23.0 %). Analogous conclusions are true for both MBA and non-MBA post-graduate degrees. These findings demonstrate that, conditional on receiving their degree from a university in the US, immigrant founders are more likely to attend a top university. This result is consistent with the idea that the population of US-educated immigrant founders is likely to be drawn from the right tail of the non-US and global academic talent distribution (Borjas, 1989; Borjas, 1991; Borjas and Bratsberg, 1996; Rosenzweig et al., 2006).

These summaries of educational attainment across our various founder groups provide several important takeaways. First, Undergraduate immigrant founders tend to appear more similar to native-born founders in terms of educational backgrounds and startup characteristics. Graduate and Post Degree immigrant founders, however, appear quite different from native-born founders along these dimensions. Second, immigrant founders are more likely to hold advanced degrees compared to native-born founders. Finally, across all three groups, immigrant founders are more likely to major in STEM fields than are native-born founders to a statistically significant degree. Altogether, the findings suggest that education-based immigration pathways could help to explain some of the heterogeneity in the quality of immigrant-founded startups. In addition, as discussed in Section 3.4, observing the pathways that immigrant founders take to arrive in the United States further enables us to evaluate the prominence of universities vs. companies as recruiters of foreign-born talent.

Finally, the breakdown of educational backgrounds across immigrant groups also foreshadows an intuitively predictable breakdown of startups' industry composition by founder group/type. Table 3C presents the industry breakdowns for native-founded and immigrant founded

<sup>21</sup> *t*-Tests of subsample differences in means verify that differences in the Indian share of Undergraduate vs. Graduate, as well as Undergraduate vs. Post Degree immigrant founders are statistically significant.

<sup>22</sup> *t*-Tests of subsample differences in means verify that differences in the East Asian share of Undergraduate vs. Post Degree, as well as Graduate vs. Post Degree immigrant founders are statistically significant.

<sup>23</sup> *t*-Tests of subsample differences in means verify that differences in the White share of Undergraduate vs. Graduate, as well as Post Degree vs. Graduate immigrant founders are statistically significant.

<sup>24</sup> Following the definition in Gompers et al. (2016), we define top universities to include Ivy League schools, Amherst College, California Institute of Technology, Duke University, MIT, Northwestern University, Stanford University, University of California (Berkeley), University of Chicago, Williams College, Cambridge University, INSEAD, London School of Economics, London Business School, and Oxford University.



**Table 3B**

Education information by immigration status and path.

The table presents education information by immigration status and path. Each observation is a founder. Undergraduate (UG) immigrant founders are those who came to the United States for undergraduate studies. Graduate (GR) immigrant founders are those who came to the United States for graduate studies. Post Degree (PD) immigrant founders are those who came to the United States for work. The summary statistics below are computed from a series of indicator variables that are defined as follows. STEM equals one if the founder holds a STEM undergraduate degree. Business equals one if the founder holds a business undergraduate degree. The number of observations is smaller because major information is not available for all founders. All founders that have a graduate degree have at least one undergraduate degree. Any Graduate Education equals one if the founder holds at least one graduate degree. This category includes professional degrees such as M.D. and J.D. MBA equals one if the founder holds an MBA degree. Non-MBA equals one if the founder holds any non-MBA graduate degree. STEM Master's equals one if the founder holds a STEM master's degree. Ph.D. equals one if the founder holds a doctoral degree. Top school is defined in the same as in [Gompers et al. \(2016\)](#). Any Top School equals one if the founder holds a college or graduate degree from a top school. College equals one if the founder holds a college degree from a top school. Any Top Graduate Education equals one if the founder holds at least one graduate degree from a top school. MBA equals one if the founder holds an MBA degree from a top school. Non-MBA equals one if the founder holds a non-MBA graduate degree from a top school.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Natives		UG immigrants		GR immigrants		PD immigrants	
	N	mean	N	mean	N	mean	N	mean
STEM Major	19,451	0.64	1883	0.72	1252	0.88	806	0.77
Business Major	19,451	0.28	1883	0.26	1252	0.05	806	0.13
Any Graduate Education	26,439	0.55	2391	0.55	2346	1.00	1462	0.57
MBA	26,439	0.20	2391	0.19	2346	0.24	1462	0.08
Non-MBA Graduate School	26,439	0.39	2391	0.41	2346	0.90	1462	0.51
STEM Master's	26,439	0.16	2391	0.22	2346	0.41	1462	0.23
PhD	26,439	0.12	2391	0.12	2346	0.39	1462	0.18
Any Top School	26,439	0.35	2391	0.42	2346	0.38	1462	0.04
Top College	26,439	0.22	2391	0.31	2346	0.01	1462	0.01
Top MBA	26,439	0.10	2391	0.11	2346	0.13	1462	0.01
Top Graduate School (non-MBA)	26,439	0.15	2391	0.18	2346	0.27	1462	0.02
Any Top Graduate Education	26,439	0.23	2391	0.26	2346	0.38	1462	0.03

**Table 3C**

Industry breakdown by immigration path.

The table presents startup industry proportions by immigration status and path. Each observation is a founder-startup pair. Undergraduate immigrant (UG Imm.) founders are those who came to the United States for undergraduate studies. Graduate immigrant (GR Imm.) founders are those who came to the United States for graduate studies. Post Degree immigrant (PD Imm.) founders are those who came to the United States for work.

	(1)	(2)	(3)	(4)
	Natives	UG Imm.	GR Imm.	PD Imm.
Business and Financial Services	22.49 %	21.69 %	15.40 %	17.27 %
Consumer Goods	3.23 %	2.70 %	1.21 %	1.70 %
Consumer Services	18.16 %	17.91 %	7.83 %	10.48 %
Energy	1.41 %	1.12 %	1.54 %	0.79 %
Health	17.04 %	14.42 %	16.46 %	17.09 %
Industrials	2.05 %	2.17 %	2.08 %	1.88 %
IT	35.56 %	39.98 %	55.49 %	50.73 %
Unassigned	0.06 %	0.00 %	0.00 %	0.06 %

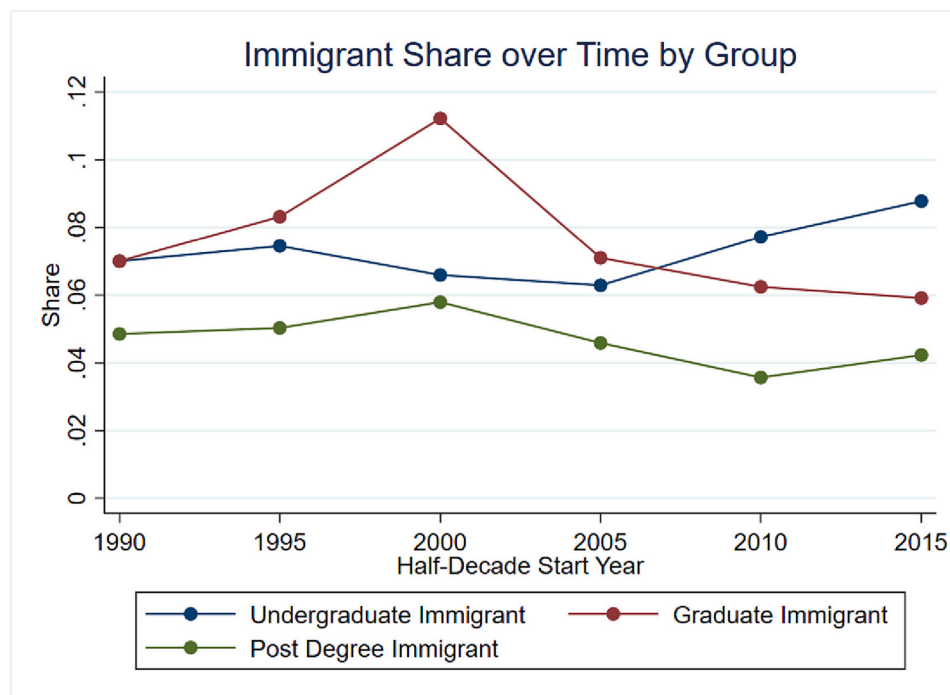
companies (across the three immigrant pathway groups). First, native-born founders and Undergraduate immigrant founders generally appear to start companies in similar proportions by industry. IT companies represent 35–40 % of their startups, Business and Finance 22 %, and Consumer Services roughly 18 %. This observation is consistent with the fact that native-born founders and Undergraduate immigrant founders have similar educational backgrounds. Second, Graduate and Post Degree immigrant founders are more likely to start companies in the IT sector (55.5 % and 50.7 %) than native-born founders (35.5 %) and are less likely to found Business and Finance (15.4 % and 17.3 %) or Consumer Services (7.8 % and 10.5 %) companies than are native-born founders (22.0 % and 18.0 % respectively). This fact suggests that Graduate and Post Degree immigrant founders, who are much more likely to have completed STEM undergraduate and post-graduate education, tend to establish more technologically focused and STEM-oriented companies.

### 3.4. Immigration paths time trends

In this section we explore temporal trends in the share of VC-backed immigrant founders belonging to each pathway-based immigrant group (Undergraduate, Graduate, or Post Degree). [Fig. 3](#) plots the breakdown of immigrant founder-startup pairs by immigration path over time, from 1990 to 2019. The figure tabulates immigrant founders by the year they started their companies, not the year of entry into the US. The proportion of Undergraduate immigrant founders was stable in the early part of the sample but increased over the past decade. On the other hand, the proportion of Graduate and Post Degree immigrant founders peaked in the 5-year period between 2000 and 2004 and has declined since then.

The time trends described above are informative of underlying trends in immigrant founders. First, the decline in the share of immigrant founders shown in [Fig. 1](#) is primarily driven by the decline in Graduate immigrant founders who initially arrived in the United States for graduate education. Compared to Post Degree immigrant founders, Graduate immigrant founders make up a much larger proportion of the sample and their relative decline from the peak period has also been larger.<sup>25</sup> This change drives the overall decline in immigrant founder shares. Second, the proportion of Undergraduate immigrant founders who initially arrived in the US for their undergraduate education has been increasing steadily over the sample period. The time trends point to the growing importance of undergraduate education as the primary channel for foreign entrepreneurial talent to enter the US. The share of foreign students in US universities saw a dramatic jump in the late 1970s

<sup>25</sup> Various explanations could account for this persistent decline. For example, the dot-com crash of 2000–2001 could have especially dissuaded Graduate immigrants from pursuing what appear to be “risky” I.T. entrepreneurship, since these immigrants may intuitively appear more financially constrained than their native and Undergraduate counterparts. Alternatively, the expanding global share of non-US VC funding during the 2000–2018 period, combined with potential improvement in graduate education abroad, may have incentivized potential Graduate immigrants against migrating to the US. Nonetheless, precisely disentangling the specific mechanisms behind this decline is beyond the scope of our paper, and we leave further investigation of the decline in the Graduate immigrant share of US VC-backed founders to future research.



**Fig. 3.** Immigrant founder share by immigration path.

The figure plots immigrant founders' immigration path breakdown over time. Immigrant groups' shares of total VC-backed founders are calculated from all founder-startup pairs in each 5-year cohorts. Undergraduate immigrant founders are those who came to the United States for undergraduate studies. Graduate immigrant founders are those who came to the United States for graduate studies. Post Degree immigrant founders are those who came to the United States for work. Number of immigrant founder-startup pairs in each immigrant group is scaled by the total founder-startup pairs in each 5-year cohort.

and again in the 2010s (Israel and Batalova, 2021). These trends have begun to reverse as the change in new foreign student enrollment in US universities turned negative in 2016–2019. While this reversal is unlikely to affect the supply of VC-backed immigrant founders in the US economy in the short run, the long-term economic implications may cause more justifiable concern for US policymakers.<sup>26</sup>

#### 4. Empirical analyses and results

The summary statistics from the previous sections provide suggestive evidence that (i) VC-backed immigrant founders make exceptional and distinctive contributions to VC-backed entrepreneurship and (ii) US universities are important contributors to the quantity and quality of foreign entrepreneurial talent. Nonetheless, in this section we explore material linkages/connections (as opposed to completely spurious correlations) between US university education and the activity of immigrant and native-founded VC-backed startups. Accordingly, we present a set of statistical analyses which more rigorously show that universities play a crucial role in adding to the quantity and quality of VC-backed entrepreneurship in the United States. Furthermore, we demonstrate that university location plays a significant role in shaping the geographic distribution of immigrant (and native) founded startups. Throughout this section, for ease of coefficient interpretation, we use linear probability (i.e., OLS) regression models for analyses where the outcome variable is an indicator variable. In the appendix, we show that all

<sup>26</sup> Given differences in the industries that immigrants in each of the groups start firms in, one alternative explanation is that changes in the industrial composition of firms funded by venture capitalists over time may account for the increasing share of Undergraduate immigrants and decreasing share of Post Degree immigrants over time. Figure D.3 in the appendix suggests this is likely not to be the case. The figure plots industry composition-implied immigrant founder group share over time. Per-period industry-implied founder group shares are calculated as the product of the full-sample industry-group shares (e. g., share of Undergraduate immigrant founder-startup pairs in the IT industry) and the per-period industry shares (share of IT founder-startup pairs). The plot shows that industry-implied group shares are relatively constant over time.

results are qualitatively similar when we use probit regressions. For cases where the outcome variable is a count variable, we use Poisson regressions (Cohn et al., 2021).

##### 4.1. Universities and local entrepreneurship activity

We begin our formal empirical analysis by exploring how universities attract and retain entrepreneurial talent locally. In other words, we ask – do universities contribute to the quantity of local/state-level entrepreneurship activity? Table 4 presents descriptive statistics on the probability that a founder who is educated in a given state also starts a VC-backed company in the same state. The top panel presents the relevant statistics for all states. The first row presents the share of founders who started at least one VC-backed company in the state that they received their highest postsecondary education degree. Across the three relevant groups (natives, Undergraduate immigrants, and Graduate immigrants), upon completion of their education, more than one-third of founders did not migrate to another state to start their high-potential venture. This headline number is in line with the finding that 45 % of foreign students in the US extend their visas to work in the same metropolitan area as their alma maters (Ruiz, 2014). The similarity may suggest that universities attract a large proportion of foreign-born would-be entrepreneurs to remain in the same state from graduation to founding.

The proportion of non-migrant (“same state”) founders is high when we compare it to the share of same state founders implied by “overall random motion” (ORM). To calculate the ORM share of same state founders, we first assume a counterfactual world in which migration patterns are uncorrelated with education state. In other words, newly graduated founders are (i) not “sticky” to their education state when deciding where to establish a startup and (ii) just as likely as any other founder to establish a startup in a given state. Hence, the probability that a founder migrates to (or stays in) a given state is effectively assumed to equal that state's overall share of total US VC-backed startups in the entire sample (by definition between 0 and 1). Given this assumption, we compute the counterfactual probability of non-migration for every founder in our final sample. We average these probabilities within each group to find the associated ORM-implied same state founder share. The

**Table 4**

Education and startup location.

The table presents statistics for the propensity for founders to start VC-backed companies in the same state where they received their final postsecondary education. The first panel presents the statistics for all states and the remaining panels present the statistics for the group of states as marked in the State column. Actual Same State Founder Share is the percentage of founders in each group that started at least one company in the state that they received their final postsecondary education. Overall Random Motion (ORM) Implied Share is the share of same state founders implied by no geographic stickiness with respect to state of education. To calculate this counterfactual share, we first assume that every individual is just as likely to establish a startup in a given state as any other founder, for all states. This assumption implies the probability that a founder is a same state founder to be the share of total startups founded in the founder's state of education. The ORM-Implied Same State Share is accordingly the average of these probabilities across all individuals within the specified education state and group. Actual - ORM t-statistic is the t-statistic from the *t*-test that tests the statistical difference between the group's Actual Same State Founder Share and the imputed ORM Implied Share. GRM Implied Share is the share of same state founders implied by no geographic stickiness with respect to state of education, conditional on group status. Calculation of this share is analogous to calculating the ORM-Implied Share, except that a founder's probability of non-migration is calculated to be the share of total startups founded by *same-group* (e.g., *native*, *Undergraduate Immigrant/UG Imm.*, or *Graduate Immigrant/GR Imm.*) founders in the relevant state of education. Actual - GRM t-statistic is the t-statistic from the *t*-test that tests the statistical difference between the group's Actual Same State Founder Share and the GRM Implied Share.

State	Statistics	Natives	UG Imm.	GR Imm.
All	Actual Same State Fdr Share	36.5 %	36.9 %	32.7 %
	ORM Implied Same State Fdr Share	11.4 %	13.4 %	13.7 %
	Actual - ORM t-statistic	88.9	27.4	23.8
	GRM Implied Same State Fdr Share	10.9 %	15.6 %	17.6 %
	Actual - GRM t-statistic	90.7	25.4	20.0
CA	Actual Same State Fdr Share	74.1 %	82.4 %	83.2 %
	ORM Implied Same State Fdr Share	42.0 %	42.0 %	42.0 %
	Actual - ORM t-statistic	51.3	25.4	26.9
	GRM Implied Same State Fdr Share	38.4 %	51.6 %	60.1 %
	Actual - GRM t-statistic	57.0	19.4	15.1
MA	Actual Same State Fdr Share	31.4 %	27.3 %	30.7 %
	ORM Implied Same State Fdr Share	9.3 %	9.3 %	9.3 %
	Actual - ORM t-statistic	28.5	7.8	7.7
	GRM Implied Same State Fdr Share	9.3 %	8.3 %	9.9 %
	Actual - GRM t-statistic	28.5	8.2	7.4
NY	Actual Same State Fdr Share	31.4 %	31.3 %	16.4 %
	ORM Implied Same State Fdr Share	9.5 %	9.5 %	9.5 %
	Actual - ORM t-statistic	22.6	7.4	2.8
	GRM Implied Same State Fdr Share	10.0 %	11.2 %	4.6 %
	Actual - GRM t-statistic	22.2	6.8	4.9
Other	Actual Same State Fdr Share	25.6 %	19.2 %	11.8 %
	ORM Implied Same State Fdr Share	1.7 %	1.8 %	1.8 %
	Actual - ORM t-statistic	65.1	15.3	10.9
	GRM Implied Same State Fdr Share	1.9 %	1.4 %	1.3 %
	Actual - GRM t-statistic	64.8	15.6	11.4

ORM-implied shares are 11 %, 13 %, and 13 % for natives, Undergraduate immigrants, and Graduate immigrants, respectively.<sup>27</sup> The actual shares are all statistically different from the ORM-implied shares as shown by the t-statistics that we present in the third row.<sup>28</sup> As a robustness check, we perform the same statistical test against the group random motion (GRM) same state share, which is the group-specific version of ORM. That is, computation of the GRM-implied same state share is identical to computation of the ORM-implied share with one difference. Specifically, under GRM, we assume that a newly graduate founder is just as likely as any other founder *within his or her group* (i.e., native, Undergraduate immigrant, or Graduate immigrant) to establish a startup in any given state when calculating non-migration probabilities for each founder. Under GRM, we find the same qualitative results.

<sup>27</sup> Differences in the ORM-implied share across each group (Native, Undergraduate, and Graduate) stem from differences in the distribution of founders' education states across the three groups. For example, a relatively higher share of Undergraduate and Graduate immigrants are educated in states that host many VC-backed startups (e.g., CA, MA), and the ORM probability of non-migration for *all* founders educated in these states is, by construction, higher. This compositional difference causes the ORM-implied share of same state founders to be higher for Undergraduate and Graduate immigrants than for natives.

<sup>28</sup> The t-statistics shown in the table summarize a *t*-test of whether the average of the Same State founder indicator variable is equal to the ORM implied Same State share value within each of the three relevant groups. The ORM implied share is constant within each of the three relevant groups (i.e., natives, Undergraduate immigrants, and Graduate immigrants).

Overall, the top panel of Table 4 suggests that universities contribute to their local economies by attracting students who will eventually start high-potential companies in the same/nearby locales.

The remainder of Table 4 presents the results of the analogous analyses that occur when one restricts the sample to founders educated in California, Massachusetts, New York, and all remaining ("Other") states, respectively. The goal of this exercise is to explore whether the magnitude of geographic stickiness towards state of education (exhibited in Table 4's top panel) might differ between hub and non-hub states. We consider the first three states to be venture capital hubs because they contain the highest concentration of venture capital investment activity in the country. Even though the same state founder share is substantially higher among hub states, especially for California, we find that the same qualitative results regarding geographic stickiness hold across hub states and non-hub states. In other words, since founders across the entire country are disproportionately likely to remain in their state of education when founding a company, research universities (not just leading research universities) play a critical role in attracting high-potential VC-backed entrepreneurs to their local economies, even in non-hub states.

Hence, regardless of immigration status, would-be founders tend to start VC-backed companies in the state where they completed their final postsecondary degree relative to ORM and GRM benchmarks. To complement this finding, we more rigorously test the differences in the propensity to start high-potential companies in their state of education across Natives, Undergraduate immigrants, and Graduate immigrants. We perform this analysis by running variants of the following OLS regression:

**Table 5**

Same state founding regression results.

This table presents OLS regression results where the same state founder indicator variable is regressed onto the founders' immigration status indicator variables. Each observation is a founder-startup pair. The dependent variable equals one if the founder started his company in the same state that he received his final postsecondary education degree. Immigrant equals one if the founder is an immigrant. In-State Native equals one if the founder is a native founder who received his SSN in the same state that he received his final postsecondary education degree. UG Imm. equals one for immigrants who came to the United States for college. GR Imm. equals one for immigrants who came to the United States for a graduate degree. Post Degree (PD) immigrants are excluded from the sample. The reference group is composed of out-of-state native founders. Demographic control variables include indicator variables for Female, Jewish, East Asian, Indian, and Hispanic founders. Refer to Appendix C for control variable definitions. Heteroskedasticity-robust standard errors are reported in parentheses and clustered at the founder level.

	(1)	(2)	(3)	(4)	(5)	(6)
	All states	All states	Hub states	Hub states	Non-hub states	Non-hub states
Immigrant	0.0029 -0.0082		0.0365*** -0.0127		-0.0257*** -0.0099	
UG Imm.		0.0201** -0.0093		0.0420*** -0.014		-0.0029 -0.012
GR Imm.		-0.0209* -0.0119		0.0292 -0.0189		-0.0604*** -0.0133
In-State Native	0.1546*** -0.0074	0.1547*** -0.0074	0.1233*** -0.0107	0.1233*** -0.0107	0.1792*** -0.0101	0.1793*** -0.0101
Top School	-0.0359*** -0.0063	-0.0373*** -0.0064	-0.012 -0.0087	-0.0125 -0.0088	-0.0861*** -0.0089	-0.0881*** -0.0089
MBA	0.0359*** -0.0068	0.0377*** -0.0069	0.004 -0.0102	0.0047 -0.0103	0.0717*** -0.009	0.0737*** -0.0091
Other Graduate Degree	-0.0290*** -0.0086	-0.0272*** -0.0086	-0.0372*** -0.0139	-0.0365*** -0.014	-0.0196* -0.0105	-0.0175* -0.0105
STEM Master's	0.0424*** -0.0097	0.0438*** -0.0097	0.0667*** -0.0152	0.0670*** -0.0153	0.0188 -0.0122	0.0211* -0.0122
PhD	0.0152 -0.0109	0.0192* -0.0109	0.0172 -0.017	0.0185 -0.017	0.0151 -0.0136	0.0204 -0.0138
Previous Start-up XP	-0.0536*** -0.011	-0.0533*** -0.011	-0.0081 -0.0179	-0.0079 -0.0179	-0.0899*** -0.0133	-0.0897*** -0.0132
Previous Founding XP	0.0257* -0.0141	0.0260* -0.0141	0.0014 -0.0228	0.0015 -0.0228	0.0405** -0.0164	0.0412** -0.0164
Founding Team Size	-0.0098*** -0.0021	-0.0099*** -0.0021	0.0094*** -0.0034	0.0094*** -0.0034	-0.0254*** -0.0025	-0.0255*** -0.0025
Observations	32,401	32,401	14,528	14,528	17,867	17,867
R-squared	0.2412	0.2414	0.2249	0.2249	0.1378	0.1384
Demographic controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Education state FE	Y	Y	Y	Y	Y	Y

\* Statistical significance at the 10 % level.

\*\* Statistical significance at the 5 % level.

\*\*\* Statistical significance at the 1 % level.

$$\text{SameStateFDR}_{it} = \alpha + \beta_1 \times \text{Immigrant}_{it} + \beta_2 \times \text{InState Native}_{it} + \gamma \times \text{Controls} + \text{FE} + \epsilon_{it}$$

Founder-company pairs are indexed by  $i$  and founding years are indexed by  $t$ .  $\text{SameStateFDR}_{it}$  equals one if the founder started a VC-backed company in the state in which he received his final postsecondary degree. Founders are divided into three groups: immigrants, in-state natives, and out-of-state natives. We use each native founder's SSN state to determine whether he is an in-state and out-of-state founder; that is, if the founder's SSN state is the same as the state in which he received his final postsecondary degree, then he is classified as an in-state founder.<sup>29</sup> Otherwise, he is considered an out-of-state founder. We use out-of-state native founders as the reference group. Naturally, since Post Degree immigrants do not have a US state of education, the sample only includes native founders, Undergraduate immigrant founders, and Graduate immigrant founders. We include a vector of control variables, detailed in Appendix C.1, that account for

<sup>29</sup> A native founder's SSN state is considered a decent proxy for his or her state of birth/childhood. Native founders in VS matched to multiple Infutor individuals are automatically classified as out-of-state, since we do not take a stand on which individual's SSN state to use in this situation. Similar logic justifies why we do not assign every matched founder a founding age. Thus, our proxy for in-state founders is, if anything, a conservative measure/underestimate of whether a native founder was an "in-state" student at any point.

differences across founders such as education and work experience. We also include industry, founding year, and state of education fixed effects. The regression allows us to compare the difference in probability of migration between these groups of founders.<sup>30</sup>

Table 5 presents the regression results. Columns 1 and 2 present results for all states. In column 1, we find that immigrant founders are not more likely to migrate, relative to out-of-state native founders. As intuition about in-state residents' geographic stickiness may suggest, in-state native founders are 15 percentage points (pp) more likely to stay in the same state, relative to the reference group. In column 2, we explore the difference between Undergraduate and Graduate immigrant founders and find that Undergraduate founders are 2 percentage points more likely to stay local, relative to the reference group, while Graduate founders are roughly 2 percentage points less likely to stay.<sup>31</sup> Compared to the unconditional non-migration probability of 0.338, in-state founders are approximately  $0.15/0.338 = 44\%$  more likely to stay, Undergraduate founders are roughly 6 % more likely to stay, and Graduate founders are roughly 6 % less likely to stay. Columns 3 and 4 present analogous results that restrict the sample to hub states (California, Massachusetts, and New York). Within this group of states, immigrant

<sup>30</sup> OLS regressions are used for ease of interpretation. Appendix Table B.2 presents analogous probit regression results, which are qualitatively similar to the OLS regression results.

<sup>31</sup> This final result is only statistically significant at the 10 % level.



founders are roughly 3.5 pp. more likely to stay and start their first companies than out-of-state native founders, even after conditioning on the state's average immigrant share and capacity for VC-backed entrepreneurship via the state FE. Compared to the non-migration probability of 0.516 within hub states, immigrants are roughly 7 % more likely to stay. Results from column 4 suggest that Undergraduate immigrants are the main contributors to the coefficient on the immigrant variable in column 3. Columns 5 and 6 present the results for non-hub states. We find that (i) immigrant founders are 2.6 percentage points less likely to stay and start companies in these states, relative to out-of-state native founders and (ii) this result is mainly driven by Graduate immigrants. Compared to the non-migration probability of 0.21 in these non-hub states, immigrants are roughly 12 % less likely to stay. Taken together, the results presented in columns 3 through 6 suggest that non-hub states are, at the margin, losing immigrant would-be founders to hub states, relative to out-of-state native would-be founders. Nonetheless, the results also show that Undergraduate immigrant founders educated in non-hub states do not behave in a statistically significantly different manner than out-of-state native founders when deciding whether to migrate before founding their startup.<sup>32</sup>

So far, the results from this section suggest that universities contribute to their local economies by retaining students who eventually start VC-backed companies within the same state. Therefore, to provide additional evidence that there is a direct relationship between universities and local entrepreneurial activity, we run variants of the following Poisson regression over the 2006–2018 timeframe:

$$N_{it} = \alpha + \beta_1 \times E_{i,t-5}^N + \beta_2 \times E_{i,t-5}^F + \gamma \times \text{Controls} + FE + \epsilon_{it}$$

Each observation in the regression is a state-year pair.  $N_{it}$  is the number of VC-backed startups that were founded in each year. The main variables of interest are the number of native students enrolled and the number of foreign students enrolled, both undergraduate and graduate, at all universities in the state in year  $t - 5$ . The goal of this regression is to provide suggestive evidence that student enrollment today is associated with local high-potential entrepreneurship in the future. Of course, the number of VC-backed startups founded in a state in a given year is likely to be correlated with local and national economic conditions. To this end, we add time-varying controls for the state's population, labor force participation rate, unemployment rate, income per capita, White population share, and native-born population share in each year (see Appendix C.2 for further details). Furthermore, we include state and founding year fixed effects.<sup>33</sup>

Table 6 presents the regression results. Column 1 presents the result for the association between all startup births and lagged enrollment. We

<sup>32</sup> Internal robustness checks to our regressions compare Post Degree (PD) immigrants' likelihood to found a startup in their state of entry (i.e., SSN State) to other immigrants' and natives' propensity to found a startup in their state of education. These checks show that PD immigrants are more likely to found a startup in their SSN state than out of state natives are to found a startup in their education state. This result likely reflects the fact that PD immigrants enter the US later than other immigrants and therefore have less time to move within the country before founding their startup. However, data limitations prevent us from reliably identifying the full job location history of individuals in our sample, so we are unable to conduct an apples-to-apples comparison of geographic stickiness across all immigrant groups by comparing PD immigrants' entry states to other immigrants' initial or immediate pre-founding job locations. Since comparing PD immigrants' entry states to other immigrants' education states is an incongruent comparison, we do not include PD immigrants in our analysis of geographic mobility.

<sup>33</sup> The number of observations drops when we consider immigrant-founded startups because an increased number of "zero" observations exacerbates statistical separation issues which prevent maximum likelihood/Poisson estimates from correctly converging. More detailed information on statistical separation can be found on the following website: [https://github.com/sergiocoreia/ppmlhdfc/blob/master/guides/separation\\_primer.md](https://github.com/sergiocoreia/ppmlhdfc/blob/master/guides/separation_primer.md).

find that both total native and foreign student enrollment are associated with future VC-backed startup formation. Specifically, interpreting the Poisson coefficients, we can conclude that admitting 1000 additional international students in universities within a state is correlated with an approximately  $\exp(1000 \times 1.24e - 05) - 1 = 1.2$  percentage points increase in the rate of new startup creation in that state 5 years into the future. In contrast, admitting an additional 1000 native-born students in universities within a state only correlates with an approximately  $\exp(1000 \times 1.20e - 06) - 1 = 0.12$  percentage points increase in the rate of new startup creation 5 years into the future in that state.

The remaining columns add further nuance to column 1's finding. Results from columns 2 suggest that greater native student enrollment today leads to a higher rate of native-founded startup formation in the future. The coefficient on foreign student enrollment is not statistically different from zero, which suggests that immigrant student enrollment likely has at most a minimal crowding out effect on native VC-backed entrepreneurship. Likewise, results in column 3 show that greater foreign student enrollment today leads to a higher rate of immigrant-founded startup formation in the future. Regression results presented in columns 4 through 6 explore whether the baseline results are driven by undergraduate or graduate enrollment. They suggest that undergraduate enrollment is the primary contributor to the relationship between university enrollment and VC-backed entrepreneurship presented in columns 1–3. Of course, throughout these regressions we cannot track which student enrollees become VC-backed entrepreneurs at the individual level, and it remains unclear whether students' entry into a founder vs. employee role within the startup space, among other factors, primarily accounts for our results. Nonetheless, when considered together with the statistical analyses presented in Table 4, Table 5 suggests that there is a direct connection between university student enrollment and future local entrepreneurial activity, and it thus further highlights the role that universities play as contributors to the sourcing and development of locally impactful entrepreneurial talent.

#### 4.2. Venture performance by immigration status – financial success

Up to this point, our empirical results strongly suggest that universities are important contributors to the *quantity* of immigrant VC-backed entrepreneurs. In this section we examine whether the supply of foreign entrepreneurial talent that enters the United States through universities is of comparable quality to the country's native entrepreneurial talent. We estimate versions of the following OLS regression equation to address this quality issue:

$$Y_{it} = \alpha + \beta \times \text{Immigrant}_{it} + \gamma \times \text{Controls} + FE + \epsilon_{it}.$$

Each observation in the sample is a founder-startup pair indexed by  $i$ . Time in years is indexed by  $t$ . We use the year of founding to assign a founder-startup pair to a given year. In this subsection, the outcome variable  $Y_{it}$  is a placeholder for various measures of financial success that we examine, which includes IPO probability and successful (i.e., profitable) acquisition probability.<sup>34</sup> The variable of interest is the immigrant indicator, which equals one if the founder is identified as an immigrant by the method described in Section 2. In the regression, we condition on the founder's demographics such as race, ethnicity, and gender. We follow the venture capital literature and condition on the founder's schooling background, which has been shown to be highly correlated with venture success (Gompers et al., 2016). In all specifications, we include industry fixed effects to difference out variation in success rates across industry, and we include year fixed effects to difference out temporal variation in success rate. Standard errors are clustered at the founder level because our variable of interest, immigrant

<sup>34</sup> IPO and successful acquisition probabilities are standard measures of startup financial success in the venture capital literature (Hochberg et al., 2007; Gompers et al., 2010; Gompers et al., 2016; Amornsiripanitch et al., 2019).

**Table 6**

University enrollment and future entrepreneurship.

This table presents Poisson regression results where measures of VC-backed startup activity in year  $t$  are regressed onto measures of students enrolled at local universities in year  $t - 5$ . Each observation is a state-year pair. All Firms is the number of VC-backed startups in state  $i$  that were started in year  $t$ . Native Firms is the number of native-founded VC-backed startups in state  $i$  that were started in year  $t$ . Imm Firms is the number of immigrant-founded VC-backed startups in state  $i$  that were started in year  $t$ . Total Native Students is the number of native students enrolled in universities in state  $i$  in year  $t - 5$ . Total Foreign Students is the number of foreign students enrolled in universities in state  $i$  in year  $t - 5$ . Native Graduate Students is the number of native graduate students enrolled in universities in state  $i$  in year  $t - 5$ . Foreign Graduate Students is the number of foreign graduate students enrolled in universities in state  $i$  in year  $t - 5$ . Native Undergraduate Students is the number of native undergraduate students enrolled in universities in state  $i$  in year  $t - 5$ . Foreign Undergraduate Students is the number of foreign undergraduate students enrolled in universities in state  $i$  in year  $t - 5$ . Refer to Appendix C for control variable definitions. Heteroskedasticity-robust standard errors are reported in parentheses and clustered at the state level.

	(1)	(2)	(3)	(4)	(5)	(6)
	All firms	Native firms	Immigrant. Firms	All firms	Native firms	Immigrant. Firms
Total Native Students	1.20e-06*** (4.07e-07)	1.52e-06*** (4.66e-07)	-2.68e-07 (4.40e-07)			
Total Foreign Students	1.24e-05** (5.95e-06)	7.27e-06 (5.81e-06)	3.04e-05*** (8.65e-06)			
Native Graduate Students				-1.60e-07 (2.40e-06)	-8.44e-07 (2.36e-06)	2.81e-06 (4.76e-06)
Foreign Graduate Students				4.69e-06 (1.12e-05)	4.45e-06 (1.13e-05)	3.44e-06 (1.69e-05)
Native Undergraduates				1.27e-06*** (4.41e-07)	1.65e-06*** (5.03e-07)	-4.32e-07 (6.06e-07)
Foreign Undergraduates				1.57e-05 (1.15e-05)	8.19e-06 (1.14e-05)	4.18e-05*** (1.33e-05)
Population	-7.24e-08** (3.68e-08)	-9.09e-08** (3.89e-08)	1.63e-08 (4.91e-08)	-7.40e-08** (3.56e-08)	-9.85e-08** (3.88e-08)	3.78e-08 (5.23e-08)
LFPR	-0.00500 (0.0252)	-0.00532 (0.0241)	0.000539 (0.0507)	-0.000170 (0.0239)	0.000668 (0.0237)	0.000145 (0.0506)
Unemployment Rate	0.0790*** (0.0210)	0.0842*** (0.0231)	0.0719** (0.0313)	0.0834*** (0.0202)	0.0883*** (0.0231)	0.0779** (0.0330)
Income per Capita	-1.11e-05 (1.56e-05)	-7.55e-06 (1.70e-05)	-1.56e-05 (2.41e-05)	-1.30e-05 (1.57e-05)	-8.54e-06 (1.76e-05)	-2.07e-05 (2.30e-05)
White Population Share	6.280* (3.219)	5.045* (2.732)	10.27* (6.226)	6.042* (3.197)	4.484 (2.829)	11.16* (6.218)
Native-Born Population Share	2.748 (5.675)	0.545 (6.189)	12.07 (10.36)	1.935 (5.992)	0.269 (6.579)	8.693 (9.998)
Observations	650	650	559	650	650	559
State FE	Y	Y	Y	Y	Y	Y
Founding year FE	Y	Y	Y	Y	Y	Y

\* Statistical significance at the 10 % level.

\*\* Statistical significance at the 5 % level.

\*\*\* Statistical significance at the 1 % level.

status, is assigned at the founder level.<sup>35</sup> We also account for the variation in founding team size by directly conditioning on this information. For ease of interpretation, we estimate linear probability (i.e., OLS) models.

Table 7 presents the regression results. Columns 1 and 2 present the results for *Success*, an indicator variable that equals one if the startup goes public in an IPO or was acquired for more than its total investment value by 2019. In column 1, the coefficient on the immigrant indicator is positive and statistically significant. The economic magnitude is also sizable: immigrant-founded companies are 1.7 percentage points more likely to succeed than native-founded companies. Compared to the unconditional success rate of 16.2 % in the sample, the coefficient on the immigrant indicator represents a 10.5 % relative increase. It is important to note that one advantage of our immigrant classification method is highlighted in this set of results. If we were to define immigration status using founders' race and ethnicity, the regression results would likely have been very different, as none of our specification's coefficient estimates on the race and ethnicity variables are statistically different from zero.

In column 2, we explore whether immigration pathway groups explain immigrants' aggregate startup success. We find that the positive immigrant coefficient in column 1 is largely driven by excess success from Graduate and Post Degree immigrant founders, though

Undergraduate immigrants also perform, at worst, similarly to native-born founders. The remaining columns present regression analyses that further disaggregate and decompose our definition of startup financial success. Columns 3 and 4 present results on the relationship between immigration status and startup IPO probability. We find that, compared to natives, immigrant founders are not more likely to start companies that eventually go public in an IPO. Results in columns 5 and 6 suggest that, although the immigrant indicator variable is not statistically different from zero, Post Degree immigrants are 1.2 percentage points more likely to start companies that will eventually go public. Once again, the economic magnitude is large. The coefficient is equivalent to a 25 % increase in IPO rate relative to the sample mean IPO rate of 4.8 %. Results presented in columns 5 and 6 show that the positive correlation between *All Successes* and the immigrant indicator variable is driven by success via acquisition.<sup>36</sup>

There are several key takeaways from this set of regression results. First, on average, immigrant founders are more likely to start financially successful companies as measured by going public in an IPO or through a successful acquisition. This finding implies that immigrant entrepreneurs who enter the US do not seem to lower the quality of US startups.

<sup>36</sup> The number of observations differ in columns 5 and 6 because we exclude companies that reached the IPO stage. Appendix Table B.3 presents probit regression results as a robustness check. We find that the results are qualitatively similar.

<sup>35</sup> All of our results are robust to clustering at the startup level.

**Table 7**

Immigration status and venture success – Financial success.

This table presents OLS regression results where measures of startup financial success are regressed onto founder's immigration status. Each observation is a founder-startup pair. Success equals one if, by 2019, the startup reached the IPO stage or was acquired for a larger amount than the total funds invested, adjusted for inflation. IPO and Acquisition follow the same logic. Immigrant equals one if the founder is an immigrant. UG Imm. equals one if the founder is an immigrant who came to the United States for college. GR Imm. equals one if the founder is an immigrant who came to the United States for graduate school. PD Imm. equals one if the founder is an immigrant who came to the United States for work. Refer to Appendix C for control variable definitions. Heteroskedasticity-robust standard errors are reported in parentheses and clustered at the founder level.

	(1) Success	(2) Success	(3) IPO	(4) IPO	(5) Acquisition	(6) Acquisition
Immigrant	0.0171*** (0.0055)		0.0022 (0.0032)		0.0164*** (0.0051)	
UG Imm.		0.0072 (0.0074)		−0.0034 (0.0041)		0.0102 (0.0069)
GR Imm.		0.0222** (0.0089)		0.0018 (0.0051)		0.0214** (0.0083)
PD Imm.		0.0299*** (0.0101)		0.0124** (0.0062)		0.0211** (0.0095)
Top School	0.0177*** (0.0042)	0.0191*** (0.0042)	0.0075*** (0.0025)	0.0084*** (0.0026)	0.0124*** (0.0039)	0.0131*** (0.0039)
MBA	−0.0019 (0.0049)	−0.0021 (0.0049)	0.0046 (0.0031)	0.0046 (0.0031)	−0.0057 (0.0045)	−0.0060 (0.0046)
Other Graduate Degree	0.0069 (0.0061)	0.0063 (0.0061)	0.0064* (0.0038)	0.0063* (0.0038)	0.0017 (0.0056)	0.0013 (0.0057)
STEM Master's	0.0055 (0.0072)	0.0054 (0.0072)	−0.0033 (0.0045)	−0.0032 (0.0045)	0.0086 (0.0067)	0.0084 (0.0067)
PhD	0.0089 (0.0078)	0.0081 (0.0078)	0.0060 (0.0051)	0.0059 (0.0052)	0.0043 (0.0071)	0.0035 (0.0072)
Previous Start-up XP	0.0610*** (0.0104)	0.0607*** (0.0104)	0.0399*** (0.0071)	0.0397*** (0.0071)	0.0310*** (0.0096)	0.0308*** (0.0097)
Previous Founding XP	−0.0377*** (0.0119)	−0.0377*** (0.0119)	−0.0256*** (0.0078)	−0.0255*** (0.0078)	−0.0188* (0.0112)	−0.0188* (0.0112)
Founding Team Size	0.0222*** (0.0017)	0.0222*** (0.0017)	0.0142*** (0.0011)	0.0142*** (0.0011)	0.0117*** (0.0015)	0.0117*** (0.0015)
Female	−0.0067 (0.0059)	−0.0067 (0.0059)	−0.0017 (0.0037)	−0.0016 (0.0037)	−0.0056 (0.0054)	−0.0056 (0.0054)
Jewish	0.0063 (0.0050)	0.0064 (0.0050)	−0.0035 (0.0029)	−0.0035 (0.0029)	0.0094** (0.0046)	0.0094** (0.0046)
East Asian	−0.0017 (0.0070)	−0.0016 (0.0069)	0.0019 (0.0039)	0.0022 (0.0039)	−0.0035 (0.0064)	−0.0036 (0.0064)
Indian	0.0064 (0.0071)	0.0049 (0.0072)	−0.0030 (0.0042)	−0.0032 (0.0042)	0.0084 (0.0067)	0.0072 (0.0068)
Hispanic	0.0109 (0.0085)	0.0109 (0.0085)	−0.0052 (0.0042)	−0.0052 (0.0042)	0.0154* (0.0081)	0.0154* (0.0081)
Observations	36,333	36,333	36,333	36,333	34,576	34,576
R-squared	0.1271	0.1272	0.0979	0.0980	0.0714	0.0714
Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

\* Statistical significance at the 10 % level.

\*\* Statistical significance at the 5 % level.

\*\*\* Statistical significance at the 1 % level.

Second, immigration pathways and educational backgrounds seem to influence immigrant founders' startup success. Specifically, immigrants who come to the US via graduate school or work seem to be particularly apt in starting financially successful companies. The result on Graduate immigrants suggests that universities' graduate programs are particularly important at producing high quality entrepreneurs.

#### 4.3. Venture performance by immigration status – patenting

In this subsection, we consider how a founder's immigration status correlates with the innovative production and potential of a startup as opposed to the startup's financial success. To measure a startup's propensity for technical scientific invention, we use the USPTO patent data merged onto VentureSource to identify (i) whether a startup filed a successful patent within 2 years of founding and (ii) how many patents (unweighted or citation-weighted) a startup filed within this 2-year period. As we have discussed in Section 3, within two years of establishment, immigrant founder startups are more likely to patent and file more patents than their native-born counterparts. Given that there are differences in the propensity to found companies in various industries

between immigrants and native founders, it is necessary to control for industry effects on patenting. We first provide preliminary descriptive evidence that within-industry differences account for some portion of the observed difference in patenting between immigrant and native-founded startups, but not all of the difference. Specifically, Appendix Table B.4 tabulates the immigrant and native share of startups within each of the eight industry groups used in our analysis as well as the average within-industry patenting likelihood (patent rate) and patent counts for immigrant and native-founded startups. Immigrant-founded startups appear to (i) be more likely to patent and (ii) file more patents than their native-born counterparts within seven of the eight industries. In addition, the share of immigrant-founded startups in more patent-intensive industries (as measured by overall patent rates and count in columns 8 and 9) appears to be higher than the corresponding share in less patent-intensive industries. To formally test the effect of these industry patterns, we use a regression specification analogous to the one presented in the previous section to explore the conditional correlation between a startup's innovation output and founders' immigration status. Table 8 presents the regression results.

In Table 8 we first consider the extensive margin of patenting: are

**Table 8**

Immigration status and venture success – Patenting.

This table presents OLS and Poisson regression results where measures of startup patenting success are regressed onto founder's immigration status. Patent >0 equals one if the startup filed at least one patent within the first two years of founding. Patent Count is the number of patent applications that the startup filed within the first two years of founding. Citation-Weighted Patents is defined as the number of patents, weighted by forward patent citations, filed by the founder's startups within two years of founding. Immigrant equals one if the founder is an immigrant. UG Imm. equals one if the founder is an immigrant who came to the United States for college. GR Imm. equals one if the founder is an immigrant who came to the United States for graduate school. PD Imm. equals one if the founder is an immigrant who came to the United States for work. Refer to Appendix C for control variable definitions. Heteroskedasticity-robust standard errors are reported in parentheses and clustered at the founder level.

	(1)	(2)	(3)	(4)	(5)	(6)
	Patent > 0	Patent > 0	Count	Count	Cite-W count	Cite-W count
Immigrant	0.0121** (0.0054)		0.1764** (0.0727)		0.0173 (0.1063)	
UG Imm.		−0.0064 (0.0071)		0.1470 (0.2174)		−0.0088 (0.1831)
GR Imm.		0.0281*** (0.0090)		0.2267* (0.1245)		0.0755 (0.1356)
PD Imm.		0.0244** (0.0097)		0.1397 (0.1086)		−0.0363 (0.1984)
Top School	0.0121*** (0.0041)	0.0139*** (0.0042)	0.1184* (0.0678)	0.1184** (0.0596)	0.1931** (0.0881)	0.1912** (0.0882)
MBA	−0.0030 (0.0048)	−0.0041 (0.0048)	−0.0860 (0.0710)	−0.0910 (0.0676)	−0.1371 (0.1156)	−0.1422 (0.1157)
Other Graduate Degree	0.0173*** (0.0058)	0.0158*** (0.0058)	0.0931 (0.0902)	0.0881 (0.0867)	−0.0139 (0.1348)	−0.0188 (0.1358)
STEM Master's	0.0207*** (0.0070)	0.0201*** (0.0070)	0.1784* (0.0914)	0.1761* (0.0914)	0.3977** (0.1566)	0.3958** (0.1573)
PhD	0.0483*** (0.0081)	0.0458*** (0.0082)	0.4382*** (0.1030)	0.4308*** (0.1127)	0.3253** (0.1438)	0.3188** (0.1426)
Previous Start-up XP	0.0335*** (0.0101)	0.0331*** (0.0101)	0.4785*** (0.1026)	0.4789*** (0.1026)	0.4227** (0.1746)	0.4238** (0.1747)
Previous Founding XP	0.0160 (0.0122)	0.0159 (0.0122)	0.1648 (0.1396)	0.1632 (0.1411)	0.1883 (0.1881)	0.1857 (0.1884)
Founding Team Size	0.0109*** (0.0017)	0.0109*** (0.0017)	0.0631*** (0.0237)	0.0634*** (0.0238)	0.0744*** (0.0286)	0.0747*** (0.0287)
Female	−0.0108* (0.0061)	−0.0107* (0.0061)	−0.3278*** (0.0790)	−0.3269*** (0.0798)	−0.5206*** (0.1551)	−0.5199*** (0.1549)
Jewish	0.0006 (0.0048)	0.0006 (0.0048)	−0.0089 (0.0689)	−0.0090 (0.0691)	0.1569 (0.1145)	0.1566 (0.1145)
East Asian	0.0031 (0.0077)	0.0027 (0.0077)	−0.0733 (0.1020)	−0.0799 (0.1021)	0.0742 (0.1837)	0.0648 (0.1924)
Indian	0.0001 (0.0073)	−0.0039 (0.0074)	0.1761 (0.1422)	0.1639 (0.1699)	0.0703 (0.1248)	0.0532 (0.1247)
Hispanic	−0.0022 (0.0094)	−0.0021 (0.0094)	−0.4359*** (0.1093)	−0.4339*** (0.1106)	−0.6069** (0.2661)	−0.6062** (0.2661)
Regression type	OLS	OLS	Poisson	Poisson	Poisson	Poisson
Observations	36,333	36,333	36,013	36,013	36,013	36,013
R-squared	0.0691	0.0695				
Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

\* Statistical significance at the 10 % level.

\*\* Statistical significance at the 5 % level.

\*\*\* Statistical significance at the 1 % level.

immigrant founders more likely establish startups that file at least one successful patent? Column 1 accordingly presents OLS regression results for the conditional correlation between a startup's probability of patenting (patenting status) and immigration status. The outcome variable is an indicator variable that equals one if the startup associated with the founder filed at least one ultimately successful patent within the first two years of founding. In this definition, we evaluate whether a patent has a grant date to determine whether it is “ultimately successful.” Although startup-level patent counts and citation-weighted patent counts (presented in columns 3–6) may present more complete information about the scientific and economic value of a startup's innovative activity, our indicator measure of a startup's patenting status better characterizes a startup's decision on whether to engage in any successful patentable innovative activity. Indeed, prior work on immigrant

invention and entrepreneurship (cf., [Hunt, 2011](#); [Brown et al., 2019](#)) has complemented information on patent counts with this binary measure of inventor/startup patenting to portray a more complete description of innovative activity differences based on immigration status. In addition, while we focus on patents assigned to startups as opposed to patents directly invented by the founder, our focus on patents filed within two years of startup founding helps to ensure with reasonable likelihood that the founder was substantially involved in the creation and development of the patented invention, even if not directly listed as an inventor on the patent application. Column 1 results suggest that immigrant founders are 1.2 pp. more likely to start companies that produce at least one successful patent in the early stages of their development. Extending column 1's findings, column 2 presents OLS regression results that explore the differences in startup patenting status across immigration



paths. The regression shows that the baseline result is driven by Graduate and Post Degree immigrants, which is consistent with the financial success results.<sup>37</sup>

Since immigrant-founded startups are more likely to enter patenting on the extensive margin than their native-born counterparts, a natural follow-up question lies in evaluating how immigrant startups compare to native startups on the intensive margin. In other words, do immigrant-founded startups produce more patents and higher-quality patents than their native counterparts? Addressing this question could suggest the extent to which differences in innovative ability, as opposed to uneven VC-signaling incentives, could explain immigrant founders' higher propensity to enter patenting on the extensive margin. If immigrant founders only have greater incentive to file at least one patent solely as a signal of entrepreneurial quality to VCs than their native-born counterparts, they may be more inclined to file a single patent of mediocre scientific quality. One might expect the aggregate intensive-margin quantity and quality of immigrant founders' patents to be lower. However, if immigrant founders' entry into startup patenting arose due to their innovative abilities, the aggregate quantity and scientific quality of their startups' patents would be similar to or higher than those of native founders' startups. Therefore, we test whether immigrant-founded startups' unweighted and citation-weighted patent counts differ from those of native-founded startups in columns 3–6 of Table 8 (and Appendix Table B.5).<sup>38</sup>

Columns 3 and 4 present Poisson regression results where the dependent variable is patent count, which is defined as the number of ultimately successful patent applications that the company filed within the first two years of founding. The results are qualitatively similar. Immigrant founders' startups produce, on average,  $\exp(0.1764 - 1) = 19\%$  more patents (as measured by unweighted count) than their native-born counterparts' startups. Similar to the financial success result, the immigrant indicator variable tends to be positively correlated with patenting (though the relationship is less statistically significant, and it provides more information than founders' race and ethnicity indicator variables because the race and ethnicity variables are negative or statistically indistinguishable from zero). Columns 3 and 4 of Appendix Table B.5 display analogous results to columns 3 and 4 of Table 8, except that the regression sample is restricted to startups with at least one patent.

Likewise, columns 5 and 6 present Poisson regression results where the dependent variable is citation-weighted patent count. This outcome variable is defined similarly to patent count, except that each ultimately successful patent is weighted by its associated number of forward citations when being counted. This citation-weighted patent count follows standard methodology used in the innovation economics literature to measure the scientific value of a firm's patented inventions (Hall et al., 2001), and our specifications' inclusion of industry and year fixed effects resolves concerns about (i) differences in industry citation practices and (ii) temporal truncation bias, both of which are often used to criticize citation-weighted patent counts' value as a measure of scientific quality. Accordingly, the goal of this exercise is to explore the difference in the aggregate quality of patents produced by immigrant- and native-founded startups. We find that there is no statistical difference in citation-weighted patent counts between the two groups.

Overall, these results suggests that immigrants tend to start companies that are more innovative on the extensive margin, i.e., companies

that are more likely to produce patents, even when controlling for broad industry classifications via industry fixed effects. However, conditional on being in a more innovative sector of the economy, immigrant-founded companies do not innovate decisively more on the intensive margin. That is, though immigrant-founded startups do produce a moderately higher number of patents, the patents that they produce are of comparable, not higher, scientific impact, as measured by citations.<sup>39</sup> Though immigrant founders do not conditionally outperform native founders in the scientific impact of their patenting activity, the similar scientific quality of startup patents across these two founder groups, combined with immigrant founders' higher propensity to enter patenting on the extensive margin, suggest that immigrant founders critically and substantially contribute to the aggregate innovative output of VC-backed startups.

#### 4.4. Founding team composition and venture performance

We have so far considered immigrant and native founders separately, and we have abstracted away from the founding team synergies that immigrant and native founders might experience with themselves as well as each other. Nonetheless, since founding team composition does matter for venture outcomes (Delmar and Shane, 2006), we estimate the conditional correlation between founding team composition and venture performance. We do so by running variants of the OLS and Poisson regressions from Sections 4.2 and 4.3. Specifically, we regress the same venture outcome variables used in the previous two sections on a vector of founding team composition indicator variables:

$$Y_{it} = \alpha + \beta_1 \times \text{Immigrant Min}_{it} + \beta_2 \times \text{Immigrant Maj}_{it} + \beta_3 \times \text{Immigrant All}_{it} + \gamma \times \text{Controls} + FE + \epsilon_{it}.$$

For this regression specification, the unit of observation is a startup  $i$  that was started in year  $t$ .<sup>40</sup> The first variable of interest is *Immigrant Min*, which equals one if fewer than half of the founding team is composed of immigrants. The second variable of interest is *Immigrant Maj*, which equals one if more than half of the founding team is composed of immigrants. This second variable excludes cases where every member of the founding team is an immigrant. The last variable, *Immigrant All*, equals one if every member of the founding team is an immigrant. The sample naturally excludes startups where the immigration status of every founder cannot be determined. Altogether, the reference group is composed of startups that have all-native founding teams. Founder-level control variables such as demographic and schooling variables are converted to their respective startup-level

<sup>37</sup> Appendix Table B.5 (columns 1–2) presents probit regression results as a robustness check. We find that the results are qualitatively similar.

<sup>38</sup> The number of observations drops when we consider patent count variables because “zero” observations cause statistical separation issues which prevent maximum likelihood/Poisson estimates from correctly converging. More detailed information on statistical separation can be found on the following website: [https://github.com/sergiocorreia/ppmlhdf/blob/master/guides/separation\\_primer.md](https://github.com/sergiocorreia/ppmlhdf/blob/master/guides/separation_primer.md).

<sup>39</sup> Columns 3 through 6 of Appendix Table B.5 present Poisson regression results for patent count and citation-weighted patent count using the sample of founder-company pairs that have at least one patent. All four regressions show that, conditional on having at least one patent, there is no difference in patenting quantity or (citation-measured) quality between immigrant- and native-founded companies, which indicates that patenting activity does not significantly differ on the intensive margin. The similarity of immigrant and native founders' conditional patenting behaviors on the intensive margin suggests that immigrant founders' higher propensity to enter into patenting on the extensive margin is likely not completely driven by the VC-signaling efforts mentioned in the previous paragraph.

<sup>40</sup> Of the 23,675 founding teams/startups used in our empirical analysis, 17,729 (~75 %) are founded entirely by natives (single or multiple), 3507 (~15 %) are founded entirely by immigrants (single or multiple), and 2439 (~10 %) are co-founded by natives and immigrants. Of the 2439 co-founded startups, 602 (~25 %) have a founding team that is made up of mostly natives, 226 (~9 %) have a founding team that is made up of mostly immigrants, and the remaining 1611 (~66 %) have a founding team with an identical number of natives and immigrants. Since there are a small number of startups that are truly majority immigrant co-founded, we categorize firms equally co-founded by natives and immigrants as majority immigrant in our Table 9 analysis.

**Table 9**

Founding team composition and venture success.

This table presents OLS and Poisson regression results where measures of startup success are regressed onto founding team composition indicator variables. Each observation is a startup. Immigrant Min equals one if fewer than half of the founding team members are immigrants. Immigrant Maj equals one if more than half of the founding team members are immigrants. Immigrant All equals one if every member of the founding team is an immigrant. Demographic control variables include indicator variables for Female, Jewish, East Asian, Indian, and Hispanic founders. Refer to Appendix C for control variable definitions. Heteroskedasticity-robust standard errors are reported in parentheses and clustered at the founder level.

	(1)	(2)	(3)	(4)	(5)	(6)
	Success	IPOs	Acquisitions	Patent > 0	Pat. count	Cite-W count
Immigrant Min	0.0223 (0.0178)	0.00826 (0.0117)	0.0156 (0.0165)	0.0106 (0.0168)	0.0861 (0.173)	-0.0223 (0.256)
Immigrant Maj	0.00983 (0.00937)	0.000777 (0.00552)	0.0109 (0.00881)	0.0122 (0.00955)	-0.00475 (0.115)	-0.319* (0.178)
Immigrant All	0.0213*** (0.00705)	0.00261 (0.00387)	0.0200*** (0.00662)	0.0173** (0.00702)	0.251* (0.130)	0.162 (0.164)
Founding Team Size	0.0195*** (0.00367)	0.0132*** (0.00228)	0.00966*** (0.00341)	0.00279 (0.00362)	0.00404 (0.0627)	0.0304 (0.0833)
Top School	0.0179*** (0.00481)	0.00726*** (0.00272)	0.0127*** (0.00449)	0.0129*** (0.00481)	0.158* (0.0954)	0.236** (0.115)
MBA	-0.00564 (0.00525)	0.000856 (0.00305)	-0.00636 (0.00488)	-0.00683 (0.00518)	-0.146* (0.0845)	-0.192 (0.123)
Other Graduate Degree	0.00471 (0.00665)	0.00215 (0.00389)	0.00324 (0.00619)	0.0190*** (0.00638)	0.223** (0.0925)	0.0972 (0.162)
STEM Master's	0.00843 (0.00726)	0.00193 (0.00439)	0.00723 (0.00676)	0.0184** (0.00719)	0.0847 (0.0922)	0.313** (0.157)
PhD	0.00500 (0.00788)	0.00690 (0.00493)	-0.000747 (0.00729)	0.0498*** (0.00830)	0.398*** (0.0997)	0.150 (0.143)
Previous Start-up XP	0.0570*** (0.0112)	0.0361*** (0.00749)	0.0303*** (0.0104)	0.0283*** (0.0109)	0.519*** (0.123)	0.399** (0.190)
Previous Founding XP	-0.0369*** (0.0126)	-0.0242*** (0.00816)	-0.0200* (0.0117)	0.0232* (0.0129)	0.214 (0.164)	0.273 (0.196)
Regression type	OLS	OLS	OLS	OLS	Poisson	Poisson
Observations	23,641	23,641	22,643	23,641	23,367	23,367
R-squared	0.122	0.093	0.069	0.070		
Demographic controls	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

\* Statistical significance at the 10 % level.

\*\* Statistical significance at the 5 % level.

\*\*\* Statistical significance at the 1 % level.

counterparts. For example, the East Asian founder indicator variable for the founder-startup pair regression would be converted to equal one if at least one East Asian founder was found on the team. Finally, we also condition on founding team size.

Table 9 presents the regression results. Columns 1 through 3 present OLS regression results for financial success outcomes. We find that the results presented in Section 4.2 are largely driven by all-immigrant founding teams. Columns 4 through 6 present OLS and Poisson regression results for patenting outcomes. In line with the financial success results, we find that the results presented in Section 4.3 are also largely driven by all-immigrant founding teams. Overall, conditional on teams' demographic information, the results show that all-immigrant founding teams produce the most productive startups, suggesting that shared immigration experience may influence founding team performance to a greater extent than other factors, such as shared ethnic or racial background.<sup>41</sup>

#### 4.5. Selection and omitted variable bias

The regression results presented in Sections 4.2 through 4.5 suggest that immigrant founders are more likely to start financially successful startups, as measured by the probability of (i) going public in an IPO or (ii) being acquired at a value greater than total VC investment. The results also suggest that immigrant founders start more innovative companies, as measured by the likelihood of patenting and patent count.

Immigrant founders are therefore comparable, if not superior, in quality and entrepreneurial ability. Thus, foreign entrepreneurial talent that enters the US venture capital ecosystem does not seem to dilute the quality of startups that get funded.

It is important to note that the results reported above are conditional correlations and not causal estimates of the treatment effect that immigration status has on VC-backed startup outcomes. The regression results presented above may suffer from omitted variable bias (OVB) that stems from the unobservable quality of each founder, for which we cannot control. OVB arises in this context because immigration status is not randomly assigned and is hence not likely to be orthogonal to unobservable quality. Therefore, the coefficients of interest that we discuss above can be thought of as the combined result the selection, which manifests through the selection process of immigration, i.e., who chooses to immigrate, and treatment effects, which can be characterized as the immigration process's impact on a high-skilled individual's ultimate entrepreneurial ability.

Nonetheless, despite these OVB concerns, the regression results are still informative for the policy questions that we consider. For an immigrant-receiving country such as the United States, policymakers should care about both selection into high-skilled immigration and the treatment effect of such an immigration process, so long as their combined influence on immigrant talent and ability is positive and beneficial. In fact, understanding selection is perhaps more relevant than understanding the treatment effect because policymakers in an immigrant-hosting country may sometimes find it preferable and

<sup>41</sup> Appendix Table B.8 presents analogous probit regression results, which are qualitatively similar and show that the OLS findings are robust.

perhaps less costly to design policies that attract talented people (as opposed to policies that improve the immigration process's impact on foreigners' economic performance and contributions in the host country).<sup>42</sup> In the current context, positive selection may play a large role, given that, as shown in previous sections, immigrant founders are more likely to be better educated than native founders and, intuitively, immigrants to the US via higher education are more likely to have relatively affluent backgrounds because substantial resources are required for such undertakings. We believe that, regardless of the channel (selection or treatment), the regression results show that immigrant entrepreneurs who come to the US, mostly via higher education, are high quality individuals and, hence, universities are important contributor to both the quantity and quality of VC-backed entrepreneurial talent.

## 5. Conclusion

In this paper, we assemble new data to document several novel facts about immigrant VC-backed entrepreneurship in the United States. We show that higher education serves as an important contributor to the quantity and quality of entrepreneurial talent. First, we document that one fifth of VC-backed startup founders are immigrants. More importantly, we find that approximately 75 % of immigrant founders came to the United States via higher education. Using this second fact as motivation, we show that universities contribute to local economies by enrolling students, both native and foreign, who tend to start VC-backed companies in the same state in which they received their final post-secondary education degree. Lastly, we show that immigrant entrepreneurs are more likely to start financially successful and innovative companies, which suggests that the foreign entrepreneurial talent that enters the United States through universities is, on average, high quality. While our focus on VC-backed startups limits our ability to draw general conclusions on the overall relationship between immigration and entrepreneurship, this focus allows us to more precisely characterize immigrants' contribution to a particularly valued, innovative, and transformative subset of entrepreneurship.

Accordingly, our findings have several policy implications. Government policies that affect the flow of foreign students into the United States are likely to affect the flow of entrepreneurial talent into the country. Restricting the flow of foreign students into the United States or the ability of foreign students to stay in the country after earning degrees would restrict an important source of innovative and entrepreneurial talent that contributes to the US economy. Similarly, our results highlight that university admissions decisions to admit high-skilled foreign students also carry important implications for the broader economy in future years.

Our data's coverage of the VC-backed entrepreneurial ecosystem in the US naturally renders these implications most applicable for US policymakers. However, as other advanced economies, such as Israel, the UK, the EU, and multiple Commonwealth countries, also house vibrant VC hubs and attract substantial foreign talent through universities, these policy implications may also guide non-US policymakers seeking to harness foreign students as a source of entrepreneurial talent. Nonetheless, we encourage future work to more rigorously examine the extent to which our findings generalize to other advanced economies, as our data do not allow us to directly analyze the linkages between immigration and VC-backed entrepreneurship outside of the US.

Within the US, the beneficiaries of immigrant entrepreneurship in our sample have primarily been coastal states. One driver of this fact is the presence of leading research universities, which tend to have a larger

share of immigrant students, on the coasts. A sizable proportion of immigrant founders tend to start firms in the same states that they received their education. These results suggest that a potential lever that can contribute to local economic growth is attracting high-skilled immigrant students to local universities. However, given the long average lag between arrival in the US and starting a firm, it would likely take a sustained effort over an extended period in order to observe the benefits of such a policy. Policies targeted at attracting immigrant students are also likely not sufficient on their own. Broader policy changes to attract capital and other resources must be implemented concurrently in order make a location attractive for immigrants to remain in for the longer term.

Our findings on the importance of US universities in attracting VC-backed entrepreneurs also contains important implications for US policy surrounding work visas. Specifically, whereas US policy discourse surrounding high-skilled work visas (e.g., the H-1B visa) largely frame such visas primarily as tools to recruit new talent directly from abroad. Nonetheless, our findings show that US universities represent a primary entry point for high-potential immigrant entrepreneurs. If our findings for future VC-backed immigrant entrepreneurs generalize to the broader population of high-skilled immigrants in the US, it may be sensible to recharacterize high-skilled work visas as tools to retain US-educated talent in policy discourse. Such a reframing would naturally promote consideration of how to smooth the transition from student to work visas and may encourage the increase in the number of work visas that are given out each year, which, currently, is a major bottleneck for immigrants' ability to remain in the US.

Future research can address several remaining questions related to the importance of immigrant entrepreneurs. First, the datasets (e.g., Infutor and Emsi) that are assembled in this paper provide a viable alternative to the Census Bureau's data for entrepreneurship and innovation researchers who are interested in the intersection between these topics and immigration policy. Although the coverage of our data is generally not as comprehensive as Census data, they are simpler to merge with other datasets, and, in some panel-related dimensions (e.g., schooling, job titles, and work history), they contain more granular information that may be useful to this line of research. Thus, we hope other researchers will use and perhaps further enrich our alternative data platform for additional researchers. For example, future researchers could extend our dataset by adding more comprehensive information on the locations of individuals' pre-entrepreneurship jobs, and they could use the additional information to study the role of incumbent corporations (in addition to universities) as magnets and incubators of entrepreneurial talent. Second, we have presented useful conditional correlations. Specifically, we have established a positive relationship between foreign student enrollment and local entrepreneurship activity as well as a positive correlation between immigration status and various positive VC-backed startup outcomes. Future research can address the causal relationships between these variables as well as the channel for these effects. Similar analyses can also be performed to study the immigration path and economic contribution of foreign inventors, scientists, and other high-skilled workers. Finally, because immigration pathways appear to influence immigrant founders' startup outcomes, future research should more fully explore the educational and vocational mechanisms that may drive this result.

## CRedit authorship contribution statement

**Natee Amornsiripanitch:** Conceptualization, Methodology, Software, Writing – review & editing. **Paul A. Gompers:** Supervision, Conceptualization, Methodology, Writing – review & editing. **George Hu:** Conceptualization, Methodology, Software, Writing – review & editing. **Kaushik Vasudevan:** Conceptualization, Methodology, Software, Writing – review & editing.

<sup>42</sup> It is important to note that if the coefficients of interest are purely driven by the selection effect, then it is not necessarily true that doubling the number of foreign students will double the amount of entrepreneurial talent in the US because, eventually, such doubling might deplete the stock/available number of talented immigrants.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.respol.2023.104782>.

## References

- Akcigit, Ufuk, Dinlersoz, Emin, Greenwood, Jeremy, Penciakova, Veronika, 2019. Synergizing Ventures. No. w26196. National Bureau of Economic Research.
- Akcigit, Ufuk, Kerr, William R., 2018. Growth through heterogeneous innovations. *J. Polit. Econ.* 126 (4), 1374–1443.
- Amornsiripanitch, Natee, Gompers, Paul A., Xuan, Yuhai, 2019. More than money: venture capitalists on boards. *J. Law Econ. Org.* 35 (3), 513–543.
- Åstebro, Thomas, Bazzazian, Navid, Braguinsky, Serguey, 2012. Startups by recent university graduates and their faculty: implications for university entrepreneurship policy. *Res. Policy* 41 (4), 663–677.
- Audretsch, David B., Lehmann, Erik E., Warning, Susanne, 2017. University spillovers and new firm location. In: *Universities And the Entrepreneurial Ecosystem*. Edward Elgar Publishing.
- Azoulay, Pierre, Jones, Benjamin F., Daniel Kim, J., Miranda, Javier, 2020. Immigration And Entrepreneurship in the United States. No. w27778. National Bureau of Economic Research.
- Azoulay, Pierre, Jones, Benjamin F., Daniel Kim, J., Miranda, Javier, 2020. Age and high-growth entrepreneurship. *Am. Econ. Rev.: Insights* 2 (1), 65–82.
- Balasubramanian, Natarajan, Sivadasan, Jagadeesh, 2011. What happens when firms patent? New evidence from US economic census data. *Rev. Econ. Stat.* 93 (1), 126–146.
- Babinaetal., forthcoming Tania Babina Alex X. He Sabrina T. Howell Elisabeth Perlman Joseph Staudt. "Cutting the innovation engine: how federal funding shocks affect university patenting, entrepreneurship, and publications." *Q. J. Econ.*, forthcoming.
- Baptista, Rui, Lima, Francisco, Mendonça, Joana, 2011. Establishment of higher education institutions and new firm entry. *Res. Policy* 40 (5), 751–760.
- Bartik, Timothy J., 1985. Business location decisions in the United States: estimates of the effects of unionization, taxes, and other characteristics of states. *J. Bus. Econ. Stat.* 3 (1), 14–22.
- Bell, Alex, Chetty, Raj, Jaravel, Xavier, Petkova, Neviana, Van Reenen, John, 2019. Who becomes an inventor in America? The importance of exposure to innovation. *Q. J. Econ.* 134 (2), 647–713.
- Bernstein, Shai, Diamond, Rebecca, McQuade, Timothy, Pousada, Beatriz, 2020. The contribution of high-skilled immigrants to innovation in the United States. In: *Stanford Graduate School of Business Working Paper* 3748, en (2018), pp. 19–20.
- Bernstein, Shai, Giroud, Xavier, Townsend, Richard R., 2016. The impact of venture capital monitoring. *J. Financ.* 71 (4), 1591–1622.
- Blume-Kohout, Margaret E., 2016. Why are some foreign-born workers more entrepreneurial than others? *J. Technol. Transf.* 41 (6), 1327–1353.
- Borjas, George J., 1989. Immigrant and emigrant earnings: a longitudinal study. *Econ. Inq.* 27 (1), 21–37.
- Borjas, George J., 1991. Immigration and self-selection. In: *Immigration, Trade, And the Labor Market*. University of Chicago Press, pp. 29–76.
- Borjas, George J., Bratsberg, Bernt, 1996. Who leaves? The outmigration of the foreign-born. *Rev. Econ. Stat.* 165–176.
- Bound, John, Demirci, Murat, Khanna, Gaurav, Turner, Sarah, 2015. Finishing degrees and finding jobs: US higher education and the flow of foreign IT workers. *Innov. Policy Econ.* 15 (1), 27–72.
- Bound, John, Braga, Breno, Khanna, Gaurav, Turner, Sarah, 2021. The globalization of postsecondary education: the role of international students in the US higher education system. *J. Econ. Perspect.* 35 (1), 163–184.
- Bramwell, Allison, Wolfe, David A., 2008. Universities and regional economic development: the entrepreneurial University of Waterloo. *Res. Policy* 37 (8), 1175–1187.
- Brown, J.D., Earle, John S., Kim, Mee Jung, Lee, Kyung Min, 2019. "Immigrant Entrepreneurs And Innovation in the U.S. High-Tech Sector." NBER Working Paper 25565.
- Budiman, A., Tamir, C., Mora, L., Noe-Bustamante, L., 2020. Facts on US Immigrants, 2018. Pew Research Center.
- Carlton, Dennis W., 1983. The location and employment choices of new firms: an econometric model with discrete and continuous endogenous variables. *Rev. Econ. Stat.* 440–449.
- Chaney, Thomas, Sraer, David, Thesmar, David, 2012. The collateral channel: how real estate shocks affect corporate investment. *Am. Econ. Rev.* 102 (6), 2381–2409.
- Cohn, Jonathan B., Liu, Zack, Wardlaw, Malcolm, 2021. Count Data in Finance.
- Delmar, Frédéric, Shane, Scott, 2006. Does experience matter? The effect of founding team experience on the survival and sales of newly founded ventures. *Strateg. Organ.* 4 (3), 215–247.
- Di Gregorio, Dante, Shane, Scott, 2003. Why do some universities generate more startups than others? *Res. Policy* 32 (2), 209–227.
- Diamond, Rebecca, McQuade, Tim, Qian, Franklin, 2019. The effects of rent control expansion on tenants, landlords, and inequality: evidence from San Francisco. *Am. Econ. Rev.* 109 (9), 3365–3394.
- Eckhardt, Jonathan T., Harris, Clint, Chen, Chuan, Khoshimov, Bekhzod, Goldfarb, Brent, 2022. Student regional origins and student entrepreneurship. *Reg. Stud.* 56 (6), 956–971.
- Entis, Laura, 2013. Where startup funding really comes from (Infographic). November 20. *Entrepreneur Magazine*. <https://www.entrepreneur.com/money-finance/where-startup-funding-really-comes-from-infographic/230011>.
- Etzkowitz, Henry, 1998. The norms of entrepreneurial science: cognitive effects of the new university–industry linkages. *Res. Policy* 27 (8), 823–833.
- Fini, Riccardo, Meoli, Azzurra, Sobrero, Maurizio, 2022. University graduates' early career decisions and interregional mobility: self-employment versus salaried job. *Reg. Stud.* 56 (6), 972–988.
- Gompers, Paul A., Kovner, Anna, Lerner, Josh, Scharfstein, David, 2010. Performance persistence in entrepreneurship. *J. Financ. Econ.* 96 (1), 18–32.
- Gompers, Paul A., Lerner, Josh, 2000. *The Money of Invention*. Harvard Business School Press.
- Gompers, Paul A., Lerner, Josh, Scharfstein, David, 2005. Entrepreneurial spawning: public corporations and the genesis of new ventures, 1986 to 1999. *J. Financ.* 60 (2), 577–614.
- Gompers, Paul A., Mukharlyamov, Vladimir, Xuan, Yuhai, 2016. The cost of friendship. *J. Financ. Econ.* 119 (3), 626–644.
- Gornall, Will, Strebulaev, Ilya A., 2015. The Economic Impact of Venture Capital: Evidence From Public Companies. Stanford Graduate School of Business Working.
- Grogger, Jeffrey, Hanson, Gordon H., 2015. Attracting talent: location choices of foreign-born PhDs in the United States. *J. Labor Econ.* 33 (S1), S5–S38.
- Guerrero, Maribel, Cunningham, James A., Urbano, David, 2015. Economic impact of entrepreneurial universities' activities: an exploratory study of the United Kingdom. *Res. Policy* 44 (3), 748–764.
- Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2001. "The NBER Patent Citation Data File: Lessons, Insights And Methodological Tools." NBER Working Paper 8498.
- Hall, Bronwyn H., Jaffe, Adam, Trajtenberg, Manuel, 2005. Market value and patent citations. *RAND J. Econ.* 16–38.
- Hanson, Gordon H., Slaughter, Matthew J., 2017. High-skilled immigration and the rise of STEM occupations in US employment. In: *Education, Skills, And Technical Change: Implications for Future US GDP Growth*. University of Chicago Press, pp. 465–494.
- Hausman, Naomi, 2020. University innovation and local economic growth. *Rev. Econ. Stat.* 1–46.
- Hochberg, Yael V., Ljungqvist, Alexander, Yang, Lu., 2007. Whom you know matters: venture capital networks and investment performance. *J. Financ.* 62 (1), 251–301.
- Howell, Sabrina T., Lerner, Josh, Nanda, Ramana, Townsend, Richard, 2020. "Financial Distancing: How Venture Capital Follows the Economy Down And Curtails Innovation." NBER Working Paper 27150.
- Hunt, Jennifer, Gauthier-Loiselle, Marjolaine, 2010. How much does immigration boost innovation? *Am. Econ. J. Macroecon.* 2 (2), 31–56.
- Hunt, Jennifer, 2011. Which immigrants are most innovative and entrepreneurial? Distinctions by entry visa. *J. Labor Econ.* 29 (3), 417–457.
- Israel, Emma, Batalova, Jeanne, 2021. *International Students in the United States*. Migration Policy Institute.
- Kerr, William R., 2020. The gift of global talent: innovation policy and the economy. *Innov. Policy Econ.* 20 (1), 1–37.
- Kerr, William R., 2010. Breakthrough inventions and migrating clusters of innovation. *J. Urban Econ.* 67 (1), 46–60.
- Kerr, Sari Pekkala, Kerr, William R., 2016. Immigrant entrepreneurship. In: *Measuring Entrepreneurial Businesses: Current Knowledge And Challenges*. University of Chicago Press, pp. 187–249.
- Kerr, Sari Pekkala, Kerr, William, 2020. Immigrant entrepreneurship in America: evidence from the Survey of Business Owners 2007 & 2012. *Res. Policy* 49 (3), 103918.
- Kerr, William R., Shihe, Fu., 2008. The survey of industrial R&D—patent database link project. *J. Technol. Transf.* 33 (2), 173–186.
- Kerr, William R., Lincoln, William F., 2010. The supply side of innovation: H-1B visa reforms and US ethnic invention. *J. Labor Econ.* 28 (3), 473–508.
- Kogan, Leonid, Papanikolaou, Dimitris, Seru, Amit, Stoffman, Noah, 2017. Technological innovation, resource allocation, and growth. *Q. J. Econ.* 132 (2), 665–712.
- Lee, Yong Suk, Eesley, Chuck, 2018. The persistence of entrepreneurship and innovative immigrants. *Res. Policy* 47 (6), 1032–1044.
- Masumba, Mark, Jin, Yanhong H., Mjelde, James W., 2011. Factors influencing career location preferences of international graduate students in the United States. *Educ. Econ.* 19 (5), 501–517.
- Reynolds, Paul, Storey, David J., Westhead, Paul, 1994. Cross-national comparisons of the variation in new firm formation rates. *Reg. Stud.* 28 (4), 443–456.
- Rosenzweig, Mark R., Irwin, Douglas A., Williamson, Jaffrey G., 2006. Global wage differences and international student flows [with comments and discussion]. In: *Brookings Trade Forum*. Brookings Institution Press, pp. 57–96.
- Ruiz, Neil G., 2014. The geography of foreign students in US higher education: origins and destinations. In: *Report*, Global Cities Initiative.



- Shumilova, Yulia, Cai, Yuzhuo, 2016. Three approaches to competing for global talent: role of higher education. In: *Global Perspectives And Local Challenges Surrounding International Student Mobility*. IGI Global, pp. 114–135.
- Sorenson, Olav, Audia, Pino G., 2000. The social structure of entrepreneurial activity: geographic concentration of footwear production in the United States, 1940–1989. *Am. J. Sociol.* 106 (2), 424–462.
- Stephan, Paula E., 2010. The “T” s have it: immigration and innovation, the perspective from academe. *Innov. Policy Econ.* 10 (1), 83–127.
- Stephan, Paula E., Levin, Sharon G., 2001. Exceptional contributions to US science by the foreign-born and foreign-educated. *Popul. Res. Policy Rev.* 20 (1), 59–79.
- Stuart, Toby, Sorenson, Olav, 2003. The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms. *Res. Policy* 32 (2), 229–253.
- Tartari, Valentina, Stern, Scott, 2021. More Than an Ivory Tower: The Impact of Research Institutions on the Quantity And Quality of Entrepreneurship. No. w28846. National Bureau of Economic Research.
- Uhlbach, Wolf-Hendrik, Tartari, Valentina, Kongsted, Hans Christian, 2022. Beyond scientific excellence: international mobility and the entrepreneurial activities of academic scientists. *Res. Policy* 51 (1), 104401.
- Wadhwa, Vivek, Freeman, Richard, Rissing, Ben, 2010. Education and tech entrepreneurship. *Innov.: Technol.Gov.Glob.* 5 (2), 141–153.
- Wadhwa, Vivek, Rissing, Ben A., Saxenian, Anna Lee, Gereffi, Gary, 2007. “Education, entrepreneurship and immigration: America's new immigrant entrepreneurs, Part II.” Part II (June 11, 2007).
- Wadhwa, Vivek, Saxenian, Anna Lee, Rissing, Ben, Gereffi, Gary, 2007. America's new immigrant entrepreneurs. In: *Kauffman Foundation Report*.