An Analysis of Employment Dynamics at Venture-Backed Companies Between 1990 and 2020

Gregory W. Brown, David Fisher, Lu Yi
University of North Carolina Kenan Institute of Private Enterprise & Research

Michael J. Chow
National Venture Capital Association & Venture Forward

February 2022

In this study, we build upon the existing body of literature focusing on the relationship between young companies and job creation by examining the employment dynamics of over 67,000 U.S. companies that received venture capital (VC) investment between 1970 and 2020. Importantly, we map the location of 3.8 million employees working at these 67,000-plus companies in 2020, the first attempt (that we are aware of) to analyze the geographic dispersion of employees working for VC-backed companies. We also analyze the employment dynamics of VC-backed companies between 1990 and 2020 (the range of years for which our employment data allow such analysis). Our key findings are as follows:

- **VC-backed jobs are distributed broadly across the entire U.S.**
  - 62.5% of VC-backed jobs are outside the states of California, Massachusetts, and New York.
  - This geographic footprint of employment is in contrast to how VC investments are distributed across the nation, with 73% of VC dollars invested going to those three states in 2020.
- **Employment at VC-backed companies grows much faster than jobs at non-VC-backed companies, consistent with research on employment trends at young, high-growth companies.**
  - Employment at our set of VC-backed companies grew at roughly eight times the pace of employment at non-VC-backed companies.
    - The annualized growth rate of employment at VC-backed companies in our dataset between 1990 and 2020 is 8.2%.
    - For total private sector employment, the growth rate between January 1990 and February 2020 is just 1.1%.
    - Employment at our set of VC-backed companies grew 960% between 1990 and 2020.
    - In comparison, total private sector employment increased by 40% between January 1990 and February 2020.
- **Employment growth at VC-backed companies is resilient with strong positive growth rates observed regardless of the stage of the business cycle.**
  - Even after the 2007-2008 financial crisis and during the Great Recession, annual job growth at our set of VC-backed companies exceeded 4.0%.
  - In contrast, total private sector employment shrank by 4.3% in 2009, contributing to a decrease of 7.4 million private sector jobs during the Great Recession.

Our findings are consistent with job growth at VC-backed companies playing a key role in the known important contributions that young, high-growth companies make to net job creation in the U.S. As well, our results indicate that VC investment is a catalyst for job creation in not only a select few metropolitan hubs, but across the entire nation. From a policy perspective, these findings should prompt officials focused on job creation to devise and promote policies that encourage more VC activity in the U.S. and permit the innovation economy to prosper.

---

*We thank PitchBook and Professor Jay Ritter of the University of Florida for assistance in identifying VC-backed companies. The opinions expressed in this paper are solely those of the authors. All errors are our own.

Introduction

The VC industry plays a vital role in the U.S. economy. VC investors make early high-risk investments in young, innovative startups developing leading edge technologies that show great promise in revolutionizing business processes, and in so doing increasing productivity and efficiency, and otherwise improving standards of living. VC investors take these financial risks with hopes that a fraction of these startups (most of which will fail) will yield a healthy return on investment either by being acquired by another company or through a listing on a public stock exchange. Many of America's most iconic companies received venture financing during their incipient stages, including the six largest public U.S. companies by market capitalization: Apple, Microsoft, Alphabet, Amazon, Tesla, and Meta.[1] The very strong financial returns of VC as an asset class in recent years also make venture capital important for the limited partners, such as university endowments, foundations, and public pension plans, who provide the necessary capital to VC firms to fund promising startups.

The importance of the VC industry has resulted in a substantial body of research examining various aspects of the industry. Much of this research focuses on the investment and fundraising behavior of industry participants as well as the financial performance of individual firms, funds, or VC as an asset class. Less studied are the business dynamics of the portfolio companies and the contributions the VC industry makes to the broader economy. This is somewhat surprising since many of the nation's most influential companies, which touch the lives of virtually all Americans in myriad ways, at some point received venture financing. A recent paper by Gornall and Strebulaev (2021) might provide the best effort to comprehensively capture the overall significance of VC-backed companies.[2] Among Gornall and Strebulaev's findings are that VC-backed companies[3] constitute 50% of all U.S. public companies, 77% of total market capitalization on U.S. public stock exchanges, take in $2.4 trillion in annual revenue, have annual net income of $257 billion, and employ 4.7 million employees.

These are important figures that underline the importance of VC-backed companies to the U.S. public equity and labor markets. This latter relationship between venture capital and labor markets may be of particular interest to empirical researchers that approach the industry with a broader lens than ones solely focused on financial returns and investment, exit, and fundraising activity. This is because a considerable body of literature has found that young companies are essential to net job creation in the United States.[4] In particular, researchers have identified that *high-growth* young companies disproportionately contribute to job creation in the U.S. Such companies are also referred to as "knowledge-intensive" companies in the literature, which include startups that receive VC investment.[5]

[1] As of January 27, 2022. Source: TradingView.
[2] Gornall, Will and Ilya A. Strebulaev, "The Economic Impact of Venture Capital: Evidence from Public Companies (June 2021). Available at SSRN: https://ssrn.com/abstract=2681841 or https://dx.doi.org/10.2139/ssrn.2681841.
[3] Gornall and Strebulaev examine only public companies that previously received venture financing and limit their analysis to a post-ERISA sample of companies that were founded after 1968 and went public after 1978. They exclude analysis of private companies that received venture financing but were never listed on a public exchange.
[4] For example, see:

Haltiwanger, John, Ron S. Jarmin, Robert Kulick, and Javier Miranda, "High Growth Young Firms: Contribution to Job, Output, and Productivity Growth," Measuring Entrepreneurial Businesses: Current Knowledge and Challenges (University of Chicago Press: September 2017).

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda, "Who Creates Jobs? Small Versus Large Versus Young," The Review of Economics and Statistics 95, no. 2 (May 2013): 347-361.

[5] Crews, Jonas, Ross DeVol, Richard Florida, and Dave Shideler, "Young Firms and Regional Economic Growth: Knowledge-Intensive Entrepreneurs Critical," Heartland Forward, May 2020.

Data and Methodology

These facts motivate our current research which attempts to advance knowledge of the relationship between VC-backed companies and employment dynamics in two primary ways. First, we estimate the geographic distribution of jobs at U.S. companies that have received venture financing. Second, we examine the employment dynamics of jobs at VC-backed companies over time and see if the observed trends are consistent with the literature on job creation at young, high-growth companies. To achieve our goals, we utilize four different data sources:

1. A dataset of 66,318 companies, courtesy of PitchBook, that received VC investment between 2000 and 2020.
2. A dataset of 3,876 companies, courtesy of Professor Jay Ritter of the University of Florida, that received venture financing but underwent an IPO. The companies in this dataset date back to approximately 1970.
3. Data from the National Establishment Time Series (NETS) database which contains longitudinal administrative data on employment for companies at the establishment level dating back to 1990.
4. Wharton Research Data Services.

A detailed description of our methodological approach is provided in the appendix to this paper, but we provide an overview here. To the extent possible, we utilized DUNS Numbers (unique 9-digit identification numbers provided by Dun & Bradstreet) to link company records from the PitchBook and University of Florida datasets with same-company records from the NETS database. The PitchBook dataset included 33,095 companies that had matching DUNS Numbers in the NETS data, among which 31,100 companies have employment data when they appeared as headquarters. For these companies, we adopted the longitudinal employment data in the NETS data as the actual employment over time for these companies. Among the remaining 35,218 PitchBook companies for which we could not identify a matching DUNS Number, we imputed employment data based on company industry and age by estimating a panel regression with industry and age fixed effects. Of the 3,876 companies in the University of Florida dataset, we acquired DUNS Numbers for over 1,400 of these companies through Wharton Research Data Services data files. We matched these companies with the corresponding data in the NETS database and, as with the other companies with matching DUNS Numbers, accessed the longitudinal employment data in the NETS dataset. We did not impute employment for companies in the University of Florida dataset for which we could not acquire DUNS Numbers.

Both the PitchBook and NETS datasets include a zip code field associated with each company or establishment in their respective datasets. However, neither dataset includes a similar field for the congressional district in which a company or establishment is located. Allocating a company's employment at a particular establishment to local congressional districts consequently requires a mapping algorithm to impute geographic dispersion of employment for companies without a DUNS number. We first mapped companies with matching NETS data to congressional districts based on their zip code using the U.S. Department of Housing and Urban Development's USPS Zip Code Crosswalk File. Using this first set of mapped companies, we subsequently constructed a mapping matrix of employment by company by congressional district and applied the aggregated geographic dispersion to the companies from the PitchBook dataset for which we imputed employment, thus allocating those jobs to the 435 congressional districts with voting members as well as Washington, DC and Puerto Rico. The result is a dataset with employment allocations to congressional districts for over 67,000 companies derived directly from the NETS data and imputed values based the large sample of firms with NETS data. For more details on our methodology, please refer to the **Appendix**.

Key Findings

One of the key findings from our analysis of estimated employment at VC-backed companies is that VC-backed jobs are spread broadly across the entire nation. We find that 62.5% of employment at VC-backed companies in 2020 was in states other than California, Massachusetts, or New York. This pattern stands in contrast to the geographic dispersion of VC investment in the U.S., which is more heavily concentrated in those three hub states. In 2020, 73% of VC dollars invested in the U.S. occurred in California, Massachusetts, or New York. This geographic concentration of investment has been the norm for decades: since 2004, the annual share of total VC investment attributed to these three states has never fallen below 59% (**Chart 1**).[6]

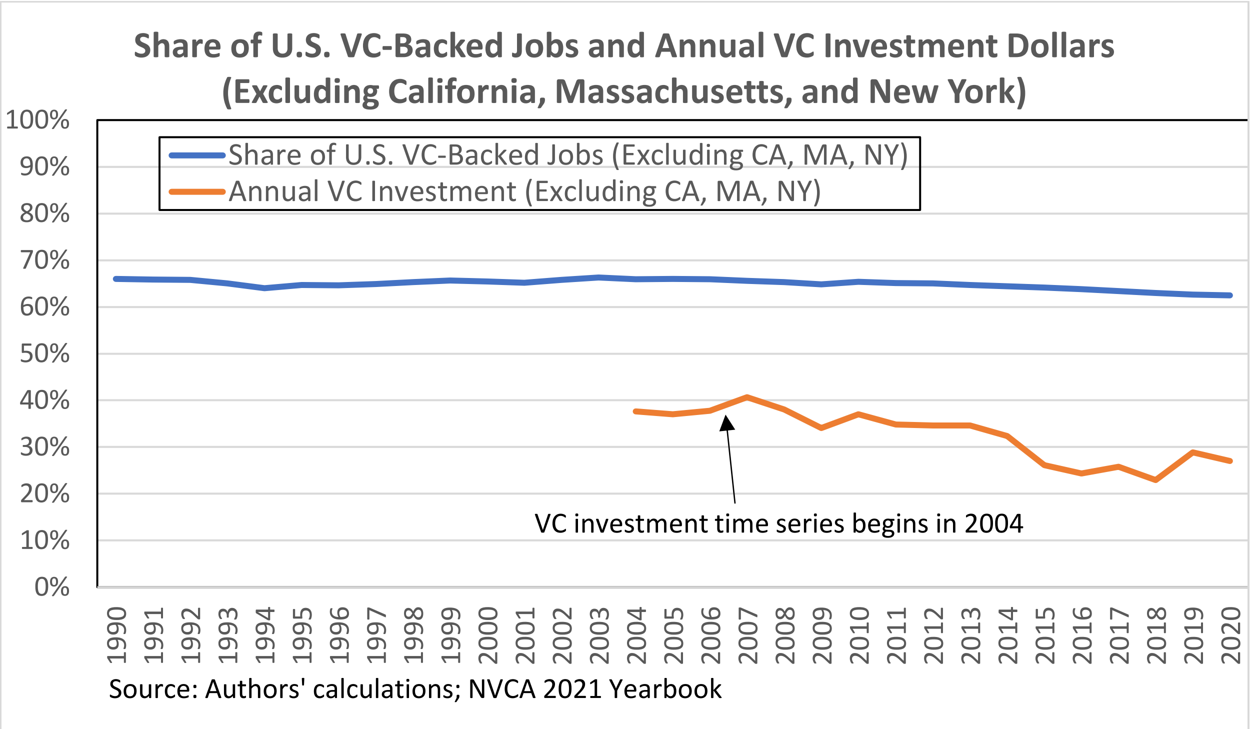

Chart 1

Our results also reinforce the findings of prior research that young, knowledge-intensive companies increase their labor force more rapidly than other young, so-called "Main Street" businesses which often lack similar ambitions to evolve into large companies. Job growth at young, knowledge-intensive companies is important because such businesses are responsible for most aggregate job creation. This occurs despite the fact that as many as 50% of jobs at startups will be lost within the first five years due to business exits. Put differently, higher than average job creation at high-growth companies like the ones we examine compensates for the majority of losses associated with an annual startup cohort, so that the cohort retains 80% of its original employment after five years.[7,8] Our cohort of VC-backed companies

[6] NVCA 2021 Yearbook.
[7] Crews et al.
[8] Horell, Michael and Robert Litan, "After Inception: How Enduring is Job Creation by Startups?", Ewing Marion Kauffman Foundation, September 9, 2010.

demonstrate much faster employment growth compared to total private sector employment during the period between 1990 and 2020.[9] In particular, we find:

- The annualized growth rate of employment at VC-backed companies in our dataset between 1990 and 2020 is 8.2%.
- For total private sector employment (as published by the U.S. Bureau of Labor Statistics), the average annual growth rate between January 1990 and February 2020 is 1.1%.
- Employment at our set of VC-backed companies grew 960% between 1990 and 2020.
- In comparison, total private sector employment increased by 40% between January 1990 and February 2020.

Chart 2 demonstrates the large gap in employment relative to equivalent initial starting points in employment levels that result from the annualized growth rate we discovered at VC-backed companies (8.2%) versus the growth rate for total private sector employment (1.1%). Beginning with an index value of 100, a growth rate of 8.2% applied for 30 consecutive years (consistent with our period of analysis ranging from 1990 to 2020) grows to 1,060. By comparison, a growth rate of 1.1% applied to an initial index value for 30 consecutive years grows to only 140. Put another way, if one observed two groups of equally-sized employees—one that worked only for VC-backed companies and the other working for a group of companies representative of the broader universe of private employers—after 30 years, one could expect that employment at the first group of VC-backed companies to grow to a number roughly eight times larger than employment at the second set of companies.

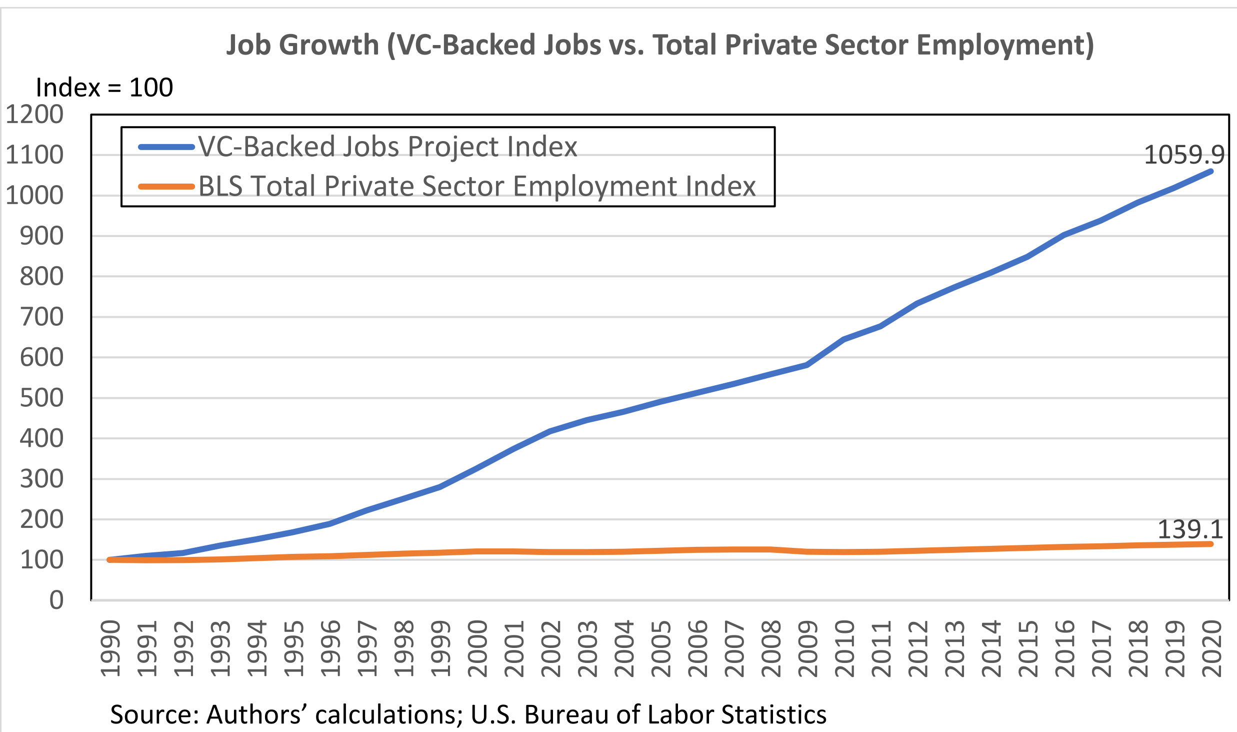

Chart 2

[9] While our set of VC-backed companies include companies that received venture financing as far back as 1970, our analysis of employment dynamics is limited to the years 1990 to 2020 due to data limitations associated with the NETS dataset (the NETS employment data only date back to 1990).

A third key finding of our analysis is that job creation at VC-backed companies is highly resilient and largely insulated to fluctuations in the business cycle. **Chart 3** shows the year-over-year percent change in employment at our cohort of VC-backed companies compared to total private sector employment. Consistent with our earlier findings on annualized growth rates of employment, the average year-over-year percent change for employment at our set of VC-backed companies between 1990 and 2020 was 8.3%. In contrast, the average year-over-year percent change for total private sector employment was only 1.1%.

This should be expected given our earlier results. Of more interest is that job growth at VC-backed remained robust and positive every single year between 1990 and 2020, ranging from a low of 3.7% to a high of 17.8%. This means that VC-backed companies increased employment at a fast clip even during and after (unemployment is a lagging indicator) the recessions of 1990:Q2 to 1991:Q1, 2001:Q1 to 2001:Q4, and the Great Recession spanning 2007:Q4 to 2009:Q2. While we do notice some dips in the rate of growth following the 1990-1991 and 2001 recessions, job growth at our set of VC-backed companies never fell below 4% during the Great Recession and actually surged to 10.9% in 2010 following the end of, what was then, the sharpest U.S. economic contraction since the Great Depression.

Such buoyant performance throughout even the contractionary phases of the boom-bust cycles of the economy stands in contrast to the behavior of total private sector employment. As Chart 3 indicates, U.S. private sector employment shrank during each of the three aforementioned recessions. The sharpest year-over-year dip occurred in 2009, coincidental with the Great Recession when total private sector employment shrank by 4.3% for the year, contributing to a decrease of 7.4 million private sector jobs during that bleak episode of U.S. economic history.

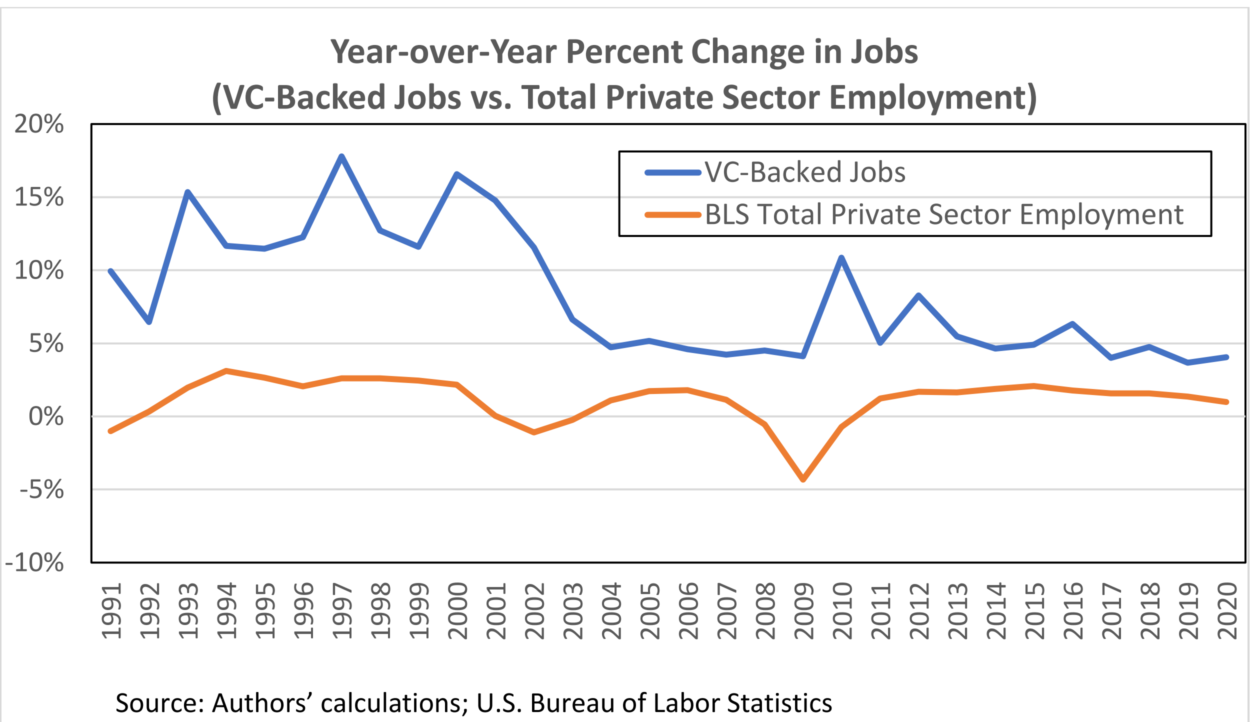

Chart 3

VC-Backed Employment Dynamics by Geography

Finally, we offer some insights on the geographic dimensions of our analysis of employment dynamics at our set of VC-backed companies. As mentioned previously, VC-backed companies appear to create jobs broadly across the nation despite the concentration of VC investment in a select few states and metropolitan locations (primarily in California, Massachusetts, and New York). Despite typically receiving less than 30% of total annual VC investment in the U.S.,[10] states apart from California, Massachusetts, and New York (so-called "non-hub" states) have accounted for at least 62% of jobs at VC-backed companies every single year dating back to 1990 (see Chart 1 above).

A breakdown of each state's share of total employment in 2020 for our set of VC-backed companies is provided in **Table 1** below.[11] We also examine annualized growth rates in VC-backed employment for the last 5- and 10-year periods. We find that VC-backed jobs are located not only in all 50 states, but also all 435 congressional districts with voting members as well as Washington, DC and Puerto Rico. Excluding California, Massachusetts, and New York, states with noticeable shares of employment (at least 2.5% of total employment) include Colorado, Florida, Georgia, Illinois, New Jersey, Pennsylvania, Texas, Virginia, and Washington. States exhibiting the most rapid growth rates in VC-backed employment in recent years (at least an annualized growth rate of 4.0% since 2015) include Arizona, California, Connecticut, Delaware, the District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Maine, Massachusetts, Missouri, Montana, Nevada, New York, North Carolina, Pennsylvania, South Dakota, Tennessee, Texas, Utah, Washington, and Wyoming.

Heatmaps of employment at VC-backed companies in our dataset are provided in **Figure 1** and **Figure 2** below. We map the geographic distribution of employment across the U.S. by both congressional district and by state. Readers interested in exploring the geography of employment at VC-backed companies in a more dynamic format may do so by visiting our interactive online dashboard available on the websites of the National Venture Capital Association, here, or Venture Forward, here.

Concluding Remarks

Our analysis of over 67,000 U.S. companies that received VC investment and which employ 3.8 million workers indicates that VC-backed companies create jobs broadly across the entire nation and play a key role in the known important contributions that young, high-growth companies make to net job creation in the U.S. The much quicker growth rates in employment and the resilience of such growth at VC-backed companies suggest that encouraging greater VC investment may be a boon to overall U.S. employment and ought to be of interest to policymakers interested not only in creating jobs in general, but in creating durable jobs that can weather the vicissitudes of economic cycles. Our findings suggest that the promotion of policies which encourage more VC activity in the U.S. and permit the innovation economy to prosper will benefit the U.S. labor market and economy.

[10] According to data from the NVCA 2021 Yearbook, startups in California, Massachusetts, and New York have received at least 70% of all VC investment in the U.S. since 2015. Between 2004 and 2014, startups in these three states never received less than 59% of all VC investment in the U.S.

[11] We also include the District of Columbia and Puerto Rico which, per our analysis, are home to a number of VC-backed jobs.

Table 1: Percent Shares (in 2020) and Annualized Growth Rates (Between 2010 and 2020) of VC-Backed Jobs, by State or Region

State/Region	% of VC-Backed Jobs in 2020	Annualized % Change Between 2010 and 2020	Annualized % Change Between 2015 and 2020
Alabama	0.9%	1.9%	2.2%
Alaska	0.1%	3.7%	2.3%
Arizona	1.7%	5.4%	4.9%
Arkansas	0.3%	4.9%	3.9%
California	24.7%	6.0%	5.3%
Colorado	3.3%	4.1%	3.6%
Connecticut	1.0%	4.5%	4.0%
Delaware	0.3%	8.1%	8.8%
District of Columbia	0.3%	6.7%	5.2%
Florida	4.0%	5.6%	4.7%
Georgia	2.6%	5.3%	4.1%
Hawaii	0.2%	6.2%	4.6%
Idaho	0.3%	6.7%	6.8%
Illinois	3.3%	4.7%	4.4%
Indiana	1.5%	4.2%	4.3%
Iowa	0.5%	5.1%	3.1%
Kansas	0.6%	3.2%	2.6%
Kentucky	0.8%	4.0%	3.4%
Louisiana	0.6%	3.6%	2.0%
Maine	0.2%	5.9%	5.3%
Maryland	1.9%	3.6%	3.3%
Massachusetts	6.2%	4.5%	4.4%
Michigan	1.6%	4.7%	3.5%
Minnesota	1.6%	4.3%	3.6%
Mississippi	0.5%	3.1%	1.1%
Missouri	1.5%	4.6%	4.7%
Montana	0.1%	5.9%	5.9%
Nebraska	0.4%	3.9%	2.7%
Nevada	0.6%	5.5%	5.0%
New Hampshire	0.5%	3.8%	2.6%
New Jersey	2.8%	3.5%	2.9%
New Mexico	0.3%	4.3%	3.6%
New York	6.6%	7.3%	7.3%
North Carolina	2.0%	4.5%	4.1%
North Dakota	0.1%	3.4%	2.4%
Ohio	2.4%	4.1%	3.7%
Oklahoma	0.6%	5.2%	3.7%
Oregon	1.4%	4.7%	3.4%
Pennsylvania	3.4%	4.4%	4.1%
Puerto Rico	0.2%	2.7%	2.2%
Rhode Island	0.3%	4.1%	3.6%
South Carolina	0.7%	4.7%	3.2%
South Dakota	0.2%	8.9%	4.6%
Tennessee	1.6%	5.1%	5.2%
Texas	7.3%	5.0%	4.1%
Utah	1.1%	7.1%	7.5%
Vermont	0.1%	5.8%	3.3%
Virginia	2.8%	4.2%	3.1%
Washington	3.0%	5.5%	5.2%
West Virginia	0.1%	1.8%	1.1%
Wisconsin	0.9%	4.5%	3.5%
Wyoming	0.1%	4.7%	5.5%

VC-Backed Jobs by Congressional District in 2020

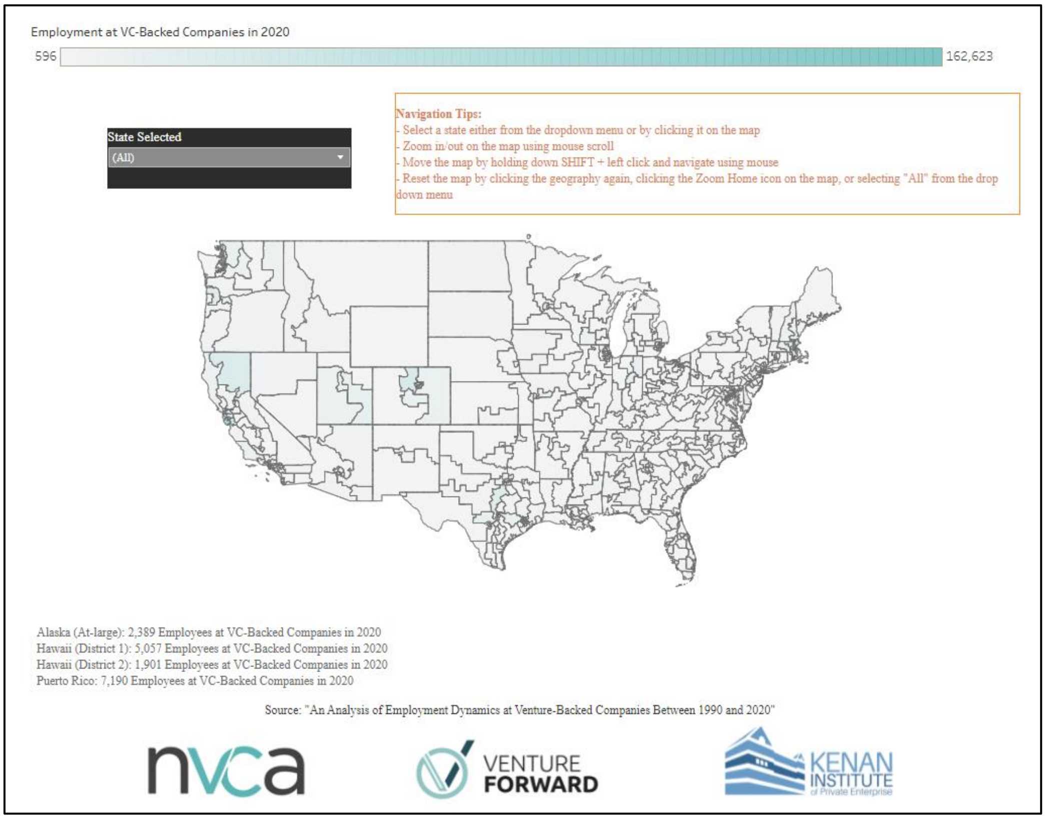

Figure 1

VC-Backed Jobs by State in 2020

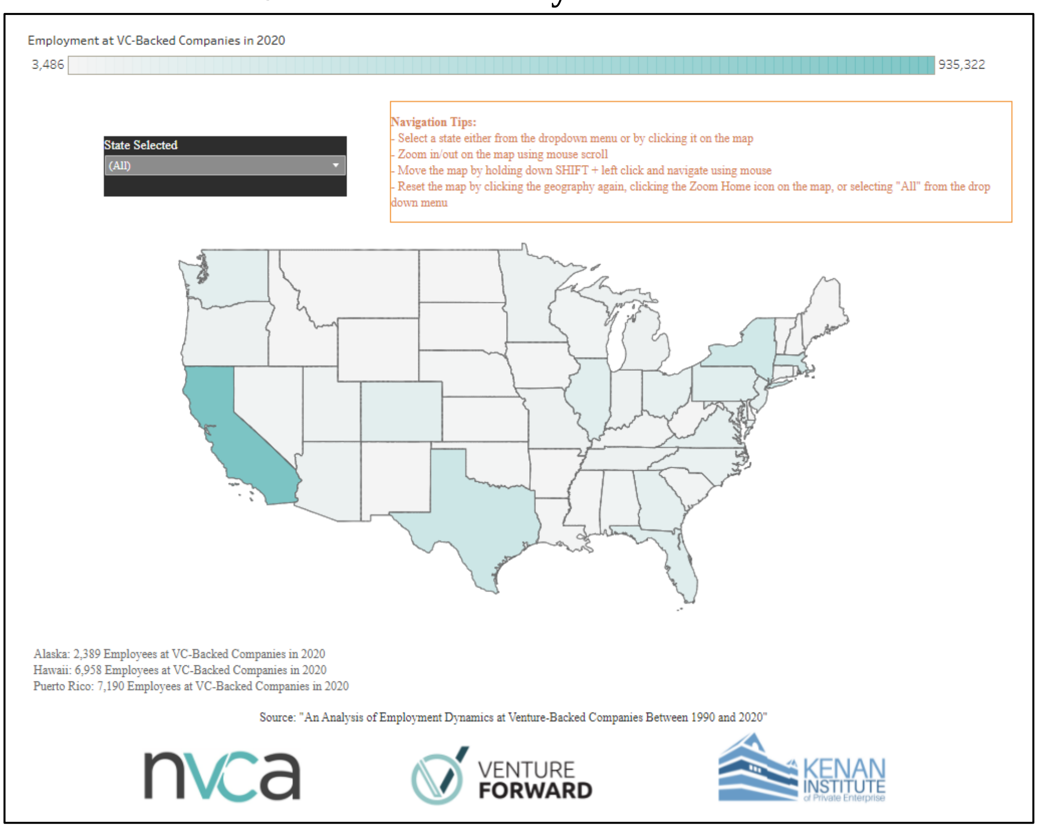

Figure 2

Appendix: Methodology

Overview

The purpose of this project is to estimate employment of VC-backed companies headquartered in the United States at the national, state, and congressional district level. To accomplish this task, we utilized three datasets: (1) a list, courtesy of PitchBook, of 66,318 companies that received venture capital investment from 2000 to the present; (2) a list, courtesy of Professor Jay Ritter at the University of Florida, of 3,876 VC-backed companies that underwent an initial public offering from approximately 1970 onward; and (3) a dataset consisting of a subset of the National Establishment Time Series (NETS) dataset that contains employment time series data dating back to 1990 for establishments linked to the companies contained in the PitchBook or the University of Florida datasets. By matching the companies contained in the PitchBook and the University of Florida datasets with same-company data from the NETS dataset, we are able to link high-quality administrative data on company employment disaggregated at very fine levels of geographic detail.

Matching Company Records Between PitchBook and NETS Datasets

The first stage in this project consisted of matching company records between the PitchBook and NETS datasets. This required an initial identification step which consisted of identifying DUNS Numbers (unique 9-digit identification numbers provided by Dun & Bradstreet and the main identifiers in the NETS data) for each PitchBook company. Identifying DUNS Numbers for companies is a vital component of this project, since DUNS Numbers serve as the unique identifier that links company records in the PitchBook and NETS datasets. Optimally, we would obtain DUNS Numbers for all companies in the PitchBook dataset which could then be linked to employment administrative data in the NETS dataset. This task was not straightforward for two primary reasons:

First, the data did not cleanly match up. For example, company names rarely matched exactly and sometimes differed substantially between PitchBook and NETS. Or, in the case where company names were similar across the two datasets, the datasets may have had different zip codes or cities, raising doubts about the match quality. Additionally, one dataset might refer to the original name of a VC-backed startup while the other referred to a corporation that acquired and now owns the startup. Few records, even those we confirmed as positive matches, matched on every characteristic between the two datasets.

The second obstacle was that the NETS dataset is enormous. The sheer size of the NETS dataset rendered manual hand-matching between the two datasets to be practically useless due to long filtering times. It also makes “all versus all” comparisons, or text comparisons between every company name in NETS with every company name in our PitchBook sample, computationally infeasible.

While perfect matches were rare, strong matches between “features” (characteristics of the data such as names, location, phone numbers, etc.) indicate higher likelihood of a business entity match. While name similarity on its own may not demonstrate an entity match, as additional data characteristics are added to the matching process, our confidence of match, or non-match, increased.

We used several different data characteristics from the datasets (City, Zip Code, Phone Number, Phone Area Code, SICS classification, NAICS classification, and Founding Year) to identify subsets that were potential match candidates. We then used name similarity to identify matches between companies. We used multiple different company names when available (common names, trade names, legal names) to improve match quality.

In our final sample, we obtained about 43,000 DUNS Numbers for the PitchBook sample. Of these, around 30,000 were provided from the original PitchBook set, and the remainder were obtained from name matching.

Deeper Dive into Matching Processes

The NETS dataset contains about 71.5 million establishments. Our PitchBook dataset contains a much more modest 66,318 companies, implying approximately 4.8 trillion possible matches in an "all names versus all names" comparison between the two datasets, more if we consider multiple possible names (trade, common, legal) for each business. To make the matching computationally feasible, we identified common characteristics, such as zip code, that allowed us to get subsets of data that were more likely to contain matches. For instance, when we filter by a single zip code, we may get 20 PitchBook companies and 30,000 NETS entries, requiring 30,000 times 20, or 600,000 comparisons for that zip code, rather than the trillions from the 71.5 million against 66,318 comparisons.

Zip Code Filter

Our first filter utilized zip codes. We filtered the NETS data to any zip codes either equivalent or adjacent to a PitchBook entry, and then performed a name match algorithm on the subset of companies that share a zip code. This proved to be our most successful filter. We found that high accuracy scores strongly indicate a match. In testing this filter, we also discovered that many companies in the PitchBook dataset did not match well with any of the NETS companies in that same zip code or even adjacent zip codes. This is most likely due to headquarter differences between datasets (NETS focuses on where workers are located, not where a company is registered), and changes of names via acquisition or rebranding.

Phone Number Filter

Our second filter utilized phone numbers. We compared companies that had the same first 6-digits in their phone numbers and provided match scores based on similarity between phone numbers, with the highest scores going to perfect matches. We found that identical phone numbers generally produced high-quality matches, but less-than-identical phone numbers resulted in quite weak outcomes. This step was often useful more for identifying parent companies which we could then match based on other characteristics.

Industry Filters

Our third filter utilized industry activity. The PitchBook dataset contains information on the industry of each startup company, while the NETS search database has both NAICS and SIC codes. The PitchBook descriptions often matched fairly poorly to the NAICS/SIC codes, but we used "true" matches (PitchBook entries with known DUNS Numbers) to create a crosswalk between PitchBook descriptions and NAICS/SICS codes as another potential filter for the datasets, again matching companies by name afterwards.

Cleaning and Matching Features

Both datasets of PitchBook companies (the one with DUNS Numbers and the other consisting of unmatched companies) were combined with NETS data at some point during the project to evaluate the accuracy of matches. The different fields examined, and an overview of our process, are provided below.

- **Company Name –** This was our primary matching tool. We started with the most commonly filled-in name fields in each dataset and filled in empty values from other name columns. For instance, if "Company Name" was blank in PitchBook, we filled in the value for this variable using a combination of the "Common Name" and "Official Name" variables. We then removed a limited number of stop words (e.g., "corporation") to improve matching quality. Name fields were then vectorized into 3-letter ngrams for fast string comparison with other names. Finally, strings were compared using a Cosine similarity algorithm.
- **State** – We used a tool to standardize to the two-letter abbreviations and performed a simple equality match.
- **City** – An equality match was performed to determine whether city names were identical. We opted not to use a string comparison for this, worrying that we would give too much credit when two

different cities had very similar names. Instead, we used a simple equality to perform this check (PitchBook city equals NETS city).

- **Zip –** We checked to see if a PitchBook zip code matched either exactly to the same zip code provided in the NETS data or to a neighboring zip code in the NETS data.
- **Phone Number & Area Code** – Checks on phone numbers and area codes were generally less reliable. This derived from a lack of one-to-one matches between the PitchBook and NETS datasets. When we examined phone numbers by DUNS Number within the NETS database, we found that companies regularly had multiple numbers—sometimes up to 12 numbers for one company in the database. If any of these multiple phone numbers within NETS matched the PitchBook number directly, then we scored it as a full phone match.
- **NAICS and SIC Classifications** – Rather than provide a NAICS or SIC code as the NETS dataset does, PitchBook provides—in levels of increasing detail, respectively—classifications as to a business's industry through the Industry Sector, Industry Group, and Industry Code variables. The disparate classification systems mean there is no clean crosswalk from either NAICS or SIC to PitchBook industry category. Many NAICS or SIC codes could reasonably be mapped to several different Industry Groups and Industry Sectors. For example, a single NAICS code in the NETS data could match to everything from computers to durable goods in the PitchBook dataset. As a workaround, we recorded every match between Industry Code and NAICS or SIC in our positive match set, creating a type of fuzzy crosswalk. We then employed this crosswalk on proposed matches. This allowed us to determine whether the Industry Code/NAICS or Industry Code/SIC combination has been seen before. If the combination has been seen, then we set the corresponding binary variables (NAICS for a NAICS match in the crosswalk, SIC for a SIC match) to true.
- **Founding Year –** The absolute value of the difference in founding years (NETS versus PitchBook) was used as an input to the matching algorithm.

Matching Records Between University of Florida and NETS Datasets

In addition to the list of 66,318 companies that we obtained from PitchBook, we also acquired a list of 3,876 VC-backed companies that underwent an initial public offering from approximately 1970 onward from Professor Jay Ritter at the University of Florida. To the extent that these companies still exist independently today (some may have been acquired while others may have gone out of business), they are likely larger than the typical company in the PitchBook dataset, the vast majority of which are still privately held and have never experienced a liquidity event.

While the data provided in the University of Florida dataset did not include our typical matching information (Zip Code, Phone Number, NAICS/SIC, etc.), these were substantially better-known companies, and for many we could find matching information within the Compustat database accessed through the Wharton Research Data Services (WRDS). We then matched similarly to the PitchBook dataset. We ended up cleanly matching over 1,400 of the companies in the University of Florida dataset to approximately 3,700 different establishments. The DUNS Numbers for these companies were also included within the list of DUNS Numbers we obtained from NETS.

Acquiring NETS Employment Data

Once we had a combined set of DUNS Numbers from PitchBook and the University of Florida dataset, we accessed the related NETS data to obtain employment administrative data. This was done by matching to headquarters, acquiring companies, related establishments, and grabbing all subsidiaries of that new, larger list. This gave us a list of around a million establishments, with comprehensive employment information, NAICS/SIC codes, and location information by year for each establishment. With these data, we filtered back, dropping any results that were not directly owned or related to originally VC-backed startups. This insured that we were not counting unrelated employment of non-VC-backed acquiring companies in our totals, but instead were only including employment at VC-backed companies. (E.g., assume our data included a non-VC-backed parent company with several subsidiary businesses that were acquired over time but which did not begin as VC-backed startups, and that this parent company also acquired one VC-backed startup. We only counted the employment associated with the VC-backed startup, which may have multiple establishments across the U.S.)

We also made some minor adjustments and cleaned the dataset for use in further analysis. Our main adjustment was for currently operating companies not reporting recent employment information. For companies coded by NETS as still in business (i.e., "Out of Business" flag set to false) but missing recent employment information, we backfilled employment for up to three years to increase our estimates' accuracy. This is consistent with the sometimes-infrequent updating of small companies in the NETS dataset.

In NETS, an establishment is a specific line of business at a specific location (see below), and employment includes all workers at an establishment, potentially including proprietors, independent contractors, and temporary workers supplied by outside organizations. This group of workers constitutes a superset of workers that include employees as traditionally defined and measured in datasets produced by government statistical agencies like the County Business Patterns (CBP) and Quarterly Census of Employment and Wages (QCEW) datasets. However, findings that (a) correlation of NETS employment counts and CBP employment counts across U.S. counties can be in excess of 0.99, (b) correlations across state-industry-size class cells are above 0.9 on restricted samples, and (c) zip code-level correlations are also remarkably high provide confidence that workers in the NETS data meet the conventional definition of an employee as defined in government datasets. Additionally, while NETS records are annual, information is collected throughout the year, and the timing of measurement for each establishment is not reported in the data. Our estimates of employment should therefore not be thought of as a point estimate of employment at all the VC-backed companies in our dataset on a specific date (in contrast to how some government surveys collect and measure employment data), but rather as more general measures of employment throughout the year. For further details on how employment is defined in NETS, please refer to the NETS documentation and two Federal Reserve research papers available here and here.

Imputing Employment

Of the 66,318 VC-backed companies in the PitchBook dataset, our matching algorithm captured 33,095 unique companies in the NETS data that had DUNS Numbers. 31,100 of these companies have employment data when they appeared as headquarters (HQDUNS) in the NETS data. Among the other 35,218 companies in the PitchBook dataset that either do not have DUNS Numbers or the DUNS Numbers are not in the NETS data, there are 33,503 companies that have information on both Industry Sector and Founding Year. We imputed the employment of these companies using estimated industry and company age fixed effects estimators obtained through panel regression analysis. This method is equivalent to calculating by-age, by-industry averages across all companies for which we have NETS data.

To obtain the fixed effects estimators, we classified the companies into 7 industries and 20 age groups, ranging from age 1 to age 20+. The resulting 140 fixed effect coefficients were estimated using the employment data available for the 30,163 headquarters. However, 632 of the 30,163 headquarters had missing Founding Year values in the PitchBook data, which are required when calculating company ages. We used the Year Start variable in the NETS data to fill in these missing founding years. In addition, there were 1,762 headquarters that have employment data in the NETS data before the companies are founded according to the Founding Year variable in the PitchBook data, while the Year Start variable in NETS data is consistent with the first year of employment in general. For consistency, we used the Year Start variable of the NETS data to replace the Founding Year variable for these 1,762 headquarters whenever Year Start was available. We further removed 26 outliers as their company ages were less than or equal to zero, or their founding years were not available after the backfill process discussed above, resulting in 30,137 companies with headquarters data available for the estimation of the fixed effects coefficients.

We constructed the employment data by company age for the 30,137 headquarters by aggregating the employment over all branches belonging to the same company headquarter by company age. All companies with a company age greater than or equal to 20 years old were grouped into a single category labeled as "20+". The average employment of companies in this "20+" category was assigned as the employment for all companies in this group. This averaging allowed us to characterize mature companies older than 20 years without having to generate fixed effects for every age year. Having tied company-wide employment to each headquarter location and adjusted the data for companies established 20 or more years ago, the 140 industry and company age fixed effects coefficients were then estimated using Ordinary Least Squares. The missing employment data of the 33,503 VC-backed companies were then imputed using the estimated fixed effects according to each company's industry and company age. For companies that were acquired, we locked employment at the level in the year of acquisition and in the geographies in which these workers were employed (at the establishment level). We do this primarily because once a company is acquired, we have no visibility in the NETS data on future changes in employment for the acquired company post-acquisition. In some cases, establishments are absorbed into the parent company, and so the VC-backed company's employment becomes zero. However, on average, this is likely a conservative assumption, because most establishments subject to acquisition that remain independently measured show growths in employment.

Mapping Employment to Congressional Districts

Both PitchBook and NETS datasets include a zip code field associated with each company or establishment in their respective datasets. However, neither dataset includes a similar field for the congressional district in which a company is located. Allocating a company's employment at a particular establishment to local congressional districts consequently requires a mapping algorithm. This algorithm was executed in two stages to address the two types of record data we had available: (1) records of employment data with linked zip codes that were pulled directly from the NETS database, and (2) records of imputed employment (imputed because we could not find a proper match within the NETS administrative data) that also had linked zip codes.

To impute how employment for a company headquartered in a certain congressional district may be distributed across multiple congressional districts, we produced and ran an algorithm based on the NETS data that estimates the shares of employment for companies headquartered in each congressional district that may ultimately lie in other congressional districts. Applying this algorithm to non-headquarter locations allowed us to estimate how employment at a branch that is geographically distinct from the company's headquarters is spread across other congressional districts.

To develop the algorithm, we first assigned companies to congressional districts based on their zip code and state information in the NETS data using the HUD USPS Zip Code Crosswalk File for this mapping. We note that the zip codes do not have a one-to-one relationship with congressional districts, and specifically there are 6,077 zip codes that map to more than one congressional district. For zip codes that span multiple congressional districts, we assigned the first congressional district that appeared in the Crosswalk File for each of these zip codes. We also removed companies from 8 zip codes that were not in the Crosswalk data. Apart from companies located in these 8 zip codes, we assigned congressional districts to each branch and headquarters for every company for which we had administrative data for employment and associated zip codes and states pulled directly from the NETS data. This assignment addresses the first set of record data mentioned.

This first step provides the ability to address the second set of records consisting of companies with imputed employment and to allocate these companies' employment to each of 437 congressional districts (we include Washington, DC and Puerto Rico in addition to the 435 congressional districts with voting representatives). In essence, we assumed that the allocation across congressional districts for the companies with DUNS Numbers also applied to companies with imputed data. We limited allocations of branch employment to other districts solely to branches in districts that had at least 1% of employment outside the headquarter district. However, this allocation is quite modest in practice. On average about 98% of employment was found to be in the headquarter district, so the typical imputed company is allocating only a small fraction of its employment to just one or two other districts. We augmented the final imputed employment by distributing the total employment to all congressional districts based on the congressional employment matrix.