UC Santa Cruz

UC Santa Cruz Previously Published Works

Title

Measuring Job Creation, Growth, and Survival among the Universe of Start-ups in the United States Using a Combined Start-up Panel Data Set

Permalink

<https://escholarship.org/uc/item/7p90k10t>

Journal

ILR Review, 72(5)

ISSN

0019-7939

Authors

Fairlie, Robert W
Miranda, Javier
Zolas, Nikolas

Publication Date

2019-10-01

DOI

10.1177/0019793919862764

Peer reviewed

Measuring Job Creation, Growth and Survival among the Universe of Startups in the United States using a Combined Startup Panel Dataset

Robert W. Fairlie
University of California, Santa Cruz and NBER
rfairlie@ucsc.edu

Javier Miranda
Center for Administrative Records Research & Applications
U.S. Census Bureau
javier.miranda@census.gov

Nikolas Zolas
Center for Economic Studies
U.S. Census Bureau
nikolas.j.zolas@census.gov

June 10, 2019

Abstract

The field of entrepreneurship is growing rapidly and expanding into new areas. Our goal in this paper is to present a new compilation of administrative panel data on the universe of business startups in the United States that will be useful for future research in entrepreneurship. To create the U.S. startup panel dataset we link the universe of non-employer firms to the universe of employer firms in the Longitudinal Business Database (LBD). Startup cohorts of more than five million new businesses per year creating roughly three million jobs can be tracked over time. To illustrate the potential of the new startup panel dataset for future research we provide descriptive statistics for a few examples of research topics using a representative startup cohort.

Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. We thank participants at the Labor Market and Human Resource Management Implications of Entrepreneurship Conference, National Academies for comments and suggestions. The DRB codes for this project are: DRB-B0053-CED-20190516.

1. Introduction

Many policymakers and organizations around the world have called upon entrepreneurs to create new jobs (OECD 2017). Starting with the seminal study by Birch (1979) showing that small businesses are the principal driver of job creation in the U.S. economy, there has been considerable interest in job creation among entrepreneurs. Recent evidence indicating that young and high-impact businesses (defined as having high rates of growth in sales and employment) disproportionately contribute jobs in the economy (Haltiwanger, Jarmin and Miranda 2013; Kulick, Haltiwanger, Jarmin; Miranda 2016; Tracy 2011) has further heightened interest in the job creating potential of entrepreneurs.

The previous research on job creation among businesses focuses almost exclusively on employer firms (i.e. firms at the stage of hiring their first employee), mainly due to conceptual challenges and data limitations. However, this only tells part of the story. Businesses without employees (i.e. non-employer firms) comprise the majority of both startups and total businesses in the United States. While most of these non-employer businesses start and fade quickly without ever hiring an employee, many of these same businesses transition to employer businesses. The exclusion of non-employer businesses from this line of research, may represent a limitation for understanding early-stage job creation among startups.¹ For

¹A few recent studies examine the relationship and growth patterns between non-employer to employer businesses. These studies find, for example, that non-employers have startup rates that are nearly three times the startup rates of employer firms, a significant number of new employer firms start as non-employer firms, and if non-employer startups hire, the bulk of hiring occurs in the first few

example, employer firms that start as non-employers may be misclassified as new businesses even if they actually started several years earlier. Another concern is that total levels of business and job creation per startup cohort are miscalculated if non-employer startups are excluded. In all cases, ignoring the non-employer history of businesses might miss important early entrepreneurial dynamics that are important for understanding the relationships between entrepreneurship, job creation, and growth.

To address these concerns, we create and present a new compilation of administrative panel data on the universe of business startups in the United States. The new startup panel dataset can be used to study many additional questions around entrepreneurial job creation and growth. We link the universe of non-employer businesses to the universe of employer businesses in the Longitudinal Business Database (LBD). To create the necessary links between non-employer and employer definitions, we update and expand the Integrated Longitudinal Business Database (ILBD). Beta versions of the ILBD have been used in previous work (e.g. Davis et al 2007; Fairlie and Miranda 2016) to study transitions between non-employer and employer business units, but these analyses were only for one or two years, a subset of industries, and/or not focused on startups.

Our goal in this paper is to demonstrate how our new startup panel dataset can be used for research in the rapidly growing field of entrepreneurship. The discussion focuses on how the dataset can be used to

years of existence (Davis et al. 2007; Acs, Headd and Agwara 2009; Fairlie and Miranda 2016).

identify the year of startup for non-employer business units and follow annual cohorts of startups to examine their time series properties and dynamics. After introducing the new startup panel dataset we present some descriptive statistics for an example startup cohort. The goal is to generate interest by researchers in using the new compilation of administrative data on the universe of U.S. startups to study questions about entrepreneurship, job creation, growth, and other related topics.

2. U.S. Startup Panel Dataset

We describe a new compilation of administrative panel data on the universe of business startups in the United States. To create this U.S. startup panel dataset, we link the universe of non-employer businesses to the universe of employer businesses in the Longitudinal Business Database (LBD). Instrumental for these links is the U.S. Census Bureau Integrated Longitudinal Business Database (ILBD), which is sourced from administrative income and payroll filings. The ILBD provides information on the universe of non-employer businesses as well as a set of identifiers that allow connecting them to the universe of employer businesses in the Longitudinal Business Database (LBD).² Combining the ILBD and LBD then allows researchers to explore the connections between the two universes including the transitions between the two. Starting in 1994 and in combination with the LBD, the ILBD provides annual information that allows all businesses, employer and non-

² For details about the LBD see Jarmin and Miranda (2002).

employer, to be followed over time. Because the combined dataset on startups contains the links between the employer and non-employer universes it is possible to accurately identify the point of startup for business units and each annual cohort of startups can be followed over time.³

The new startup panel dataset provides a comprehensive view of patterns of non-employer and employer businesses including the non-employers that will hire employees in some future year. These data provide information at the point in time of hiring instead of less reliable retrospective information collected through survey forms.⁴ Another advantage of the new startup panel dataset relative to many other datasets is that it contains administrative data on the universe of business units and thus suffers little from attrition problems.⁵

The transition between non-employer and employer status is especially important for examining job creation among all startups.⁶ Identification of these transitions is challenging and explains why the employer and non-

³ The Census Bureau's Longitudinal Business Database (LBD) addresses this issue by linking establishments with an inactive (non employment) spell of up to 5 years. Establishments that exit the employer universe and have not reentered it are considered exits in the LBD. Left censoring at the beginning of the series prevents us from identifying the start year for all businesses in the data for early years

⁴ Employment in tax filings covers the payroll period of March 12th.

⁵ A non-government source of panel data on startups is the Kauffman Firm Survey (KFS). The KFS follows a sample of roughly 5,000 non-employer and employer startups from the 2004 cohort. For this sample of startups, roughly 90 percent survive one year and 55 percent survive 5 years after startup (Robb and Farhat 2013). The KFS has also been used to study non-employer to employer transitions (Fairlie and Miranda 2016). The underlying sampling frame of the KFS, however, is skewed towards including employer startups.

⁶ We focus on the transitions between existing business entities. A new public-use data product recently released by the U.S. Census Bureau as a research (beta) series, the Business Formation Statistics (BFS) measures new employer business applications in the United States to combine with employer business formations.

employer administrative data are kept separate at the Census Bureau. Businesses in the United States are required to file separately income and payroll (employment) taxes. However, businesses may or may not use the same tax identifiers when filing their income and payroll reports.⁷ Since the transition to employer status can only be identified by linking the income filing to the payroll filing, this can lead to broken linkages between the two. To resolve this problem employer and non-employer businesses units are linked by a variety of identifiers including the EIN, the Protected Identification Key (PIK), and the name and address of the owner or business.⁸ These enhanced linkages are tracked in the ILBD through longitudinal business identifiers, ILBDNUMs. When a new firm enters into the dataset, they are automatically assigned a new ILBDNUM, which is used to track the performance of the firm in subsequent years. Firms who exit lose their ILBDNUM and that ILBDNUM is not assigned elsewhere. For firms who exit and then re-enter at a later date, those firms will have two separate ILBDNUMs.

For EINs, the linkage is straightforward, as that information is contained in the Business Register. Only EINs that have zero payroll are included in the non-employer data (ILBD) and once they have positive

⁷ Non-employers do not file payroll taxes.

⁸ The PIK is an individual identifier that replaces the SSN in all files at the U.S. Census Bureau. The PIK ensures no Census employee or researcher has access to SSNs. The linkage between the non-employer and the employer universes makes use of the name of the business and the tax identifiers; the EIN and the PIK. If any of these change then it might not be possible to form a link. This is more likely when there is a change in legal form of organization at the time of the transition. For additional discussion of these issues see Davis et al. (2007).

payroll, they move to the LBD, where the linkage is direct. For PIKs, the linkage is less straightforward. We use exact literal matches on business names, or the SSN that is found when the business applies for an EIN code from the SS-4 data. SS-4 data contains the business name, EIN and SSN of the business owner, which is contained in the BR. Roughly 17% of matches rely on the exact name match of business names. See Davis et al. (2007) for more details.

Every non-employer business that files taxes or a tax report is included in the ILBD. The ILBD contains information on the legal form; sole proprietor, partnership or corporation, the type of tax identifier; EIN or PIK, their revenue, and industry of activity. The file does not contain owner information although this could be obtained from other sources.⁹

One concern is that the startup panel dataset contains a large number of business activities that have no intention of hiring employees and represent consulting, contracting or hobby activities because of the inclusion of the ILBD. But, these data provide a useful view of the universe of non-employer business units, and there is the potential to identify more growth-oriented businesses by conditioning on a few of the administrative variables such as removing sole proprietorships.

⁹ The U.S. does not maintain a register of business owners. Administrative records allow identification of most but not all business owners. Owners of private corporations are particularly difficult to identify in administrative data. There are also data access limitations to some of these identifying files so that currently only Census Bureau employees with a need to know can access them.

There are essentially two data-driven primary classifications of business entities by the U.S. Census Bureau and four legal form sub-classifications within the non-employer universe:

1) Employer startups

2) Non-employer startups

2.1) Incorporated: business granted a charter recognizing it as a separate legal entity having its own privileges and liabilities distinct from those of its members.

2.2) S-Corporation: A form of corporation where the entity does not pay any federal income taxes, and its income or losses are divided among and passed to its shareholders (which are reported on their own individual income tax returns).¹⁰

2.3) Partnership: unincorporated business where two or more persons join to carry on a trade or business with each having a shared financial interest in the business.

2.4) Sole proprietorship: unincorporated business with a sole owner.

Information on these classifications can be used to tailor the definition of an entrepreneur. Instead of including or excluding all non-employer startups the definition of an entrepreneur could use only a subset of non-employer startups. For example, a more restrictive definition of an entrepreneur could include all non-employer startups that are incorporated, S-corporations or partnerships, but exclude sole proprietorships. In each of these cases there is a much stronger business registration requirement than for sole proprietorships. Also, consultants and contract workers will typically show up as sole proprietorships because these work arrangements are technically

¹⁰ Even though S-Corps may not pay any income taxes, they are legally required to file taxes for earnings and expenses. If the firm does not receive any revenues, they are still legally required to return form 1120S as long as they are incorporated and intend to make a profit. Companies may decide to be taxed as an S-Corp for tax purposes as income and losses are passed onto the owner's tax returns. This allows the owner to take money out of the business without paying employment tax.

classified as business entities in the data because of treatment in the tax code. But, using this approach will result in the loss of some growth-oriented sole proprietorships that eventually hire employees.

The ILBD contains information on many business characteristics and outcomes of interest including revenue location and industry of non-employers. When combined with the LBD it is possible to identify for the superset (including employers) the startup year, non-employer/employer status, number of employees, payroll, industry, exits, and revenue, but much more information is available. Table 1 includes a listing of the variables contained in the ILBD and the LBD and whether the information is specific to non-employer or employer businesses.¹¹

The new compilation startup panel dataset and underlying ILBD and LBD are confidential and restricted-access. Researchers can apply for access to the code and data through Federal Statistical Research Data Centers (FSRDC). Basic instructions for accessing the compilation startup panel data and the code that creates it are described in the Appendix as well as locations of FSRDCs.

3. Underlying Non-Employer and Employer Data

We first examine the underlying non-employer and employer data. Table 2 reports various statistics from the universe of non-employer and employer firms that comprise the new startup panel dataset. Employer firm

¹¹ A full listing of the variables contained in the LBD can be found at <https://dev.ncrn.cornell.edu/ced2ar-web/codebooks/lbd/v/v2>

counts are restricted to the private non-farm economy as defined in the LBD and its public use aggregates, the Business Dynamics Statistics (BDS). Non-employer counts are restricted to tabulated cases as defined in the published non-employer statistics.¹² The discrepancy between establishment and business numbers in the employer universe is due to many businesses having multiple establishments. The total number of non-employers in the United States is substantially higher than the number of employers. On average there are four times as many non-employers as there are employer firms in the United States. Including the large number of non-employers might be especially important in studying early-stage business activity.

4. Transitions, Job Creation and Survival among the Universe of Startups

The new compilation of panel data on the universe of startups can be used to study a wide range of topics. In particular, the startup panel dataset is extremely useful for examining the early-stage dynamics of entrepreneurship for job creation and survival. In this section, we briefly present some descriptive statistics on a handful of the many topics that can be explored.

4.1 Non-employer and Employer Startups

¹² The Business Dynamics Statistics can be found at <http://www.census.gov/ces/dataproducts/bds/index.html>. Non-employer statistics can be found at <https://www.census.gov/econ/nonemployer/index.html>.

Table 3 reports the number of non-employer and employer startups over time. The number of non-employer startups dwarfs the number of employer startups. The proportion has steadily climbed over time with non-employer startups representing roughly 95 percent of all startups in the most recent year. The total number of startups in each cohort is roughly 6-7 million. Table 3 makes it clear that leaving out non-employers obscures a relevant dimension to entrepreneurial dynamics in the United States.

4.2 Hiring Dynamics among Non-Employer Startups

The startup panel dataset can be used to examine non-employer businesses that transition to employer status (i.e. hire their first employee). We start from the population of non-employer startups in 1997.¹³ We follow the life cycle of non-employer startups as they either continue on or make their way out of the non-employer population as either a permanent exit or a transition into employer status. A permanent exit requires that the EIN, PIK, or business name does not appear again in the data either as an employer or as a non-employer.¹⁴ To this end, we can follow the 1997 startup cohort

¹³ Non-employer startups are identified as businesses with no activity in the four years prior to it being classified as a non-employer (either as a non-employer or as an employer).

¹⁴ The question of the period of inactivity that one should allow for the reentering firm to still be considered the same firm does not have an easy answer. The OECD Business Demography Statistics allow for a two year gap. The LBD allows for a window of 5 years. The underlying assumption is that knowledge that is embodied in the firm through the accumulation of experience depreciates such that after a period of inactivity the reentering firm is considered a new firm. Here we follow a conservative strategy that typifies a firm as reentry rather than a startup if it had some activity in the 4 years prior. Similarly we treat them as permanent exits if they are not in the data for a minimum of 7 years.

through 2011. A transition to employment status takes place when we are able to identify a non-employer business unit in the employer universe. We track transitions up to seven years after startup. We include the complete universe of non-employer firms across all industries, revenues and business types.

Table 4 reports the distribution of all non-employer startups across hiring their first employee in the seven subsequent years, not hiring their first employee by the end of the seven-year period, or exiting (also see, Fairlie and Miranda 2016). Among all non-employer startups, 2 percent hire their first employee in the first year after startup. Very low percentages of non-employer startups hire their first employee after that year. A large percentage of non-employer startups (12.7 percent) continue as non-employers by the end of the seven-year period. Nearly 85 percent of non-employer startups exit over the sample period before ever hiring their first employee. These results highlight the extreme version of the up-or-out dynamics documented in the employer universe (see Haltiwanger, Jarmin and Miranda 2013).

A very large percentage of these non-employer businesses are consulting, contracting or small-scale business activities. The legal form of organization and the tax identifier used to file taxes appear to be correlated with the intention of the business to either create jobs or remain a non-employer entity. In unreported results focusing on samples of only incorporated businesses or EIN filers we find much higher rates of hiring

employees in the first year at 16.2 and 11 percent respectively. These findings highlight the need for using caution in working with the non-employer startup universe. Interestingly, however, the industry distribution of non-employers does not differ substantially from the industry distribution of employers, as evidenced in Table 5.

4.3 Job Creation among Startups

The startup panel dataset can be used to examine job creation among a cohort of non-employer and employer startups. As an example, Figure 1 reports cumulative job creation for the 1997 cohorts of non-employer and employer startups through the first five years. For this exercise we track the firm identifier associated with the startup in the LBD. Identifiers can change due to single to multi establishment firm expansions. We resolve this by following establishments over time and identifying the relevant multi establishment firm identifier. Measures of job creation are imperfect due to ownership changes and M&A activity but are still indicative of relative magnitudes.¹⁵ Five years after startup, 2.8 million jobs were created in total by the 1997 startup cohort. Although the bulk of these jobs were created by businesses that started as employers a large number of jobs are created by businesses that initially started as non-employers. Roughly 700,000 jobs are

¹⁵ It is conceptually difficult to track job creation after a firm changes its firm identifier. Identifiers change as a result of changes in ownership or M&A activity. We do not attempt to rectify this problem here. In addition, our measures of job creation reflect a mix of greenfield job creation as well as job creation from acquisitions and divestitures. The further we attempt to measure job creation after startup the noisier our computed measures become.

created by non-employer startups in the 1997 startup cohort in year 2002. Focusing on only employer startups would capture these non-employer startups only later after they made the transition to employer status. As a result job creation is attributed to later cohorts of employer startups obscuring the connection to the 1997 startup cohort universe.

4.4 Exit Rates among Startups

The data can be used to track exit rates among a cohort of non-employer and employer startups. As an example, Figure 2 reports the number of startups from the 1997 cohort surviving over a five-year period. Non-employer startups have an exceptionally high rate of exit over the period. Fewer than 1 million of the 4.7 million non-employer startups have survived by year five. Part of the high exit rate among non-employer startups is due to the inclusion of business entities that might represent temporary work arrangements. The survival rate of employer startups is much higher, but still relatively low as 46% of employer start-ups drop out by year five.

5. Summary and Future Directions

Job creation is one of the most important aspects of entrepreneurship, but we know relatively little about the early-stage hiring patterns and decisions of startups. Our compilation of administrative data on the universe of startups discussed here provides a new and exciting startup panel dataset for studying questions about entrepreneurship, job creation, and other

subjects. The advantages of the startup panel dataset for studying these questions include having the universe of startups (both non-employer and employer), information on job creation (employees, payroll), panel data, the ability to follow annual startup cohorts over time, and other related outcomes (employees, revenues, survival, etc...). The data are administrative so there is little attrition. Additionally, as the dataset is further developed there will be the possibility to link founder characteristics and employee characteristics to startups (e.g. similar to studies using Danish and Swedish data). Already there is the possibility of limiting the definition of entrepreneurship using information on classifications of businesses such as non-employer and employer, and incorporated, S-Corporation, partnership or sole proprietorship. Most of the examples presented here use the full universe of startups that includes all classifications, but there are several more restrictive samples that can be created using the microdata.

Many future questions about entrepreneurship can be addressed using these data. For example, who makes up the population of startups each year? What drives the up-or-out dynamics of non-employer businesses? What is the contribution of continuers and high growth non-employer startups to revenue growth? What is the nature and growth patterns of high-tech startups compared to other startups? How do firms that start off without employees differ from firms that start with employees? What are the differences in their growth patterns over time? What are the employment dynamics of different startup cohorts?

Another area of promise for future research is the ability to merge the startup panel dataset with other Census datasets. For example, the startup panel dataset can be merged with the LEHD to examine the workers at startups in more detail. Another example, is that the American Community Survey, which provides demographic and educational information, can be merged to a large share of startup owners. These are just a few examples of the plethora of questions in entrepreneurship that can be addressed with the new compilation U.S. startup panel dataset introduced here.

References

Acs, Zoltan J., Brian Headd and Hezekiah Agwara. 2009. *The Nonemployer Start-up Puzzle*, U.S. Small Business Administration, Office of Advocacy Working Paper.

Birch, David L. 1979. *The Job Generation Process*, Report prepared for the U.S. Department of Commerce, Economic Development Administration, Washington, D.C.

Davis, Steven J., John Haltiwanger, Ron Jarmin, C.J. Krizan, Javier Miranda, Al Nucci and Kristen Sandusky. 2007. "Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes," NBER Working Paper No. 13226.

Fairlie, Robert W., and Javier Miranda. 2016. "Taking the Leap: The Determinants of Entrepreneurs Hiring their First Employee." *Journal of Economics & Management Strategy* (forthcoming).

Haltiwanger, John C., Ron S. Jarmin and Javier Miranda. 2013. "Who Creates Jobs? Small vs. Large vs. Young" *The Review of Economics and Statistics*, 2013, vol. 95, issue 2, pages 347-361.

Jarmin, Ron and Javier Miranda. 2002. "The Longitudinal Business Database." CES Working Papers 02-17, Center for Economic Studies, U.S. Census Bureau.

Kulick, Robert, John C. Haltiwanger, Ron S. Jarmin and Javier Miranda. "High Growth Young Firms: Contribution to Job, Output and Productivity Growth" Forthcoming NBER book Measuring Entrepreneurial Businesses: Current Knowledge and Challenges, John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar, editors

OECD. 2017. *SME and Entrepreneurship Policy Series*. Organisation for Economic Co-operation and Development Press, Paris.

OECD. 2017. *Entrepreneurship at a Glance 2017*. OECD Publishing, Paris.

Tracy, Spencer L. 2011. *Accelerating Job Creation in America: The Promise of High-Impact Companies*, U.S. Small Business Administration, Office of Advocacy Working Paper.

Appendix

Accessing the Data

The underlying ILBD and LBD and the new compilation startup panel dataset are confidential and restricted-access. At this point, the programming code by the authors that creates a new universe startup panel dataset from the underlying LBD and ILBD is available. Researchers can apply for access to the code and data through Federal Statistical Research Data Centers. Federal Statistical Research Data Centers are partnerships between federal statistical agencies, and universities or other research institutions. They are managed by the U.S. Census Bureau to provide secure access to a range of restricted-use microdatasets.

FSRDCs are located throughout the United States. Appendix Figure 1 displays the locations of each FSRDC in the country. New locations include FSRDCs at the University of Illinois Urbana-Champaign, and the University of Utah in Salt Lake City, Utah (denoted with red dots in Appendix Figure 1). Appendix Table 1 reports a full list of FSRDCs and their host universities or research institutions.¹⁶

According to submission guidelines provided by the U.S. Census Bureau proposed projects must satisfy four criteria: i) provide benefit to Census Bureau programs, ii) demonstrate scientific merit, iii) require non-public data, iv) be feasible given the data, and v) pose no risk of disclosure.¹⁷

In terms of developing and submitting a research proposal, researchers are

¹⁶ For the most up-to-date list see <https://www.census.gov/about/adrm/fsrdc/locations.html>.

¹⁷ See <https://www.census.gov/ces/rdcresearch/howtoapply.html>.

instructed to consult the FSRDC administrator at the location where they want to base their project. They should also consult with the FSRDC administrator regarding access fees and the content and form of a proposal. The FSRDC administrator will then assist the researcher in developing the proposal and identifying potential benefits to Census Bureau programs.

After the proposal is developed with the FSRDC administrator, it will be submitted to the Center for Economic Studies for Census review and then get further approval from IRS.¹⁸ The Census review process seeks to minimize potential issues before the proposal reaches IRS. Researchers on approved projects must also complete a background investigation, including fingerprinting. The process is lengthy: it can take a number of months to complete all of the required steps before a researchers might expect to begin using the data at an FSRDC. Because the startup panel data include Federal Tax Information (FTI) additional criteria will apply in reviewing and approving projects by the Census Bureau and the Internal Revenue Service (IRS).¹⁹ More details of the process can be found at <https://www.census.gov/ces/rdcresearch/howtoapply.html>.

Once the project is approved and the researcher has access to the microdata, the researcher can work with Census data staff to move external datasets into the project space. Before any statistical or economic results can leave the project space, the researcher will need to work with the FSRDC administrator and the Disclosure Review Board to disclose the results and

¹⁸ See https://www.census.gov/ces/pdf/Research_Proposal_Guidelines.pdf.

¹⁹ See https://www.census.gov/ces/pdf/IRS_Criteria_Document.pdf.

ensure that no FTI information or identifiable information is released to the public.

Appendix Table 1
Locations and Partnerships of Federal Statistical Research Data
Centers as of May 2019

Atlanta: Founded in 2011, the Atlanta RDC is located in the Federal Reserve Bank of Atlanta and attracts researchers from across the Southeast.

Boston: Founded in 1994, the Boston RDC is located at the NBER in Cambridge and provides secure data access to researchers from across New England.

California – Berkeley: Founded in 1998, the California RDC at Berkeley is located on the UC Berkeley campus, and serves researchers from the Bay Area to Northern California.

California – Irvine: Opened in late 2014, the California RDC at Irvine is located in the Social Sciences Gateway Building on the University of California, Irvine campus.

California – Stanford: Founded in 2009, the Stanford RDC operates in partnership with the California RDC, serving San Francisco peninsula and Northern California researchers.

California – UCLA: Founded in 1998, the California RDC at UCLA is located on the UCLA campus and provides data access to researchers in southern California.

California – USC: Founded in 2014, the California-USC RDC is located on the campus of the University of Southern California.

Census Bureau Headquarters: The Census Bureau Headquarters RDC is located in Suitland, MD and serves researchers throughout the Washington, DC metro area.

Central Plains: Opened in November 2015, the Central Plains RDC serves researchers at universities within the Central Plains region.

Chicago: Founded in 2003, the Chicago RDC consortium includes the Chicago Fed, Northwestern University, the University of Chicago and the University of Illinois.

Dallas-Fort Worth: Opened in 2018, the Dallas-Fort Worth FSRDC is located in the Federal Reserve Bank of Dallas and provides access to researchers in the region.

Georgetown: Opened in 2017, the Georgetown RDC is located at the Massive Data Institute at Georgetown University's McCourt School of Public Policy.

Kansas City: Opened in February 2016, the Kansas City RDC is located at the Federal Reserve Bank of Kansas City and attracts researchers from across the region.

Kentucky: Opened in 2017, the Kentucky RDC is located on the University of Kentucky campus in Lexington, and will serve researchers in the East-Central region.

Maryland: The Maryland RDC is a joint project of the Robert H. Smith School of Business, the College of Behavioral and Social Sciences, and the School of Public Health.

Michigan: Founded in 2003, the Michigan RDC is located at the University of Michigan's Institute for Social Research in Ann Arbor.

Minnesota: Founded in 2010, the Minnesota RDC supports researchers at the University of Minnesota and welcomes other researchers from the Twin Cities and Upper Midwest.

Missouri: Opened in November 2015, the Missouri RDC is located at the Ellis Library on the University of Missouri campus, and serves all researchers in the region.

New York – Baruch: Founded in 2005, the NYRDC-Baruch is located at Baruch College-CUNY in New York City and serves researchers in New York, New Jersey and Connecticut.

New York – Cornell: Founded in 2005, the NYRDC-Cornell is located in Ithaca and provides data access to researchers in the Finger Lakes region of New York State.

Northwest: Founded in 2012, the Northwest RDC is located in Seattle and provides secure access to restricted data for researchers in the region.

Penn State: Founded in 2014, the Penn State RDC is located in the Paterno Library in the center of the University Park campus.

Philadelphia: Opened in 2017, the Philadelphia FSRDC is located in the Federal Reserve Bank of Philadelphia and attracts researchers from across the East.

Rocky Mountain: Opened in 2017, the Rocky Mountain RDC is located at the Un. of Colorado Boulder and serve researchers along the Front Range of the Rocky Mountains.

Texas: Founded in 2012, the TXRDC is located at Texas A&M University in College Station, Texas and provides data access to researchers in the region.

Texas - UT Austin: Opened in 2017, UT-Austin provides researchers in central Texas secure environment to access federal statistical data.

Triangle – Duke: Founded in 1999, the Triangle RDC serves the social science research community across North Carolina and southern Virginia.

University of Illinois Urbana-Champaign (Opening 2018): Opening in 2018 the UIUC FSRDC provides researchers with the opportunity to conduct approved research projects using confidential federal data.

Wasatch Front: To be opened in 2019, The WFSRDC is located at the University of Utah in Salt Lake City, Utah.

Wisconsin: Opened in September 2015, the WiscRDC is located in the heart of the UW Madison campus, and serves a strongly interdisciplinary community.

Yale: Founded in 2015, the Yale RDC serves researchers in New York, New Jersey, and Connecticut.

Table 1 Variables Contained in the ILBD and the LBD

ILBD	LBD
Non-employer ID	
Employer ID (EIN,PIK)	Employer ID (EIN,PIK)
ILBDNUM	LBDNUM
Legal Form	Legal Form
Industry	Industry
Geographical Location	Geographical Location
Business Name	Business Name
SpouseID	
Year	Year
	Startup Year
	Age Of Startup
	Number Of Employees
	Payroll
Revenue (\$ 000s) ¹	
Interest ¹	
Rents ¹	
Assets ¹	
Revenues From 1 st , 2 nd And 3 rd Record	
Industry For 1 st , 2 nd And 3 rd Record	
¹ Equivalent information is available for the employer universe but only at the EIN level. This information is not on the LBD since this is an establishment level dataset.	

Table 2 Employer and Non-Employer Tax Units

	Total Tax Units	Non Employer Tax Units	Employer Tax Units	Number of Establishments	Percent Non-Employer
1997	20,191,864	15,439,609	4,752,255	6,041,547	76
1998	20,504,956	15,708,727	4,796,229	6,108,324	77
1999	20,977,087	16,152,604	4,824,483	6,173,747	77
2000	21,366,535	16,529,955	4,836,580	6,218,784	77
2001	21,861,087	16,979,498	4,881,589	6,308,530	78
2002	22,554,772	17,646,062	4,908,710	6,352,721	78
2003	23,612,191	18,649,114	4,963,077	6,415,020	79
2004	24,563,220	19,523,741	5,039,479	6,497,414	79
2005	25,531,480	20,392,068	5,139,412	6,633,143	80
2006	25,948,343	20,768,555	5,179,788	6,736,378	80
2007	26,948,040	21,708,021	5,240,019	6,842,829	81
2008	26,551,385	21,351,320	5,200,065	6,827,573	80
2009	26,723,431	21,695,828	5,027,603	6,651,272	81
2010	27,064,053	22,110,628	4,953,425	6,578,147	82
2011	27,403,338	22,491,080	4,912,258	6,553,920	82
2012	27,715,365	22,735,915	4,979,450	6,625,111	82
2013	28,028,494	23,005,620	5,022,874	6,674,357	82
2014	28,894,955	23,836,937	5,058,018	6,721,326	82

Source: U.S. Census Bureau Business Dynamics Statistics and Non-employer Statistics. Non-employer Statistics are compiled from published yearly totals. Employer firm statistics come from the BDS Firm Characteristics Data Tables and Establishment Characteristics Data Tables.

Table 3: Employer and Non-Employer Startup Totals, 1997-2010

	Total Startups (MM)	Non-Employer Startups (MM)	Employer Startups (MM)	Percent Non- Employer
1997	4.89	4.37	0.52	89
1998	5.07	4.56	0.51	90
1999	5.21	4.71	0.50	90
2000	5.38	4.90	0.48	91
2001	5.41	4.94	0.47	91
2002	5.39	4.89	0.50	91
2003	5.70	5.20	0.50	91
2004	5.92	5.40	0.52	91
2005	6.23	5.69	0.54	91
2006	6.67	6.11	0.56	92
2007	7.20	6.67	0.53	93
2008	6.05	5.56	0.49	92
2009	5.96	5.55	0.41	93
2010	6.70	6.31	0.39	94

Source: ILBD and Business Dynamic Statistics (BDS). Non-Employer Startups are from Authors' calculations and is defined as the first year a non-employer appears in the ILBD. BDS Statistics come from published annual firm totals as of June 2019.

Table 4: Distribution across Years in which Non-Employer Startup Hired Its First Employee - All Non-Employers Integrated Longitudinal Business Database (ILBD)

	Percent	Universe
Hired first employee at:		
1 year after startup	1.9%	84,500
2 years after startup	0.2%	10,200
3 years after startup	0.1%	4,900
4 years after startup	0.1%	2,900
5 years after startup	0.1%	2,500
6 years after startup	0.0%	1,500
7 years after startup	0.0%	1,200
Has not hired employee by end of study period	12.7%	556,200
Exit before hiring employee by end of study period	84.8%	3,704,800
		4,368,700
Total number of non-employer startups:		00

Notes: (1) The data consists of the universe of non-employer startups in 1997. (2) Non-employer startups are defined as non-employers appearing in the non-employer data for the first time in at least three years.

Table 5: Sectoral Composition of Non-Employer and Employer Firms, 1997-2010

Sector	Non-Employer	Employer (Unweighted)	Employer (Employment Weighted)
Construction	11.8%	10.9%	6.1%
Entertainment	6.8%	9.9%	11.3%
Finance, Real Estate	9.9%	10.2%	6.6%
Health and Education	12.5%	11.4%	17.5%
Information	1.8%	1.5%	2.5%
Management	8.9%	5.6%	9.8%
Manufacturing	1.5%	4.6%	12.2%
Other Industries	1.7%	0.9%	1.2%
Other Services	14.1%	10.5%	5.1%
Professional Services	14.3%	11.6%	6.6%
Retail Trade	10.3%	14.3%	12.5%
Transportation and warehousing	4.5%	2.8%	3.3%
Wholesale Trade	1.9%	5.7%	5.1%

Notes: Totals based on Authors' calculations. Sectors are defined using 2-digit NAICS Assignments: Construction (NAICS 23), Entertainment (NAICS 71), Finance and Real Estate (NAICS 52 and 53), Health and Education (NAICS 61 and 62), Information (NAICS 51), Management (NAICS 55), Manufacturing (NAICS 31-33), Other Industries (NAICS 11, 21 and 22), Other Services (NAICS 56, 72 and 81), Professional Services (NAICS 54), Retail Trade (NAICS 44-45), Transportation and warehousing (NAICS 48-49), Wholesale Trade (NAICS 42).

Source: Startup panel dataset created from underlying LBD and ILBD data.

Source: Startup panel dataset created from underlying LBD and ILBD data.