New Public Use Data to Study Entrepreneurship from Linked Employer-Employee Data

Christopher Goetz[1], Henry Hyatt[2], Erika McEntarfer[1], and Kristin Sandusky[1]

[1]*U.S. Census Bureau*

[2]*U.S. Census Bureau and IZA*

December 1, 2014

Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All figures have been reviewed to ensure that no confidential information is disclosed. We thank Douglas Walton and Alexandria Zhang for assistance in preparing tabulations for this paper. This draft is preliminary - please do not cite.

1 Introduction

New businesses continually enter the marketplace to meet unmet demand for goods and services, or to provide goods and services in better ways. While this notion of creative destruction dates back to Schumpeter (1939), new data sources now allow better measurement of business entry, growth, and failure. These improvements in the statistical infrastructure in the U.S. have spawned new interesting research on entrepreneurship. For example, we now know that only a small share of startups accounts for much of the job creation from new businesses – with most new businesses remaining small or exiting the market quickly (Haltiwanger et al., 2010). Recently, researchers have also documented a stark decline in the rate of business startups in the U.S. (Decker et al., 2014a). Yet many fundamental questions of how entrepreneurs come into existence, and why entrepreneurship in the U.S. is declining, likely cannot be answered with existing data.[1] More data on entrepreneurs and entrepreneurial firms are clearly needed.

In this paper, we describe newly available (and early prototypes of) public use statistics on new firms and entrepreneurs that fill some gaps in the available data on entrepreneurship. Specifically, we focus on new public use statistics that have become increasingly available from linked employer-employee microdata at the U.S. Census Bureau's Longitudinal

[1] See Fort et al. (2012), Decker et al. (2014a,b) and Davis et al. (2009), which provide descriptive analyses of the decline in U.S. entrepreneurship, and note that these studies conclude that little of the decline can be explained in a straightforward manner through changes in composition. Decker (2014) explores the decline in startups and the role of the decline in U.S. home equity that resulted from the precipitous decline in housing prices beginning in 2006, while Dinlersoz et al. (2014) explores the role of entrepreneurship finance, labor market search frictions, and the productivity of startups in understanding this decline.

Employer-Household Dynamics (LEHD) program. These data provide additional details about the structure and workforce of entrepreneurial firms, and about the dynamics of entrepreneurial employment. These new statistics address key questions about who starts businesses, who works for startups, and the recent employment histories of both founders and their early employees.

Our purpose here is to describe newly available data on entrepreneurship, and so the structure of the paper is as follows. Each subsequent section serves to highlight a particular dataset or set of statistics, with a focus on how the data can be used in entrepreneurial research. First, we describe the linked employer-employee LEHD microdata, the source data for the new public use statistics, which is also available to researchers through the Census Research Data Centers. We then describe new statistics on workforce demographics and worker flows at young firms developed as a recent enhancement to the Census Bureau's Quarterly Workforce Indicators. A subsequent section describes brand new data (released the week of this writing) on the flows of workers across firms – identifying flows of worker from established firms to startups. Lastly, we discuss new data products under development that combine information on sole-proprietorships (with and without employees) and their owners with the linked firm-worker data. Such worker-level data can help identify the connection between self-employment and entrepreneurship.

The linked employer-employee statistics described in this paper come in the context of an increasing number of statistics available for the study of entrepreneurship derived from business and establishment data. The Business Dynamics Statistics (BDS) is an annual data series beginning in 1976 that provides establishment-level business dynamics, based on the

U.S. Census Bureau's Longitudinal Business Database (LBD). Similar to the BDS, but distinct in several dimensions, is the Business Employment Dynamics (BED), a quarterly data series derived from the Quarterly Census of Employment and Wages (QCEW) establishment data starting in 1992, that provides establishment-level business dynamics. Together the BDS and BED publish measures of job creation and destruction, firm entry and exit, by detailed business characteristics such as industry, size, age, and geography. While these new statistics on young firms have resulted in a corresponding growth of research documenting the importance of new businesses for job creation and economic growth, they lack information on the founders of businesses, and the workforce of startups. Data on founders and their businesses is, however, a key feature of new data produced by the Kauffman Foundation. The Kauffman Firm Survey provides data on a cohort of 5,000 businesses started in 2004, and follow their outcomes such as sales, profits, and survival over time.[2] Due to the survey's small sample size and scope, however, its ability to provide detailed statistics is limited.

The linked employer-employee data that we describe here have the potential to shed new light on many questions in the literature that cannot be addressed using firm-level data alone. For example, a new firm's success or failure is not independent of their ability to attract talented workers. The Quarterly Workforce Indicators that we describe here reveal key details about the workforce at startup firms by detailed industry and geography. The granular industry and geography detail in the QWI allows investigation into agglomeration economies in business startups as observed by Carlton (1983), and the many posited explanations for the

[2] Doms et al. (2010) use these data in combination with publicly available Decennial Census data to examine the link between local workforce traits and startup success.

phenomenon, such as: the supply of entrepreneurs (Chinitz, 1961), cost efficiencies (Saxenian, 1994, Rosenthal and Strange, 2003), technology spillovers (Ellison and Glaeser, 1997), suitability of the labor force (Doms et al., 2010), thickness of labor markets (Figueiredo et al., 2014, Freedman, 2008), and other externalities (Lucas and Rossi-Hansberg, 2002, Brinkman et al., 2015). In turn, the QWI can be used to address the open question of why regional growth appears to be correlated with the presence of many small/young firms (Glaeser et al., 1992, 2010). Such detailed tabulations may also be useful in the related literature on spinoff firms (Klepper, 2001, Franco and Filson, 2006, Chatterjee and Rossi-Hansberg, 2012, Agarwal et al., 2013).

The future sole proprietor statistics will supplement the existing data sets used to describe business owners. The Survey of Small Business Owners (SBO), the Current Population Survey (CPS), the Survey of Consumer Finance (SCF), the Panel Study of Income Dynamics (PSID), The Survey of Income and Program Participation (SIPP), the National Longitudinal Survey of Youth (NLSY79), and the Health and Retirement Study (HRS) all ask a small set of questions concerning self-employment and business ownership, which are quite similar across these different surveys, but do not provide information about the workforce at their firms.[3] Those surveys with a longitudinal component also permit some understanding of the dynamics of business ownership.

Our sole-proprietor statistics will thus be relevant to the large literature focusing on the entrepreneurs themselves, particularly due to our information on employment histories and

[3]For a summary of studies using the NLSY to study entrepreneurship, see Fairlie (2005).

individual characteristics. Previous studies have explored where entrepreneurs come from in terms of previous employment, by the size of their previous employer like in Parker (2006), or from nonemployment as in Evans and Leighton (1989). Another group of studies has sought to predict the success of entrepreneurs based on demographics, education, or prior work experience (Bates, 1990, Delmar and Shane, 2006), while others point out that not all entrepreneurial endeavors are intended to grow (Schoar, 2009, Hurst and Pugsley, 2011). Finally, our measure of sole-proprietors' wages could be pertinent in the study of self-employment outcomes such as income (Kihlstrom and Laffont, 1979), wage volatility (Hamilton, 2000), and even economic mobility (Kitao, 2008).

As a whole, the linked employer-employee data products that we describe in this paper will serve to bring the business, workforce, and household sides together, thus adding a new dimension to entrepreneurship research. With these data we can observe things that are not possible with other data sources, such as the workforce of startups, flows of workers into startups from within and between industries and geographic locations, as well as founders' paths to entrepreneurship.

2 An Overview of the LEHD Data and the Recent Addition of Firm Age and Size

Until relatively recently, many interactions between workers and firms in the U.S. economy were unobservable due to an absence of a link between employer and household data in the national statistical system. The Longitudinal Employer-Household Dynamics (LEHD) program at the U.S. Census Bureau has spent the last decade filling this gap in the nation's statistical infrastructure by building a comprehensive database of longitudinally linked jobs data for the United States. The jobs data necessary to construct linked employer-employee data are voluntarily shared with Census by state governments through the Local Employment Dynamics (LED) federal-state data sharing partnership. As of this writing, all 50 states, DC, Puerto Rico, and the Virgin Islands share QCEW and UI wage data with the LEHD program as part of the Local Employment Dynamics (LED) federal-state partnership.

The LEHD data consist of quarterly worker-level earnings submitted by employers for the administration of state unemployment insurance (UI) benefit programs, linked to establishment level data collected for the Quarterly Census of Employment and Wages (QCEW) program. Demographic characteristics of the workers are derived from Census survey, census, and administrative records sources. As discussed later in this section, information on firm age and size data was recently integrated into the LEHD data from the source information for the BDS. LEHD data coverage is quite broad, with state UI covering 95% of private sector employment, as well as state and local government. A comprehensive overview of the LEHD data is found in Abowd et al. (2009).

Several features of the LEHD data should be of particular interest to researchers interested in the study of entrepreneurship. First, the potential to study startup teams as groups of workers moving from their previous employers to the newly established firm is unique to linked employer-employee data. While identifying which early employees are founders or first hires must be inferred from the data indirectly,[4] the ability to identify such moves for almost the entire universe of business startups in the U.S. compensates somewhat for the paucity of information on the role of each employee in the founding of the firm. A second interesting feature of the data is the ability to study the workforce composition of new enterprises and their early employees – examining the role that gender, age, industry experience, experience working at other new businesses plays in the success or failure of new firms.[5] The ability to identify co-workers and network effects from working in new technologies may also be interesting to researchers studying agglomeration effects and their role in forming industry clusters.

Researchers can apply for access to LEHD microdata by submitting a detailed research proposal through the Census Research Data Center (RDC) network. Applications for microdata access for research undergo a formal approval process that includes review of the proposal by the Census Bureau as well as by state and federal agencies that have supplied worker and firm data that are inputs to the LEHD linked microdata. Projects approved to conduct research on the confidential microdata are conducted in a secure research data center facility with all output undergoing a formal disclosure review process before it can leave the center. While many researchers do gain access to LEHD microdata for research through the RDCs, the high cost in terms of time for such projects – time to write a successful proposal to access the data, to obtain necessary approvals, to travel to a research data center which may

[4] As in Agarwal et al. (2013).

[5] Ouimet and Zarutskie (2011) use the LEHD microdata to describe the workforce at startups.

not be located near the researcher – is prohibitive for many, particularly graduate students, young tenure track faculty, and policy makers. Thus our focus here in this paper is on the newly available public use statistics created from the LEHD data, which can be more easily accessed by the broader research community.

2.1 Newly Added Data on the Age of National Firm

Many of the new statistics we describe here were made possible due to a recent innovation in the LEHD microdata – newly linked data on the age and size of the national firm. Firm age and firm size in the LEHD data are sourced from the same underlying microdata as the Business Dynamics Statistics. These source data contain annual longitudinal firm-level information for nearly the entire nonfarm private economy and a small portion of the public sector from 1977 through the present. Unique features of this data are the longitudinal identification and analysis of firms through ownership and structural changes such as mergers, thus yielding more accurate estimates of firm age and firm size.[6]

Firm age in the LEHD data is defined as the age of the oldest establishment in the firm, similar to the BDS. An establishment is age zero in the first year that it reports any positive payroll, and ages naturally thereafter. The firm's age in the data is robust to mergers, spinoffs, and ownership changes. For example, a new legal entity (i.e., firm) that results from some M&A activity is not necessarily considered a young firm; instead, it is assigned the age of its

[6]For additional details on how the LEHD and BDS data were linked, and how age and size information were completed in the cases where the data could not be linked, see Haltiwanger et al. (2014).

oldest establishment at the time of its birth.

2.2 Newly Added Data on National Firm Size

The BDS-LEHD linkages described above also allow new public use statistics by national firm size to be produced from the LEHD microdata. Firm size is based on the total employment in all establishments belonging to the firm on March 12th of the previous year (or the current year for new firms). For any given consecutive two-year period, size is defined as the employment-weighted sum of firm size on March 12 in year $t-1$ of all establishments that are part of an EIN on March 12 in year t. This definition automatically covers mergers, divestitures, acquisitions, etc. For instance, if a firm in year t has three establishments belonging to three different firms in year $t-1$, initial firm size in year t is the weighted sum (where the weights are based on the year t size of each establishment) of the firm sizes in year $t-1$ of each of these three establishments. Note that public-use statistics do not cross-tabulate firm size by firm age, as the very sparse off-diagonal of the matrix generates large files of predominantly empty and sparsely populated cells (in other words, almost all younger firms are smaller firms, and the QWI tabulations are already quite granular–e.g. young women in mining in Connecticut).

3 The Quarterly Workforce Indicators: New Statistics on Worker Demographics, Hiring, Turnover, and Earnings at Young Businesses

Recent enhancements to the Quarterly Workforce Indicators (QWI) provide a unique lens into the workforce of startups. The QWI are a set of thirty-two economic indicators that provide information on employment, hires and separations, business expansion and contraction, as well as earnings for the universe of unemployment insurance taxable employment in the U.S. These data are available by worker demographics and firm characteristics, as well as at great levels of detail by geography (county and Workforce Investment Board Area) and industry (NAICS Industry Groups, i.e., 4-digit NAICS codes).

Since 2013, the QWI have used the recently integrated firm age and size data from the BDS to produce workforce statistics by worker demographics or industry categories for five firm age and size tabulation levels. These workforce statistics are available at the state, metro area, and county level. Tabulations by firm age and size at the detailed industry level (3 or 4-digit NAICS) can only be done at the statewide level. While the ability to examine employment growth at startups is not a unique feature of the QWI (this can also be done with the BDS and the BED), several indicators are uniquely available in the QWI: earnings at startups, earnings of new hires at startups, hires, separations, and turnover. And as we show in the next example, the QWI are unique in allowing the composition of the startup workforce to be examined – e.g. the share of young workers, of women, of racial minorities or highly-educated workers employed at startups.

3.1 Who Works at Startups?

We now present some basic facts from the QWI, but these only scratch the surface of what can be learned from these public-use tabulations. In Table 1, we compare the workforce composition of startups to that of more established businesses.[7] If we define businesses of age 0-1 years as "startups", we can see how the workforce at startups compares to that of older firms. Comparing the percentages across the columns in Table 1a, we see that startups disproportionately employ more young workers. Females and lower educated workers are also somewhat more likely to be employed at startups. Making similar comparisons in Table 1b, we likewise notice that startups are more likely than average to employ workers who are Asian, as well as those who are Hispanic.

Table 2 shows the employment shares for startups and older firms, specifically, by NAICS supersector. The data reveal that employment at startups is concentrated in particular industries. For example, 24.9% of all employment at firms aged 0-1 years is in the Leisure and Hospitality supersector, a rate that is almost double its share of employment overall. Other Services is another sector where the employment share at startups outweighs the share at more established firms. In other industries, entry rates are much lower, and the share of employment at young businesses is correspondingly lower in these sectors.

---

[7]The data include all states except Massachusetts. For all states, the data reference 2013Q3, except for some states where the most recent quarter of data have not been released: KY (2013Q2), WY (2012Q4), CA (2012Q2), FL (2012Q4).

3.2 Trends in the Employment of Startups, 2000-2012

We now turn to an analysis of the decline in the employment at startups, a topic discussed in Haltiwanger et al. (2012), Hyatt and Spletzer (2013), Decker et al. (2014a,b), and Dinlersoz et al. (2014). Specifically, we begin the analysis in the year 2000, after which the employment share of startups began to decline and the wages they pay eroded. Another reason for starting in 2000 is that most of the states in the statistics above had entered the program as of that time, thus the analysis can be conducted on a balanced panel.[8] Our goal of this subsection is to show how the published QWI statistics can be used to account for changes in economic indicators based on the composition of the firms or the workforce. We consider the share of employment at startups, the trend in the wage penalty at startups, as well as measures of employment reallocation: job creation, job destruction, hires, and separations (these last four measures are defined below).

First, we describe the trends, although the decompositions that follow will only reference the endpoints of the trends plotted in these figures, which are from 2000Q2 to 2012Q2. Figure 1 presents the trends in employment and earnings for two age categories: "startup" firms, those aged 0-1, and all other firms, i.e., those aged 2 or older. Figure 1a shows that the employment share at young firms has declined throughout the 2000s, consistent with the evidence in the studies referenced above. The earnings series in Figure 1b shows divergent trends for

[8]Different states enter the LEHD data at different times. The year 2000 was chosen as a starting point because most of the country is in the scope of the dataset by that year. The states included are AK, CA, CO, CT, DE, FL, GA, HI, ID, IL, IN, IL, IN, IA, KS, LA, ME, MD, MN, MO, MT, NE, NV, NJ, NM, NY, NC, ND, OH, OK, OR, PA, RI, SC, SD, TN, TX, UT, VT, VA, WA, WY, and WI. Comparisons are between 2000:Q2 and 2012:Q2. The year 2000 corresponds to the start of the job-to-job flows data, as described below. Furthermore, the year 2000 is a good starting point to consider the decline in entrepreneurial employment, see Dinlersoz et al. (2014).

young and old firms. The average earnings of workers at the youngest firms have declined in real terms throughout the 2000's, but the earnings at older businesses have shown a modest increase, consistent with what is documented in Haltiwanger et al. (2010) and Dinlersoz et al. (2014).

3.2.1 The Role of Composition

Information on the composition on the workforce by firm age can be used to answer questions related to the decline of startups and of business and employment dynamics more generally, a much discussed topic. Following Hyatt and Spletzer (2013), we can measure the effect of compositional changes using a standard decomposition technique to separate between-group differences from trends within groups for shares and earnings of startups (age 0-1) and all other businesses (age 2+), as follows. Any aggregate Y_t can be written as $\Sigma_i Y_{it} S_{it}$, where i indexes groups of the workforce or businesses (such as worker age or industry sector), and S_i is the share of the group. We decompose the difference $\Delta Y_t = Y_t - Y_{t-1}$ according to:

$$\Delta Y_t = \Sigma_i \Delta Y_{it} S_{i\cdot} + \Sigma_i Y_{i\cdot} \Delta S_{it}, \tag{3.1}$$

where $Y_{i\cdot}$ denotes the mean such that $Y_{i\cdot} = (Y_{it} + Y_{it-1})/2$, and likewise $S_{i\cdot}$. In other words, the decline in employment dynamics is equal to the change in the dynamics of each group weighted by the group's average employment share (the within effect) plus the change in each group's employment share weighted by the group's average measure of dynamics (the composition effect).

The first column of Table 3 contains the results of a conventional shift-share analysis for the change in the employment at young vs. old firms. The intuition for this analysis is that different types of workers may be different inputs to the production process, or that the demands for the output of different industries may lead to the shifts in business entry/exit rates for those industries. For example, younger workers may be more productive at startups, as in Ouimet and Zarutskie (2011) and Acemoglu et al. (2014), or have fewer resources to wait until a higher wage offer from an older firm as in Dinlersoz et al. (2014). As shown in Table 3, most of the changes in composition should have increased the share of startups, not decreased it, although the effects of changes in industry composition and worker demographics are fairly small. The main exception to this is the aging of the U.S. workforce, a demographic trend that does appear tied to the decline in employment share at startups – the preference of older workers for working at established businesses explains 9.4% of the decrease in the share of employment at startups.

Figure 1b shows the average real wage for workers who worked the entire quarter at startups and established firms, between 2000 and 2012. As can be seen in the graph, earnings at established firms are rising over this period while earnings at startups are falling. In the second column of Table 3, we decompose the rising wage premium at established firms by observable characteristics of firms and workers in the QWI. The formula for this composition change is slightly different, as it compares changes in two groups with each other. We plot the percentage that the changes in the shares in each of the two categories explain, given the average earnings for the categories, as follows:

$$\frac{\sum \Delta Share_{Old,x} * \overline{Earn}_x - \sum \Delta Share_{Young,x} * \overline{Earn}_x}{\Delta \overline{Earn}_{Old} - \Delta \overline{Earn}_{Young}} \tag{3.2}$$

This provides a measure of how the change in a share for a subset of a demographic characteristic (x), as well as in the average earnings for that particular demographic characteristic, is related to the change in earnings for young vs. old firms. Unlike our results for employment shares at startups, changes in industry composition and worker demographics explain a considerable part of the apparent increased wage premium for working at an established firm. For example, changes in the industry composition across young and older firms explains about one third of the decline in relative earnings at startups. Workers at established firms are also trending older and more educated, relative to younger firms, although as these effects are measured independently of the change in the industry distribution, they may be related.

In turn, Table 4 shows how the change in the composition of employment by firm age explains the decline in four earnings dynamics measures: hires, separations, job creation, and destruction. These measures exploit the dynamic aspect of the LEHD data: workers and business size are linked longitudinally to create these measures. This decomposition is again done according to equation 3.2 above. Results show that the shift away from entrepreneurship explains a substantial portion in the decline of such dynamics, due to the fact that startups are more volatile in terms of employment dynamics. The table shows that the decline in startups explains 9.3% of the decline in hires and 6.8% of the decline in separations. Additionally, the decline in startups explains 25.8% of the decline in job creation, but only 9.5% of the decline in job destruction. These results are similar to what Hyatt and Spletzer (2013) found using the LEHD micro data, and to what Decker et al. (2014b) found using the BDS.

4 Job-to-Job Flows: New Statistics on Worker Reallocation at Young Firms

At the time of this writing, the U.S. Census Bureau has just begun releasing new beta national statistics on worker reallocation constructed from the LEHD data. Job-to-Job Flows (J2J) provide national rates of job change for workers in the U.S., as well as flows of workers in and out of nonemployment. Origin-destination data on job-to-job flows provide new data on worker reallocation across industries and the economic migration of worker across state lines. These very unique data allow a comprehensive look at the reallocation of workers and provides useful additional metrics on the overall health of the U.S. labor market. Unemployed workers are not the only labor market participants seeking new jobs – upticks and declines in the rate at which employed workers find better jobs is another key indicator of the health of the labor market. A full description of these new data, as well as the methodology for deriving the worker flow estimates from the LEHD administrative data, is available in Hyatt et al. (2014).

4.1 Where do Startups' Early Employees Come From?

Of principal interest to researchers studying entrepreneurship will be the J2J data on worker flows into startups. Worker flows into startups can be used to identify the makeup of firms' early employees – their age, gender, and race, as well as the industry and age of the firm from which they moved.[9] The data can also be used to compare how startups grow relative to more established firms. Figure 2 provides an example by comparing worker flows across three classes of employers: established firms (firms over 10 years of age), small firms (firms with fewer than 20 employees), and very young firms (in business less than two years). Employment growth in each class of employer is the net of hires and separations, here split into two types: hires and separations associated with job change (with little-to-no intervening nonemployment spell between jobs), and hires/separations associated with worker flows to and from longer nonemployment spells. This decomposition allows us to see how firms grow – either by poaching workers away from other firms or by hiring more unemployed workers. As shown in Figure 2a and 2b, both established firms and small firms grow largely by hiring more workers from nonemployment than they lay off.[10]

Compare the previous pattern to that shown in Figure 2c, which shows the same decomposition at very young firms, those less than two years old. Compared to both established and small firms, very young firms attract a sizable share of their workforce from other firms. That was particularly true in the 2000-2002 period, when half of new firm hires were workers moving from other jobs. Very young firms gain a significant share of their early employment

[9] While this level of analysis is not available in the first beta release, future releases of data will include origin-destination statistics on worker flows by age and size of the employer and more detailed tabulation levels (for example, firm age by industry).

[10] Another interesting thing to note when comparing Figures 2a and 2b is that higher worker churn at small firms is not driven by employees changing jobs at a higher rate, but driven by flows to and from nonemployment. This could be due either to a higher rate of business failure for small firms generally or a greater reliance of small business operators on marginally attached workers.

growth by poaching workers away from more established firms. However, we should note (not shown) that by the time firms are 2-3 years old, this pattern changes. At 2-3 years old, firms begin to lose employment, on net, from workers changing jobs, with net employment growth coming largely from flows to and from nonemployment. This may imply that the high early rate of growth from poaching from more established firms is being driven by the startup team itself, as they move from their previous jobs to the new firm.

5 Data on Self-Employment and its Potential for Statistics on Entrepreneurship

Many business starts (and failures) occur before a business hires its first employees. Such small owner-operated businesses are not included in statistics such as the BDS and QWI, where business birth is defined as the moment the firm hires its first worker. Business owners themselves are interesting to researchers focused on identifying the characteristics of successful entrepreneurs, and on understanding why the rate of startups has declined. Efforts are currently underway to enhance the set of available data on business founders and self-employed businesses by integrating data on sole-proprietors with the LEHD data infrastructure. A prototype micro-data file has been created which covers the universe of active U.S. sole-proprietorships, both with and without employees, from 2002 through 2012. The Census Bureau is undergoing research into using these data for new public use statistics on business founders and self-employed businesses.

Sole-proprietor businesses can be characterized as any unincorporated business owned and run by a single individual. The data that we integrate originate primarily from individual federal income tax returns, such as income filings from Schedule C, payroll tax records for employers (form 941), and applications for an Employer Identification Number (EIN) for employers (form SS-4). The scope of our data does not include partnerships and corporations, although single-owner Limited Liability Companies (LLCs) are included as long as they do not elect to be taxed as a corporation. More details on how the data are linked is provided in Garcia-Perez et al. (2013).

By combining these administrative data on the universe of sole-proprietor business owners with the universe of covered wage and salary work, the resulting dataset permits us to observe an owner's pre-ownership wage and salary work history, and thus to potentially generate statistics based on prior employment, earnings, and industry experience. Such a link should prove enlightening in the context of the well-documented decline in U.S. startups, which has sparked much interest in the underlying causes and implications of this slowdown. Although the overall trend in startups may be downward, in reality the composition of new business owners is constantly in flux, with certain types of individuals exhibiting differing and perhaps offsetting trends. To understand the decrease in startups requires knowledge of the factors that precede a business and an understanding of how these factors influence the odds of a successful startup. For example, the self-employment literature recognizes that some are "pushed" into self-employment by lack of economic opportunity while others are "pulled" into entrepreneurship by means of comparative advantage or innovative idea. Linked sole-proprietor and LEHD data offer a way to help parse such differences in the paths of potential entrepreneurs.

5.1 Don't Quit Your Day Job: A Look at Self-Employment Dynamics

What share of self-employed businesses grow enough to allow the owner to leave wage and salary employment? The left-hand panel of Table 5 shows the percentage of sole-proprietors in 2009 who are engaged in wage and salary work in the same year, as well as in the surrounding years of 2008 and 2010. The first thing that stands out is that the majority of self-employed businesses without employees do not grow enough to supplant the owner's reliance on some form of wage and salary work. Over 50% of nonemployer business owners in 2009 have wage and salary income in that year, a share that is higher for new nonemployer business owners (those in the first year of their business), at around 65%. For new employers in 2009, defined as employer businesses who were not employers in 2008, about 40% had wage and salary jobs in 2008, 35% have such employment in the 2009 year (the birth year of their employer business), and 30% retain it in the following year 2010. For more established business owners with employees, the wage and salary work rate stabilizes at just above 20%.

For employer business owners, we can also capture their experience as operators of businesses without paid employees. In the right-hand panel of Table 5, we see that amongst new employer business owners in 2009, around 36% operated a nonemployer business in the previous year. This rate falls by over half to 17% during their first year of employer business activity in 2009, suggesting that it may represent the same businesses that are transitioning as they acquire employees. Note that the percentage of new 2009 employers with nonemployer income rises again in 2010 to 24%, perhaps indicating that some new employer businesses have shed their employee within one year, but nevertheless maintained the business. Note again that the rate of nonemployer business holding amongst all employers remains in the 15-20% range, meaning that a substantial fraction of owners maintain other sources of business income simultaneous to running an employer business.

6 Conclusion

In this paper, we highlight new data on business startups, their founders and early employees, and provide some examples of how they can be used in entrepreneurship research. The Quarterly Workforce Indicators, the new Job-to-Job Flows statistics, and the future sole-proprietor data products, offer a unique perspective on new business creation by combining information on the business, workforce, and individual. With these data, we can describe the composition of the workforce at startups and their role in explaining business dynamics, the flow of workers across new and established firms, and the employment paths of the business owners themselves.

While the emphasis here has been on public-use products, further opportunities are available to those who want to conduct studies using the confidential microdata in the LEHD system. Researchers may apply to use these data at the many Census Research Data Centers (RDCs) across the country. Many of the datasets underlying the products described here are available for approved microdata research projects in the secure RDCs.

We emphasize that these data products are either new or under development. Feedback from users can lead to enhancements of these data products, and in some cases, affect the path of development of new data products.

References

J. Abowd, B. Stephens, L. Vilhuber, F. Andersson, K. McKinney, M. Roemer, and S. Woodcock. The LEHD infrastructure files and the creation of the quarterly workforce indicators. In T. Dunne, J. Jensen, and M. Roberts, editors, *Producer Dynamics: New Evidence from Micro Data*, chapter 9, pages 149–230. NBER/University of Chicago Press, 2009.

D. Acemoglu, U. Akcigit, and M. Celik. Young, restless, and creative: Openness to disruption and creative innovation. *NBER Working Paper*, 19894, 2014.

R. Agarwal, B. Campbell, A. Franco, and M. Ganco. What do I take with me?: The mediating effect of spin-out team size and tenure on the founder-firm performance relationship. *U.S. Census Bureau, Center for Economic Studies Working Paper*, 2013.

T. Bates. Entrepreneur human capital inputs and small business longevity. *The Review of Economics and Statistics*, 72(4):551–559, 1990.

N. Brinkman, D. Cohen-Pirani, and H. Stieg. Firm dynamics in an urban economy. *Internationl Economic Review (forthcoming)*, 2015.

D. Carlton. The location and employment choices of new firms - An econometric model with discrete and continuous endogeneous variables. *Review of Economics and Statistics*, 1983.

S. Chatterjee and E. Rossi-Hansberg. Spinoffs and the market for ideas. *International Economic Review*, 53(1):53–93, 2012.

B. Chinitz. Contrasts in agglomeration: New York and Pittsburgh. *American Economic Review*, 51(2): 279–289, 1961.

S. Davis, J. Haltiwanger, C.J. Krizan, J. Miranda, A. Nucci, and L.K. Sandusky. Measuring the dynamics of young and small businesses: Integrating the employer and nonemployer universes. In Roberts Dunne and Jensen, editors, *Producer Dynamics: New Evidence from Micro Data*, chapter 9. NBER: University of Chicago Press, 2009.

R. Decker. Collateral damage: Housing, entrepreneurship, and job creation. *Unpublished Manuscript: University of Maryland*, 2014.

R. Decker, J. Haltiwanger, R. Jarmin, and J. Miranda. The role of entrepreneurship in U.S. job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3):3–24, 2014a.

R. Decker, J. Haltiwanger, R. Jarmin, and J. Miranda. The secular decline in business dynamism in the U.S. *Unpublished Manuscript: University of Maryland*, 2014b.

F. Delmar and S. Shane. Does experience matter? The effect of founding team experience on the survival and sales of newly founded ventures. *Strategic Organization*, 4(3):215–247, 2006.

E. Dinlersoz, H. Hyatt, and H. Janicki. Who works for whom? Worker sorting in a model of entrepreneurship with heterogeneous labor markets. *Unpublished draft: U.S. Census Bureau*, 2014.

M. Doms, E. Lewis, and A. Robb. Local labor force education, new business characteristics, and firm performance. *Journal of Urban Economics*, 67(1), 2010.

G. Ellison and E. Glaeser. Geographic concentration in U.S. manufacturing industries: A dartboard approach. *Journal of Political Economy*, 104(5):899–927, 1997.

D. Evans and L. Leighton. Some empirical aspects of entrepreneurship. *American Economic Review*, 79 (3):519–535, 1989.

R. Fairlie. Self-employment, entrepreneurship, and the NLSY79. *Monthly Labor Review*, February:40–47, 2005.

O. Figueiredo, P. Guimares, and D. Woodward. Firm-worker matching in industrial clusters. *Journal of Economic Geography*, 14(1):1–19, 2014.

T. Fort, J. Haltiwanger, R. Jarmin, and J. Miranda. How firms respond to business cycles: The role of firm age and firm size. *Census Bureau Center for Economic Studies Papers*, 13(30), 2012.

A. Franco and D. Filson. Spin-outs: Knowledge diffusion through employee mobility. *RAND Journal of Economics*, 37(4):841–860, 2006.

M. Freedman. Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. *Journal of Urban Economics*, 2008.

M. Garcia-Perez, C. Goetz, J. Haltiwanger, and L.K. Sandusky. Don't quit your day job: Using wage and salary earnings to support a new business. *Census Bureau Center for Economic Studies Papers*, 13 (45), 2013.

E. Glaeser, H. Kallal, J. Scheinkman, and A. Shleifer. Growth in cities. *Journal of Political Economy*, pages 1126–1152, 1992.

E. Glaeser, W. Kerr, and G. Ponzetto. Clusters of entrepreneurship. *Journal of Urban Economics*, 67(1): 150–168, 2010.

J. Haltiwanger, R. Jarmin, and J. Miranda. Who creates jobs? Small vs. large vs. young. *NBER Working Paper*, 16300, 2010.

J. Haltiwanger, R. Jarmin, and J. Miranda. Where have all the young firms gone? *Kaufman Foundation Paper Series*, 2012.

J. Haltiwanger, H. Hyatt, E. McEntarfer, L. Sousa, and S. Tibbets. Firm age and size in the Longitudinal Employer-Household Dynamics data. *Center for Economic Studies Discussion Paper*, 14(16), 2014.

B. Hamilton. Does entrepreneurship pay? An empirical analysis of the returns to self employment. *Journal of Political Economy*, 108(3):604–631, 2000.

E. Hurst and B. Pugsley. What do small businesses do? *Brookings Papers on Economic Activity*, 2011.

H. Hyatt and J. Spletzer. The recent decline in employment dynamics. *IZA Journal of Labor Economics*, 2(1):1–21, 2013.

H. Hyatt, E. McEntarfer, K. McKinney, S. Tibbets, and D. Walton. Job-to-job flows: New labor market statistics from linked employer-employee data. *Joint Statistical Meetings Papers and Proceedings*, forthcoming, 2014.

R. Kihlstrom and J. Laffont. A general equilibrium entrepreneurial theory based on risk aversion. *Journal of Political Economy*, 87:719–848, 1979.

S. Kitao. Entrepreneurship, taxation and capital investment. *Review of Economic Dynamics*, 11(1):44–69, 2008.

S. Klepper. Employee startups in high-tech industries. *Industrial and Corporate Change*, 10(3):639–674, 2001.

R. Lucas and E. Rossi-Hansberg. On the internal structure of cities. *Econometrica*, 70(4):1445–1476, 2002.

P. Ouimet and R. Zarutskie. Who works for startups? The relation between firm age, employee age, and growth. *Center for Economic Studies Working Papers*, 11(31), 2011.

S. Parker. A selection-based theory of the transition from employment to entrepreneurship: The role of employer size. *IZA Discussion Paper*, 2071, 2006.

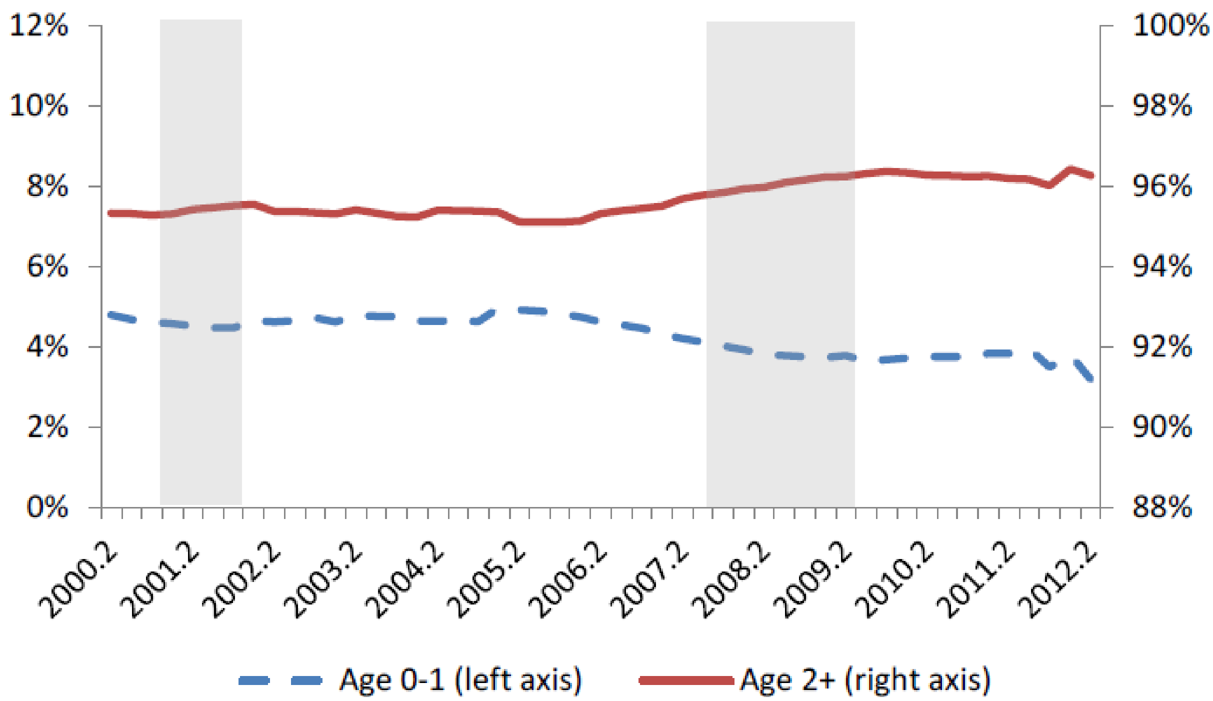
S. Rosenthal and W. Strange. Geography, industrial organization, and agglomeration. *The Review of Economics and Statistics*, 85(2):377–393, 2003.

A. Saxenian. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press, 1994.

A. Schoar. The divide between subsistence and transformational entrepreneurship. In Lerner and Stern, editors, *NBER Innovation Policy and the Economy*. University of Chicago Press, 2009.

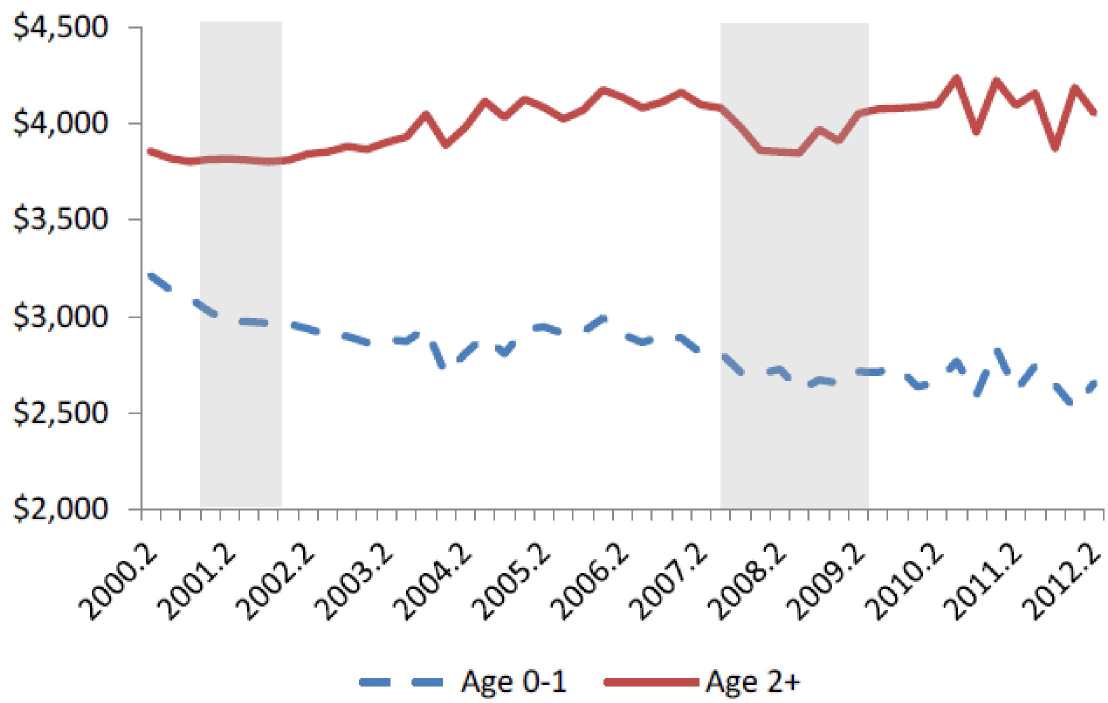
J. Schumpeter. *Capitalism, Socialism, and Democracy*. 1939.

Figure 1a: Employment Shares by Firm Age



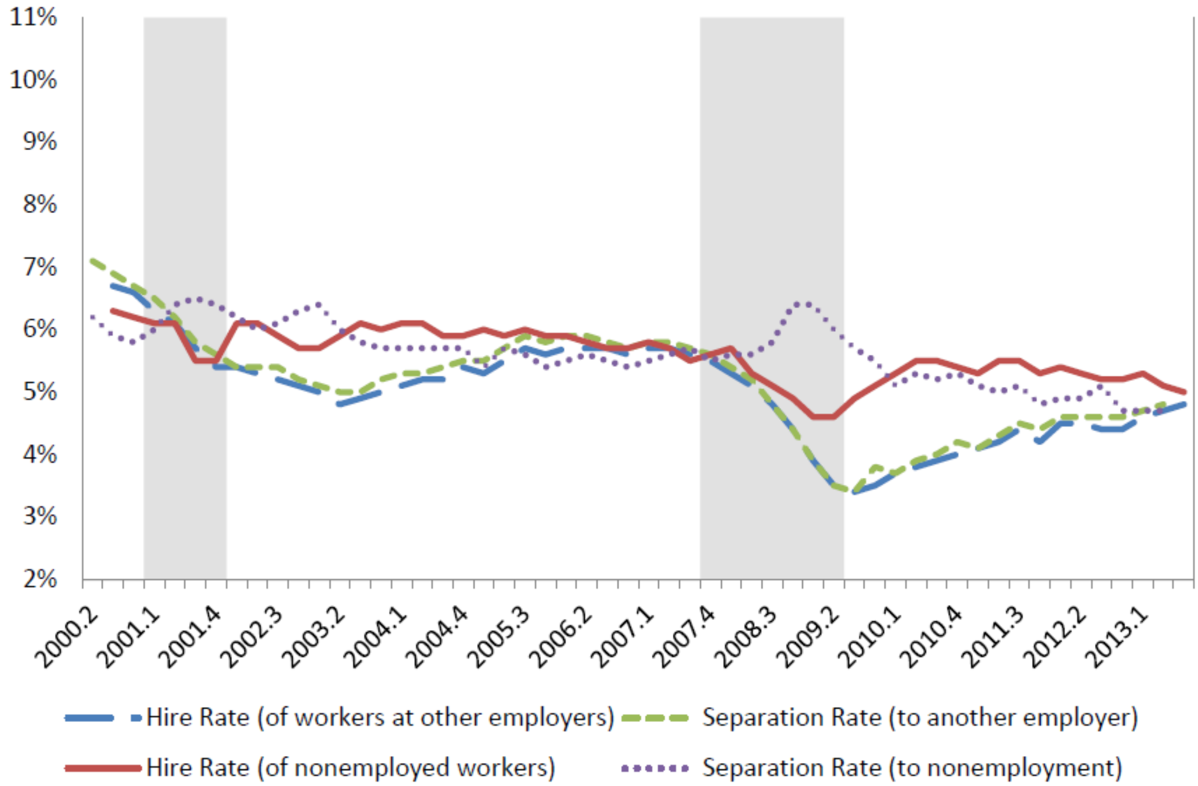
Note: Author's calculations of the Quarterly Workforce Indicators.

Figure 1b: Real Quarterly Earnings by Firm Age



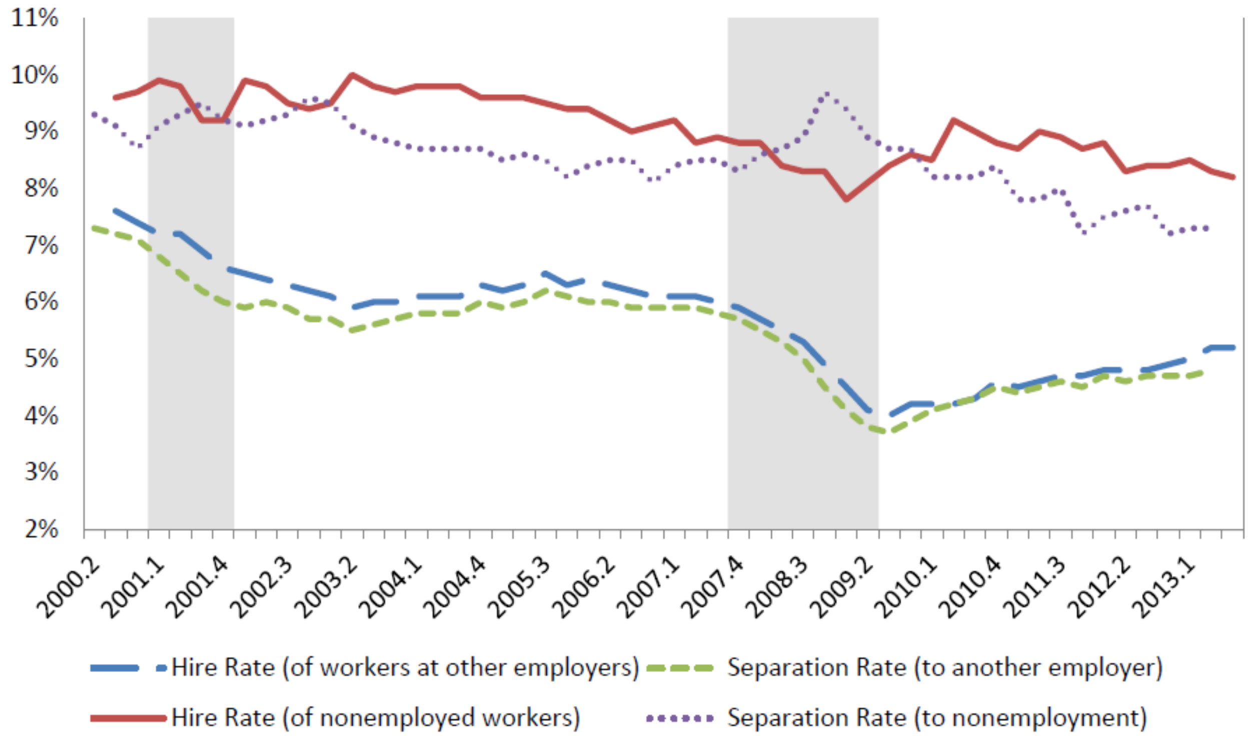
Note: Author's calculations of the Quarterly Workforce Indicators. Earnings are expressed in 2009 dollars.

Figure 2a: Hires and Separations at Established Firms (11+ years old) 2000-2013



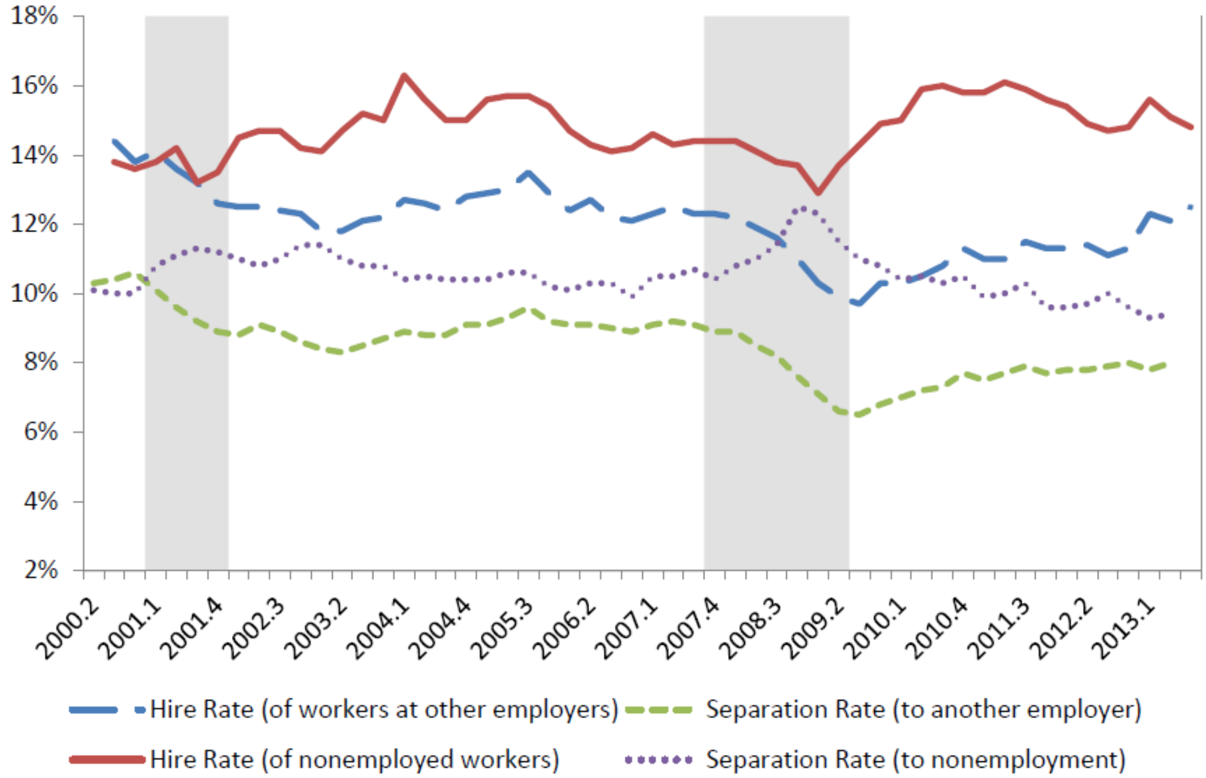
Source: Census Bureau's beta Job-to-Job Flows 2013Q3 release.

Figure 2b: Hires and Separations at Small Firms (<20 Employees) 2000-2013



Source: Census Bureau's beta Job-to-Job Flows 2013Q3 release.

Figure 2c: Hires and Separations at Young Firms (0-1 years old) 2000-2013



Source: Census Bureau's beta Job-to-Job Flows 2013Q3 release.

Table 1a: QWI Employment Shares by Demographics

	All Firms	0-1 Years	2-10 Years	11+ Years
by Age				
Age 14-24	14.5%	20.2%	17.6%	13.6%
Age 25-44	43.4%	45.0%	46.2%	42.7%
Age 45-64	37.2%	30.5%	32.6%	38.6%
Age 65-99	4.9%	4.3%	4.5%	5.0%
by Sex				
Men	52.0%	49.0%	51.2%	52.3%
Women	48.0%	51.0%	48.8%	47.7%
by Education				
Less than High School	12.2%	14.7%	13.5%	11.8%
High School	23.9%	22.3%	23.0%	24.2%
Some College	26.9%	24.2%	25.5%	27.4%
Bachelor's Degree or Higher	22.4%	18.6%	20.9%	23.0%
Education Not Available (age 24 or less)	14.5%	20.2%	17.0%	13.6%
Total All Workers	100%	3.5%	16.9%	79.5%

Note: Private sector employment counts for all U.S. states (except Massachusetts) and the District of Columbia as of 2013:Q3 or most recent quarter available by state. The "2-10 Years" column aggregates the "2-3 Years," the "4-5 Years" and the "5-10 Years" categories from the published data. The "Age 14-24" age row aggregates the "Age 14-18", the "Age 19-21" and the "Age 22-24" categories from the published data. Likewise, the "Age 25-44" row aggregates the "Age 25-34" and the "Age 35-44" categories, and the "Age 45-64" row aggregates the "Age 45-54" and "Age 55-64" from the published data. See text for additional details.

Table 1b: QWI Employment Shares by Demographics

	All Firms	0-1 Years	2-10 Years	11+ Years
by Race				
White Alone	79.4%	76.6%	78.9%	79.6%
Black or African American Alone	12.3%	11.7%	11.0%	12.6%
American Indian or Alaska Native Alone	0.9%	1.1%	1.1%	0.9%
Asian Alone	5.5%	8.3%	6.9%	5.0%
Native Hawaiian or Other Pacific Islander Alone	0.3%	0.3%	0.3%	0.2%
Two or More Race Groups	1.6%	2.0%	1.9%	1.6%
by Ethnicity				
Not Hispanic or Latino	86.1%	83.3%	84.2%	86.7%
Hispanic or Latino	13.9%	16.7%	15.8%	13.3%
Total All Workers	100.0%	3.5%	16.9%	79.5%

Note: Private sector employment counts for all U.S. states (except Massachusetts) and the District of Columbia as of 2013:Q3 or most recent quarter available by state. The "2-10 Years" column aggregates the "2-3 Years," the "4-5 Years" and the "5-10 Years" categories from the published data. See text for additional details.

Table 2: QWI Employment Shares by Industry

	All Firms	0-1 Years	2-10 Years	11+ Years
by NAICS Supersector				
Natural Resources and Mining	1.8%	2.0%	2.5%	1.6%
Construction	5.1%	6.0%	6.9%	6.5%
Manufacturing	10.7%	4.7%	5.3%	12.1%
Trade, Transportation, and Utilities	22.7%	14.1%	16.1%	24.5%
Information	2.5%	1.3%	1.5%	2.7%
Financial Activities	6.8%	4.3%	5.0%	7.3%
Professional and Business Services	16.5%	15.3%	19.6%	15.9%
Education and Health Services	17.3%	12.9%	17.1%	17.5%
Leisure and Hospitality	12.6%	24.9%	19.7%	10.6%
Other Services	4.1%	14.6%	5.6%	3.3%
Public Administration	0.0%	0.0%	0.0%	0.0%
Total All Workers	100%	3.5%	16.9%	79.5%

Note: Private sector employment counts for all U.S. states (except Massachusetts) and the District of Columbia as of 2013:Q3 or most recent quarter available by state. Note that because firm age assignment is limited to the private sector, the Public Administration Supersector contains zero employment in each age category by construction. The "2-10 Years" column aggregates the "2-3 Years," the "4-5 Years" and the "5-10 Years" categories from the published data. Supersector rows are aggregates of published NAICS sector statistics. See text for additional details.

Table 3: Employment Composition on Differences in Employment and Earnings, 2000Q2 vs. 2012Q2

	Employment	Startup Earnings Penalty
Sex	0.1%	3.5%
Age	9.4%	11.1%
Education	-0.3%	15.4%
Race	0.0%	0.8%
Ethnicity	-1.2%	2.3%
Industry	-10.9%	33.4%

Note: Author's calculations of the Quarterly Workforce Indicators. Employment shares and comparisons are of those age 0-1 in the Quarterly Workforce Indicators, versus those age 2 or older. See text for exact formulas.

Table 4: Change in Employment Dynamics due to Decline in Startups: 2000-2012

	Hires	Separations	Job Creation	Job Destruction
2000Q2	30.0%	27.1%	8.6%	5.7%
2012Q2	20.5%	17.4%	7.1%	4.0%
Change	-9.5%	-9.7%	-1.5%	-1.7%
Percent of Change explained Firm Age:	9.3%	6.8%	25.8%	9.5%

Note: Author′s calculations of the Quarterly Workforce Indicators. See text for formulas.

Table 5: Employment Status of 2009 Business Owners in Years 2008-2010

	Percentage with Wage & Salary Income			Percentage with Nonemployer Income			
Type of 2009 Business Owner	2008	2009	2010	2008	2009	2010	N
New Employers	40	35	30	36	17	24	86,011
All Employers	21	20	21	18	15	22	721,807
New Nonemployers	68	65	62	0	100	52	6,158,104
All Nonemployers	54	51	50	66	100	69	17,912,997

Note: Table reports percentages of sole-proprietor business owners in 2009 of a given type that also have positive income from wage & salary work and/or nonemployer activity in the years 2008-2010. Sample comprises all observed owner-year pairs of a given business type during 2009. “New Employers” are defined as owners who have positive income from an employer business in year 2009, but no such income in year 2008. Similarly, “New Nonemployers” are those who have nonemployer business income in 2009, but no such income in 2008.