The Promise and Potential of Linked Employer-Employee Data for Entrepreneurship Research

Christopher Goetz, Henry Hyatt, Erika McEntarfer, and Kristin Sandusky

Abstract

In this paper, we highlight the potential for linked employer-employee data to be used in entrepreneurship research, describing new data on business start-ups, their founders and early employees, and providing examples of how they can be used in entrepreneurship research. Linked employer-employee data provides a unique perspective on new business creation by combining information on the business, workforce, and individual. By combining data on both workers and firms, linked data can investigate many questions that owner-level or firm-level data cannot easily answer alone - such as composition of the workforce at start-ups 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.

Christopher Goetz, Henry Hyatt, Erika McEntarfer, and Kristin Sandusky are research economists at the Center for Economic Studies (CES) at the U.S. Census Bureau.

We thank Rajshree Agrawal, Hubert Janicki, Shawn Klimek, and participants in the NBER-CRIW Entrepreneurship Conference for very helpful comments on this chapter. We also thank Douglas Walton and Alexandria Zhang for assistance in preparing some of the tabulations. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. While most of the figures and tables in this paper are calculated from public use data, some tables use confidential Census Bureau microdata, all such tables and figures have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

Linked employer-employee data fill an important gap in the set of data to study entrepreneurship, shedding light on questions that cannot be addressed using firm or individual-level data alone. For researchers interested in start-up firms and their founders, data identifying the transition of the entrepreneur from the workforce to founder of a new firm is of inherent interest. How workers move from being employees to entrepreneurs, who they recruit for start-up teams, and what predicts starts, successes, and failures is key to understanding the dynamics of entrepreneurial activity in the United States. Policymakers are interested in entrepreneurship in part because they are interested in job growth. Linked employer-employee data show who works for new firms and whether these firms are creating "good" jobs. Labor market agglomeration effects are widely acknowledged to be important in the spatial clustering of technological or innovative industries. Yet labor market flows across firms are difficult to understand with business or household-level data alone.

In this paper, we discuss the potential of linked employer-employee data to study entrepreneurship, and provide a road map for researchers interested in using these data. We will discuss both the confidential microdata and public use data derived from linked employer-employee data. Linked employer-employee microdata for the U.S. are currently available to approved researchers working in restricted data centers. However, the Census Bureau has recently stepped up efforts to create new public use data about young firms using linked employer-employee data as part of the Longitudinal Employer-Household Dynamics (LEHD) program. The result is new public use data on workforce composition, hiring, turnover, and earnings paid to workers at young firms. Because these new statistics are sourced from administrative data, they are available at much finer geographic and industry detail than is usually available in public use statistics. While lacking the flexibility of the confidential microdata, these new statistics bring many of the benefits of the linked employer-employee data into the public domain for easier research access.

Specifically, our goals in this paper are threefold:

(1) To familiarize researchers with the U.S. linked employer-employee data and how it can be used in entrepreneurship research;

(2) To describe newly available public use statistics derived from linked employer-employee data and provide examples of how it can be used to study entrepreneurship; and

(3) To outline future plans to expand the set of available data to study entrepreneurship by linking in new administrative data sources on self-employment and partnerships, as well as identifying the employment history and human capital formation of entrepreneurs themselves.

This chapter begins with a brief overview of the current landscape of data available for empirical research on entrepreneurship. We then describe the linked employer-employee microdata in more detail, and provide information on how to access the data. Subsequent sections describe new public use data tabulated from the linked employer-employee data, and provide specific examples of how it can be used to study workforce and earnings dynamics in new firms. The last section of the paper outlines a vision for future work to build new statistical infrastructure to support entrepreneurship research from linked worker-firm administrative data.

2 An Overview of Available Data for Entrepreneurship Research

Entrepreneurship has long been acknowledged to play an important role in modern economies by spurring innovation, creating jobs, and enhancing productivity. However, only in the last few decades has entrepreneurship flourished as a research area within economics. Data on entrepreneurial activity are necessary for any empirical research on determinats of entrepreneurship and the impact of entrepreneurship on the economy. Yet the existing statistical infrastructure is in many ways inadequate to investigate questions around business formation and innovative activity. Despite several new data sources made available in the last decade, many important data gaps remain.

Currently available data to study entrepreneurship include firm-level or owner-level microdata, as well as published aggregate statistics. **Table 1** details the most commonly used publically available data in entrepreneurship research. Information on entrepreneurs typically comes from household- or business-level surveys, mostly as cross-sectional snapshots, although a few smaller panel datasets are available. The Current Population Survey (CPS), the Survey of Consumer Finance (SCF), the Panel Study of Income Dynamics (PSID), and the National

Longitudinal Surveys (NLS), and the other household surveys listed here ask a similar small set of questions concerning self-employment and business ownership.[1] Data on both founders and their businesses are available in the Census Bureau's Survey of Business Owners (SBO), and the Kauffman Firm Survey (KFS). With regard to business-level data on new firms, statistics on start-ups and established firms are available in the Business Dynamics Statistics (BDS) and the Business Employment Dynamics (BED). The creation of the BDS and BED has led to a growth of research documenting the importance of new businesses for job creation and economic growth. The Quarterly Workforce Indicators (QWI), derived from LEHD microdata, are a relatively recent addition to this list which we will describe in greater detail later in this paper.

Most existing data sources are limited in their ability to depict the interaction between start-ups and their human assets, including owner, founding team members and early employees. The omission of human capital, which can strongly influence both the nature and the success of a new business, increasingly leaves researchers of entrepreneurship at a disadvantage as the U.S. economy becomes more service-oriented and knowledge-based. Data that contain information on owners or workers are typically unable to follow the business over time, or else only provide dynamic information on a limited sample of business entrants. These shortcomings make it difficult to study the impact of factors such as owner characteristics and experience on the outcomes of start-ups, and measure the potentially changing effects over time.

The scope of entrepreneurship research is fairly broad, but there are many research questions for which longitudinally linked employer-employee data is especially useful. **Table 2** lists some of the broad questions in the field of entrepreneurship research (along with a selection of representative studies), and with some specific examples of how linked employer-employee data can be employed in the study of these topics. For instance, several researchers have noted that young firms typically hire younger workers (e.g., Ouimet and Zarutski, (2014)) spawning wider interest in exploring how labor-related factors can influence the success of new ventures. Detailed data on labor market flows across firms are well suited for investigating subjects like agglomeration economies, labor market spillovers, and spinoff firms (e.g., Agarwal et al. (2013), using LEHD microdata). Highly spatial public use data on young firms by detailed industry can

[1] For a summary of studies using the National Longitudinal Survey of Youth 1979 (NLSY79) to study entrepreneurship, see Fairlie (2005).

help explain why regional growth appears to be correlated with the presence of many small/young firms (e.g., Glaeser, Kerr, and Ponzetto (2010)). Data linking business owners and their employment histories can help identify the determinants of entrepreneurship and new business success, a large literature that includes the work of Evans and Leighton (1989), Hurst and Lusardi (2004), and Hamilton (2000). Planned integration of self-employment data with linked employer-employee data would enable further investigation into the distinction between types of entrepreneurship. As only a small subset of entrepreneurs begin their businesses with an intent to grow, identifying potential high-growth entrepreneurs is of great economic and policy interest (e.g., Hurst and Pugsley (2011) and Chaterjee and Rossi-Hansberg (2012)).

3 The Longitudinal Employer-Household Dynamics (LEHD) Data

The Longitudinal Employer-Household Dynamics (LEHD) program at the U.S. Census Bureau has built over the last decade a comprehensive linked employer-employee dataset for the United States. The result of this effort is a comprehensive longitudinal database covering over 95% of U.S. private sector jobs and most public sector employment.

The LEHD data system is extraordinarily complex, linking data across multiple agencies, blending administrative and survey data, and filling data gaps with additional source data whenever possible. The LEHD job-level data comes primarily from quarterly worker-level earnings submitted by employers for the administration of state unemployment insurance (UI) benefit programs. Information on federal jobs (not covered by state UI programs) is provided to Census by the Office of Personnel Management (OPM).[2] These job-level records are linked to establishment-level data collected for the Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW) and Census Bureau's Longitudinal Business Database (LBD) data to obtain further information about the employer. Demographic information about individual workers is obtained via links to Census surveys and Social Security administrative data. Residential information on workers comes primarily from Internal Revenue Service (IRS) address data. Ongoing work to integrate administrative data on self-employed workers is described later in this paper.

[2] State UI covers most private employment, as well as state and local government employment. There are notable exceptions to coverage, namely most small agricultural employers, religious institutions, and much of the non-profit sector. OPM federal employment data includes the civilian workforce, but not the armed forces or the postal service.

As is evident from the description above, the LEHD data relies on data sharing agreements with multiple state and federal agencies to provide critical inputs to the linked employer-employee data. Key among these are data sharing agreements between state governments and Census through the Local Employment Dynamics (LED) partnership. State agencies provide the principal job-level data (state UI records of employee-specific total quarterly wage and salary payments) as well as QCEW data. As of this writing, all 50 states, DC, Puerto Rico, and the Virgin Islands have provided data to the LEHD program through this partnership. Because states joined the partnership at different times with different amounts of data archived, the set of available states in the LEHD data varies by year; states with the longest panels have data that begin in the early 1990s, and the last state, Massachusetts, enters in 2010.

The voluntary nature of the data sharing agreements makes LEHD unique among statistical programs. While the LEHD program has been enormously successful in bringing together multiple agencies to share data to create universe-level data on jobs in the U.S., the voluntary nature of these agreements (state and federal partners receive no compensation for participation in the program) is a great risk to the long-term viability of the data program. Withdrawal of data-sharing partners from the program risks the integrity of many of the products provided from the LEHD data and the usability of the data for research. These data sharing agreements also have implications for researcher access to the confidential microdata, outlined in the next section.

The ability to identify firm age is a recent enhancement to the LEHD data, a highly valuable additional characteristic for researchers interested in entrepreneurship. Firm age is obtained via links to the microdata that underlies the Longitudinal Business Database (LBD), which also serves as the source data for the Census Bureau's Business Dynamics Statistics (BDS). As in the BDS, firm age is defined as the age of the oldest establishment in the national firm. An establishment is age zero in the first year that it reports any positive payroll, and ages chronologically thereafter. Firm age is robust to ownership changes such as mergers, spinoffs, and ownership changes. For example, a new legal entity spun off as a result of merger or acquisition activity will not be considered a new firm; instead, it is assigned the age of its oldest establishment at the time of its formation.

A comprehensive description of the LEHD data is available in Abowd et al. (2009). A detailed discussion of the methodology used to add firm age to the LEHD data is provided in Haltiwanger et al. (2014).

3.1 Researcher Access to LEHD Microdata

Researchers can apply for access to LEHD microdata by submitting a research proposal through the Federal Research Data Center (FRDC) 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 to the LEHD program. Projects approved to use the confidential microdata are conducted in a secure research data center with all output undergoing a formal disclosure review process before being permitted for dissemination outside the secure facilities.[3]

The proposal review process for LEHD confidential data access is complicated by the many data sharing agreements between data partners and the U.S. Census Bureau. Any FRDC proposal requesting access to IRS data must be approved by IRS (whether a proposal using LEHD data needs IRS approval depends on the data requested, but firm age, likely of critical interest to entrepreneurship researchers, is sourced from IRS data). State agreements vary, with some states choosing to allow their state data in pooled multi-state research samples for research projects approved by Census. Other state partners choose to review proposals and approve or deny data access on a project-by-project basis.[4]

In short, acquiring confidential LEHD microdata access for entrepreneurship research can be classified as a "high-cost/high-reward" activity. The scope of research projects that benefit from such rich microdata is vast. This is particularly true in the interdisciplinary field of entrepreneurship research, where many issues are fundamentally interactions between workers and firms. For instance, LEHD data allow identification of spin-off firms and the employment history of their start-up teams. Employment with start-up firms is considered a high-risk/high-reward career strategy – linked employer-employee data can measure both the earnings benefits

[3] More information on how to apply for confidential microdata research access through the FRDC network is available on the Center for Economic Studies website: https://www.census.gov/ces/.

[4] Under all LED data use agreements, any state or sub-state tabulation or estimate released from LEHD data must be approved by the state partner. Tables and estimates in research papers must have a minimum of three states contributing to the estimate or cell to avoid this requirement.

and risks of joining a start-up team. Acquiring talented employees is critical for start-up success – better understanding of how labor market agglomeration effects spur industry growth would help policy makers interested in spurring local entrepreneurship efforts. These examples obviously represent only a handful of possible topics for research using linked employer-employee data. Additionally, the LEHD microdata can be linked to other person and firm-level data, expanding the set of possible research questions even further.

Although LEHD microdata access offers the broadest possibilities for projects in entrepreneurship research, the relatively high cost of obtaining access to the data (writing a successful proposal, obtaining necessary approvals, possible travel to a research data center) is prohibitive for many researchers. This is especially true for younger researchers (e.g., graduate students, junior faculty). Policy makers and journalists interested in entrepreneurship often need quick answers to immediate questions. Thus, in the next few sections of this chapter we focus on new public use statistics on young firms created from the LEHD data, which can be accessed by the broader research and policy community.

3.2 LEHD Public Use Data for Entrepreneurship Research

In this section, we briefly describe three public use data products derived from LEHD microdata, with a focus on new data on firm age. In the following section, we illustrate the value of these statistics for entrepreneurship research by means of examples. **Table 3** provides an overall summary of this new data, including variables, frequency, and stratification levels, also highlighting the relative strengths of these statistics relative to other available data.

3.2.1 The Quarterly Workforce Indicators (QWI)

The Quarterly Workforce Indicators (QWI) are a set of thirty-two economic indicators providing employment, hires and separations, business expansion and contraction, as well as earnings for the universe of UI-covered employment in the U.S. Data are available by worker demographics (sex, age, education, as well as race and ethnicity) and firm characteristics (firm age, size) as well as at fine levels of detail by workplace geography (county and Workforce Investment Board area) and industry (highly detailed 4-digit NAICS codes).

QWI statistics by firm age are quite new (the first release was in 2013), made possible by the recent enhancements to the LEHD microdata discussed earlier in this chapter. The QWI provide data for five firm age tabulation levels, with the youngest firm category being firms less than two years old. While the ability to examine employment growth at young firms is not a unique feature of the QWI, several indicators are uniquely available in the QWI: earnings at start-ups, earnings of new hires at start-ups, hires, separations, and turnover.[5] Moreover, as we show in a later example, the QWI are unique in allowing the composition of the start-up workforce to be examined: for example, the share of young workers, of women, of racial minorities, or highly educated workers employed at start-ups.

3.2.2 LEHD Origin Destination Employment Statistics (LODES)

LEHD Origin Destination Employment Statistics (LODES) provide employment data by both place of work and place of residence at block-level geography. The ability to analyze employment by both place of residence as well as place of work is critical for identifying regional labor markets and understanding the interconnectedness of geographic areas that lie across state and metro area boundaries. A combination of noise infusion (similar to QWI) and synthetic data methods are used to protect worker and firm characteristics, including residential location. A web-based mapping application, OnTheMap, provides an easy-to-use interface for mapping small-area workforce characteristics. The application also provides tabulations to accompany the workforce maps on employer and worker characteristics, and allows users to create analysis of custom geographies. For researchers interested in entrepreneurship, a key feature of interest is highly detailed block-level data of employment in new firms. For example, **Figure 1** uses LODES data in OnTheMap to show the spatial concentration of new firms near the Stanford University campus in Palo Alto City, California.

3.2.3. Job-to-Job Flows (J2J)

Job-to-Job Flows (J2J) is a brand new data product from the Census Bureau on the flows of workers between employers, with data first released in December of 2014. Job-to-Job Flows is the first public use data product that exploits the ability of the linked employer-employee data to follow workers across firms, across industries, and across labor markets.

[5] Job creation and destruction for young firms and establishments can also be analyzed with the BDS and the BED.

The J2J data should prove particularly valuable to researchers of entrepreneurship. First, the potential to study start-up teams as groups of workers moving from their previous employers to the newly established firm is unique to linked employer-employee data. While there is no information about each individual's role or title in the company, strategies have been employed to identify founders, see Agarwal et al. (2013) using LEHD microdata. A second unique feature of the data is its ability to provide a dynamic view of the workforce in the early years of a business, permitting examination of the role that gender, age, industry experience, and experience working at other new businesses plays in the success or failure of new firms. Finally, the ability to identify co-workers and network effects from working in new technologies may also be interesting to researchers studying agglomeration economies and their role in forming industrial clusters.

As of this writing, the J2J data is beta, with more detailed tabulations planned for later releases. A full description of the methodology used for deriving the worker flow estimates from the LEHD data is available in Hyatt et al. (2014).

4. Some Examples of Analysis Using the Quarterly Workforce Indicators and Job-to-Job Flows

In this section, we provide some specific examples of how the public use QWI and J2J data can be used to answer questions of interest to researchers studying entrepreneurship.

4.1 Who Works at Start-ups?

We begin by presenting simple descriptive statistics from the QWI on the population of workers employed at start-ups. **Table 4** compares the workforce composition of start-ups to that of more established businesses, where start-ups are defined as businesses of age 0-1 years, and established businesses are grouped into two age categories, 2-10 years old and older than 10 years.

Comparing the percentages across the columns in **Table 4**, we see that start-ups disproportionately employ more young workers, with workers aged 14-24 representing 20.2% of the workforce at start-ups (versus 14.5% overall). Employment at younger firms also skews female (51.0%) and less educated. Young firms are also more likely to employ Asian and

Hispanic workers. Obviously, some of the differences in demographics across young and old firms are driven by industry composition (e.g., in leisure and hospitality firms, are overrepresented among young firms). These same statistics are available within detailed industries, so users can measure how the demographics of new firms in an industry compare to more established firms.

4.2 Did Changing Demographics Contribute to the Decline in Start-ups?

Next, we use the QWI to explore whether the composition of firms or the workforce can account for changes in certain economic indicators that we care about. Specifically, we turn to the important question of what has caused the documented decline in the employment at start-ups.[6] We begin the analysis in the year 2000, after which the employment share of start-ups began to decline and the earnings paid by new firms eroded.[7] We consider the share of employment at start-ups, the trend in the earnings differential between start-ups and established firms, as well as measures of employment reallocation: job creation, job destruction, hires, and separations.

We begin by describing the trends over time, although the decompositions that follow will only pertain to the endpoints of the trends plotted in these figures, which span from 2000Q2 to 2012Q2. **Figure 2** presents the trends in employment and earnings for two age categories: "start-up" firms, those aged 0-1, and all other firms, i.e., those aged 2 or older. **Figure 2a** shows that the employment share at young firms has declined throughout the 2000s, consistent with the evidence in the literature referenced above. The earnings series in **Figure 2b** shows divergent trends for young and old firms. Consistent with the evidence first documented by Brown and Medoff (2003), earnings at young firms are lower than earnings at older firms. The average

[6] This topic is discussed in a number of recent papers including Haltiwanger, Jarmin, and Miranda (2012), Hyatt and Spletzer (2013), Decker (2014), Decker et al. (2014a,b), Davis and Haltiwanger (2014), Pugsley and Sahin (2014), and Dinlersoz et al. (2015).

[7] 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. 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. (2015).

earnings of workers at the youngest firms have declined in real terms throughout the 2000s, but the earnings at older businesses have shown a modest increase, consistent with what is shown by Haltiwanger et al. (2012) and Dinlersoz et al. (2015).

Information on the composition of the workforce by firm age can be used to answer questions related to the decline of start-ups 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 start-ups (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:

(4.2.1) $\Delta Y_t = \Sigma_i \Delta Y_{it} S_{i\bullet} + \Sigma_i Y_{i\bullet} \Delta S_{it}$,

where $Yi\cdot$ denotes the mean such that $Y_{i\bullet} = (Y_{it} + Y_{it-1})/2$, and likewise $S_{i\bullet}$. 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 5** contains the results of this 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 start-ups, as in Ouimet and Zarutskie (2014) and Acemoglu, Akcigit, and Celik (2014), or have fewer resources to wait until a higher wage offer from an older firm as in Dinlersoz et al. (2015). However, as shown in **Table 5**, most of the changes in composition should have increased the share of start-ups, 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 start-ups. The increase in the share of

older workers, and their tendency to work at established businesses, explains 9.4% of the decrease in the share of employment at start-ups.

Figure 2b shows the average real earnings for workers who worked the entire quarter at start-ups 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 start-ups are falling. In the second column of **Table 5**, we decompose the rising earnings 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:

$$\frac{\sum \Delta Share_{Old,x} * \overline{Earn}_x - \sum \Delta Share_{Young,,x} * \overline{Earn}_x}{\Delta \overline{Earn}_{Old} - \Delta \overline{Earn}_{Young}}. \quad (4.2.2)$$

This provides a measure of how the change in a share for a subset of the population defined by a characteristic (x), as well as in the average earnings for that particular characteristic, is related to the change in earnings for young vs. old firms. Unlike our results for employment shares at start-ups, changes in industry composition and worker demographics explain a considerable part of the apparent increased earnings 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 start-ups. Workers at established firms are also trending older and more educated, relative to younger firms, although as these effects are measured ignoring the change in the industry distribution, they may be related and thus their effects are not necessarily additive.

In turn, **Table 6** shows how the change in the composition of employment by firm age explains the decline in four employment 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 computed according to equation 4.2.2 above. Results show that the shift away from entrepreneurship explains a substantial portion in the decline of such dynamics, due to the fact that start-ups are more volatile in terms of employment dynamics. The table shows that the decline in start-ups

explains 9.3% of the decline in hires and 6.8% of the decline in separations.[8] These results are similar to what Hyatt and Spletzer (2013) found using the LEHD microdata.

The above examples show how the demographic and industrial detail of the QWI can be used to study the composition of start-up employment, and its effects on economic dynamics. However, note that these exercises only scratch the surface of what can be learned from these statistics. All of the measures used here can be cross-tabulated on multiple levels, and are also available at narrow geographic detail, allowing for much more complex analyses.

4.3 Where do Early Employees Come From?

The new Job-to-Job Flows (J2J) data allow us to identify movements of workers into start-up firms from other employers. **Figure 3** shows a comparison of worker flows across three classes of employers: young firms (less than two years), established firms (more than 11 years), and small firms of all ages (less than 20 employees). Employment growth in each employer class is the sum of net employment flows (i.e. hires of nonemployed workers minus separations to nonemployment) and new worker reallocation (i.e. hires of workers away from other firms minus separations employees to other firms). This decomposition allows us see how firms grow, by poaching workers away from other firms or through employment flows.

Figure 3a depicts the hire and separation rates at start-up firms from 2000-2013. As can be seen in the figure, new firms obtain a significant share of their early employment growth by poaching workers away from more established firms. Flows into new firms from established firms are much higher than separations from new firms to more established employers. Poaching hires were highest during the 2000-2002 period, when half of new firm hires were of workers moving from other jobs. Overall, this decomposition shows the importance of worker moves from more established firms as a critical input to early firm growth.

As a comparison, **Figure 3b** shows this decomposition for established firms. In contrast to start-ups, net employment growth at established firms is much smaller, and occurs exclusively via employment flows. We find in other analysis (not shown) that the high contribution of job-to-

[8] 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 Decker et al. (2014b) found using the BDS.

job flows to employment growth at young firms disappears by the time firms are 2-3 years old. It may be that the high growth rate of the youngest firms from worker reallocation is driven by start-up teams transitioning from their previous jobs at older firms to the new firm.

As an additional comparison, we show the flows at businesses (of all ages) with fewer than 20 employees in **Figure 3c**. This decomposition for small businesses looks more like that for older established firms than for younger firms. Net worker reallocation to small firms from larger firms is low, although very slightly positive.[9] Haltiwanger, Jarmin, and Miranda (2013) finds that controlling for age, it is young firms rather than small firms that disproportionately drive job creation. Here we find that a pattern of employment growth through worker relocation (workers voting with their feet) characterizes new firms but not small firms generally. That workers are willing to move from established (and presumably more stable and higher-paying) employers to start-ups suggests that for early employees, working at a new firm offers opportunities for advancement and career growth not available to them at more established firms.

At press time, the J2J data are quite new, and do not yet provide as many tabulation levels as the QWI. The possibilities for analysis will only expand as the J2J statistics release more detailed tabulations.

5 Looking Forward: The Potential for New Data on Entrepreneurship

While substantial progress has been made in the last few years making linked employer-employee data more useful and accessible for entrepreneurship research, the work we have described so far represents only a fraction of possible ways to expand the frontier of data available for research. In particular, linking in additional data on business owners and creating new data on the dynamics of entrepreneurship would be an important advance in the statistical infrastructure to study new business formation. In this section, we discuss the potential for more information on entrepreneurs and their firms from linked employer-employee data and discuss some results from work to date on integrating new sources of data.

5.1 Linking Data on Business Owners

[9] Haltiwanger, Hyatt, and McEntarfer (2015) point out that the fact that worker relocation does *not* in fact redistribute workers away from small firms to large firms is inconsistent with a number of important labor market models, particularly Burdett and Mortensen (1998).

Efforts are currently underway to enhance the set of available data on business owners and the self-employed by integrating data on sole-proprietors and partnerships into the LEHD data infrastructure. A prototype microdata file is being created which covers the universe of active U.S. sole-proprietorships and partnerships, both with and without employees, from 2002 through 2012. The Census Bureau is undertaking research into using these data for new public use statistics on the dynamics of business ownership.[10]

The universe of this dataset encompasses all unincorporated businesses owned and run by one or more individuals. The data that we integrate originate primarily from individual federal income tax returns, such as income filings from Schedules C and K1, 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 includes owners of sole proprietorships, partnerships, and Subchapter S corporations. Owners of Limited Liability Companies (LLCs) and the like are included as long as they do not elect to be taxed by the IRS as a corporation. The individual business owners can then by linked via a personal identifier to the LEHD job-level database, thus providing an employment history for each owner. More details on how the data are constructed are provided in Garcia-Perez et al. (2013).

This linking of information on business ownership and employment status joins information in a way that is not available in other data sources, permitting a unique view of the path to entrepreneurship. Individuals starting businesses bring with them a pre-existing stock of human capital, through their past experience both in the labor market and also as prior business owners. The potential statistics derived from this unique data source will allow researchers to study the intersection of these two employment spheres, which has been little explored up to this point.

One challenge in the study of entrepreneurship is the lack of a cleanly defined measure of entrepreneurial activity. Measurement aside, there is in fact no consistent definition in the literature of what entrepreneurship is. At its narrowest, entrepreneurs have been identified as the founders of innovative new businesses that grow rapidly in both employment and output and thus

[10] This builds on previous work integrating the employer and nonemployer business data, see Davis et al. (2009).

drive national measures of economic growth. More broadly, the word entrepreneur has at its root "one who starts" and thus can refer to the founder of any business regardless of size or outcome.

More broadly still, entrepreneurial activity is associated with business ownership of any kind (with or without employees) and with self-employment, which is in turn equally hard to define. In fact, for tax purposes in the U.S., independent contractors are defined as self-employed and their earnings treated as self-employment earnings.

Taken independently, each of these varied concepts of entrepreneurial activity has value and each measure reveals a different facet of the economy. Rises and falls among innovative, high growth businesses have obvious implications for national employment and output. The set of all business starts with or without employees tells us, at a minimum, about the economy's capacity to support such efforts. The set of small self-owned businesses without employees combined with the pool of contract or contingent workers serves as an alternative measure of employment in a changing economy. This count may also measure what the development literature calls the informal labor market.

To better understand the implications of a rise or fall these varied measures of entrepreneurial activity, we must recognize that each of these events, the start of a new business (with or without employees) or the transition to contingent work, reflects a choice made by the owner. These choices are in turn influenced by the owner's personal pre-entry economic environment. In addition, trends in the varying concepts of entrepreneurship likely are inter-related. For example, ownership of a business without employees in many cases precedes the "birth" of an employer business. Thus, our ability to extract information from these trends is greatly enhanced by placing them in a broader context.

The linked employer-employee data constructed by the LEHD program have the potential to provide this context. Specifically, statistics released from these data may improve our understanding of entrepreneurial dynamics in three ways. First, as noted, it is the use of federal tax filings by sole proprietorships, partnerships and sub-chapter S corporations that gives the LEHD program its ability to identify business owners. Knowledge of the type of originating tax form combined with the presence or absence of employees allows us to disentangle these varied types of entrepreneurship and to separately examine trends in each. Second, by combining

administrative data on the universe of individual 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. Third, we can follow individuals as they transition between ownership of businesses without employees, employer businesses, and traditional work, and explore the interconnection between these spheres. In short, by identifying differing types of business ownership and by integrating each with employment and earning history and prior ownership experience for the owner, the program has the potential to release a set of statistics that gives insight into what each of these measures may be telling us about the vitality of the economy.

We will first describe the type of statistics the program has the ability to create to measure and explore conventional self-employment as well as self-employment as an alternate form of employment (what the literature has termed the "gig economy"). We follow with a more developed discussion of how linked employer-employee-owner data may further our knowledge of entrepreneurship by tracking the events that precede and follow the birth of a business.

5.2 Self-employment and the "Gig Economy"

The vast majority of businesses that report earnings have no employees. While self-employment counts have stagnated in survey reports in recent years, the count of these nonemployer sole proprietor businesses have continued to rise.[11] This count includes any person who receives income as a statutory employee or contingent worker or who operates a business or practice for profit with regularity and continuity.[12] Internet businesses, freelancers, contract workers, consultants, etc, all are included in this measure.

[11] In a recent interview, Laurence Katz described preliminary work with Alan Krueger to investigate the discrepancy between steady trends in self-employment in survey data and increases in self-employment suggested by tax data. Rob Wile, "There are probably way more people in the 'gig economy' than we realize." July 27, 2015, Fusion.net.

[12] Data on non-employer sole proprietors originate from filings of IRS 1949 Schedule C. The Schedule C instructions state "use Schedule C (Form 1040) to report income or loss from a business you operated or a profession you practiced as a sole proprietor. An activity qualifies as a business if your primary purpose for engaging in the activity is for income or profit and you are involved in the activity with continuity and regularity. For example, a sporadic activity or a hobby does not qualify as a business. Also use Schedule C to report (a) wages and expenses you had as a statutory employee, (b) income and deductions of certain qualified joint ventures, and (c) certain income shown on Form 1099-MISC, Miscellaneous Income."

The rise in employment arrangements of this type is linked in part to technology which has significantly lowered the entry cost for these businesses. The U.S. economy has become much more service oriented and thus the capital requirements associated with business entry are low. The pros and cons of this trend have been widely discussed and can be viewed from the perspective of the employer, the worker, or the economy as a whole. From an employer's perspective, the availability of an on-demand workforce lowers labor costs and provides flexibility. From the worker's perspective, a less formal work arrangement often precludes other benefits of employment such as stability and health insurance coverage yet does provide an alternative to conventional work when faced with unemployment or under-employment. For the economy as a whole, a rise in unemployment is one of the mechanisms through which the economy is theorized to self-correct during recessions. Thus, unlike a rise in conventional entrepreneurship which is viewed as a driving force of economic growth, it is not clear whether we should regard the rise in the numbers of nonemployer sole proprietors as a sign of economic strength.

Linked employer-employee-owner data have the potential to create statistics that provide more insight into these trends. For each new nonemployer, we observe their employment and earnings status in time periods preceding self-employment entry. The data thus give us some ability to separately identify those new nonemployers pushed into self-employment by lack of economic opportunity from those lured into self-employment by higher anticipated returns. We can identify those entrants with no wage and salary earnings, those with broken spells of employment, those previously working at a downsizing employer or those employed but earning significantly less than comparable workers. Similarly, we can identify those entrants with high, above average or rising wage and salary earnings. An understanding of the forces that may influence self-employment entry may help economists understand the nature of a rise of business ownership of this type.

5.3 Measuring Business Ownership Dynamics

The determinants of entrepreneurial success are a much studied topic, but many of these factors are determined prior to the beginning of a business. The human capital and prior experience that an entrepreneur brings to their new venture are clearly important, and may not be possible to fully encapsulate in measures such as education level. Moreover, many business

starts and business failures occur before the firm 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. In order to identify the characteristics of successful entrepreneurs, and to answer questions like why the rate of entrepreneurship is declining, it may be important to observe these potential job creators at their earliest stages.

Such a link should prove enlightening in the context of the well-documented decline in U.S. start-ups, which has sparked much interest in the underlying causes and implications of this slowdown. Although the overall trend in start-ups 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 start-ups requires knowledge of the factors that precede a business and an understanding of how these factors influence the odds of a successful start-up. 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. Statistics derived from linked sole-proprietor and LEHD data will offer a way to help parse such differences in the paths of potential entrepreneurs.

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

Researchers are interested in identifying successful transitions to entrepreneurship. One measure of success is the owner's ability to create a primary source of earnings for themselves from the business. The combined owner-work history data are well suited to explore the following question: what share of self-employed businesses grow enough to allow the owner to leave wage and salary employment?

The left-hand panel of **Table 7** 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. One of the first facts to stand out is that the majority of self-employed businesses without employees do not in fact grow large 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

businesses with employees 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 7**, 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 nonetheless 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.

This example clearly shows that there is no single path to entrepreneurship, as the relationship between wage and salary work, self-employment, and running an employer business is quite complicated. These data are uniquely suited to studying the interplay between these types of employment, and the future business owner statistics should enable new exploration into the origins of entrepreneurship.

6 Conclusion

Linked employer-employee data has enormous potential for empirical research in entrepreneurship. These data allow an ever-growing community of researchers to develop a clearer picture of how new firms come into being, obtain workers, grow, shrink, and exit, and how this dynamic process is related to employment and economic growth. In this chapter, we described the LEHD linked employer-employee microdata, public use data on start-ups tabulated from LEHD data, and highlight how they fill gaps in the set of available data for the study of entrepreneurship. We provided examples that illustrate the power of the new public data to address questions that previously required access to restricted microdata. Work to expand the

utility of this data for entrepreneurship research is still ongoing; we also outlined future plans for development of new data products for empirical research on entrepreneurship.

References

Abowd, J., B. Stephens, L. Vilhuber, F. Andersson, K. McKinney, M. Roemer, and S. Woodcock. 2009 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*, 149–230. University of Chicago Press.

Acs, Z., and P. Mueller. 2008. Employment effects of business dynamics: Mice, gazelles and elephants, *Small Business Economics*, 30(1): 85-100.

Acemoglu, D., U. Akcigit, and M. Celik. , 2014 Young, restless, and creative: Openness to disruption and creative innovation. *NBER Working Paper* No. 19894.

Agarwal, R., B. Campbell, A. Franco, and M. Ganco. 2013. 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*.

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

Birch, D. 1979. *The job generation process*. Cambridge, MA: MIT Program on Neighborhood and Regional Change.

Brown, C., and J. Medoff. 2003. Firm age and wages. *Journal of Labor Economics* 21(3): 677-698.

Burdett, K., and D. Mortensen. 1998. Wage differentials, employer size, and unemployment. *International Economic Review* 39(2): 257-273.

Carree, M. 2002. Does unemployment affect the number of establishments? A regional analysis for US states. *Regional Studies* 36(4): 389-398.

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

Combes, P., G. Duranton, and L. Gobillon. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2): 723-742.

Congregado, E., E. Carmona, A. Golpe. 2010. Co-movement and causality between self-employment, unemployment and the business cycle in the EU-12. *International Review of Entrepreneurship* 8(4): 303-336.

Davis, S., J. Haltiwanger, C.J. Krizan, J. Miranda, A. Nucci, and L.K. Sandusky. 2009. Measuring the dynamics of young and small businesses: Integrating the employer and nonemployer universes. In T. Dunne, J. Jensen, and M. Roberts, editors, *Producer Dynamics: New Evidence from Micro Data*, 329-366. University of Chicago Press.

Davis, S., and J. Haltiwanger. 2014. Labor market fluidity and economic performance. *NBER Working Paper* No. 20479.

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

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

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

Dinlersoz, E., H. Hyatt, and H. Janicki. 2015. Who works for whom? Worker sorting in a model of entrepreneurship with heterogeneous labor markets. *Center for Economic Studies Working Paper* No. 15-08.

Doms, M., E. Lewis, and A. Robb. 2010. Local labor force education, new business characteristics, and firm performance. *Journal of Urban Economics* 67(1): 61-77.

Dunne, T., M. Roberts, and L. Samuelson. 1989. Plant turnover and gross employment flows in the U.S. manufacturing sector. *Journal of Labor Economics* 7(1): 48-71.

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

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

Fairlie, R. 2005. Self-employment, entrepreneurship, and the NLSY79. *Monthly Labor Review* 128(2): 40–47.

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

Fort, T., J. Haltiwanger, R. Jarmin, and J. Miranda. 2013. How firms respond to business cycles: The role of firm age and firm size. *NBER Working Paper* No. 19134.

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

Garcia-Perez, M., C. Goetz, J. Haltiwanger, and L.K. Sandusky. 2013. 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).

Gertler, M., and S. Gilchrist. 1994. Monetary policy, business cycles and the behavior of small manufacturing firms. *Quarterly Journal of Economics* 109(2): 309-340.

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

Haltiwanger, J., H. Hyatt, and E. McEntarfer. 2015. Cyclical reallocation of workers across employers by firm size and firm wage. *NBER Working Paper* No. 21235.

Haltiwanger, J., H. Hyatt, E. McEntarfer, and L. Sousa. 2012. Business Dynamics Statistics briefing: Job creation, worker churning, and wages at young businesses. *Kauffman Foundation Paper Series*.

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

Haltiwanger, J., R. Jarmin, and J. Miranda. 2013. Who creates jobs? Small vs. large vs. young. *Review of Economics and Statistics* 95(2): 347-361.

Haltiwanger, J., H. Hyatt, E. McEntarfer, L. Sousa, and S. Tibbets. 2014. Firm age and size in the Longitudinal Employer-Household Dynamics data. *Center for Economic Studies Discussion Paper* No. 14-16.

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

Holtz-Eakin, D., D. Joulfaian, and H. Rosen. 1994. Sticking it out: Entrepreneurial survival and liquidity constraints, *Journal of Political Economy* 102(1): 53-75.

Hurst, E., and A. Lusardi. 2004. Liquidity constraints, household wealth and entrepreneurship", *Journal of Political Economy* 112(2), 319-347.

Hurst, E., and B. Pugsley. 2011. What do small businesses do? *Brookings Papers on Economic Activity* 43(2): 73-142.

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

Hyatt, H., E. McEntarfer, K. McKinney, S. Tibbets, and D. Walton. 2014. Job-to-job (J2J) flows: New labor market statistics from linked employer-employee data. *JSM Proceedings 2014*, Business and Economics Statistics Section, 231-245.

Kerr, W., and S. Kominers. 2010. Agglomerative forces and cluster shapes. *NBER Working Paper* No. 16639.

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

Lazear, E. 2005. Entrepreneurship. *Journal of Labor Economics* 23(4): 649-80.

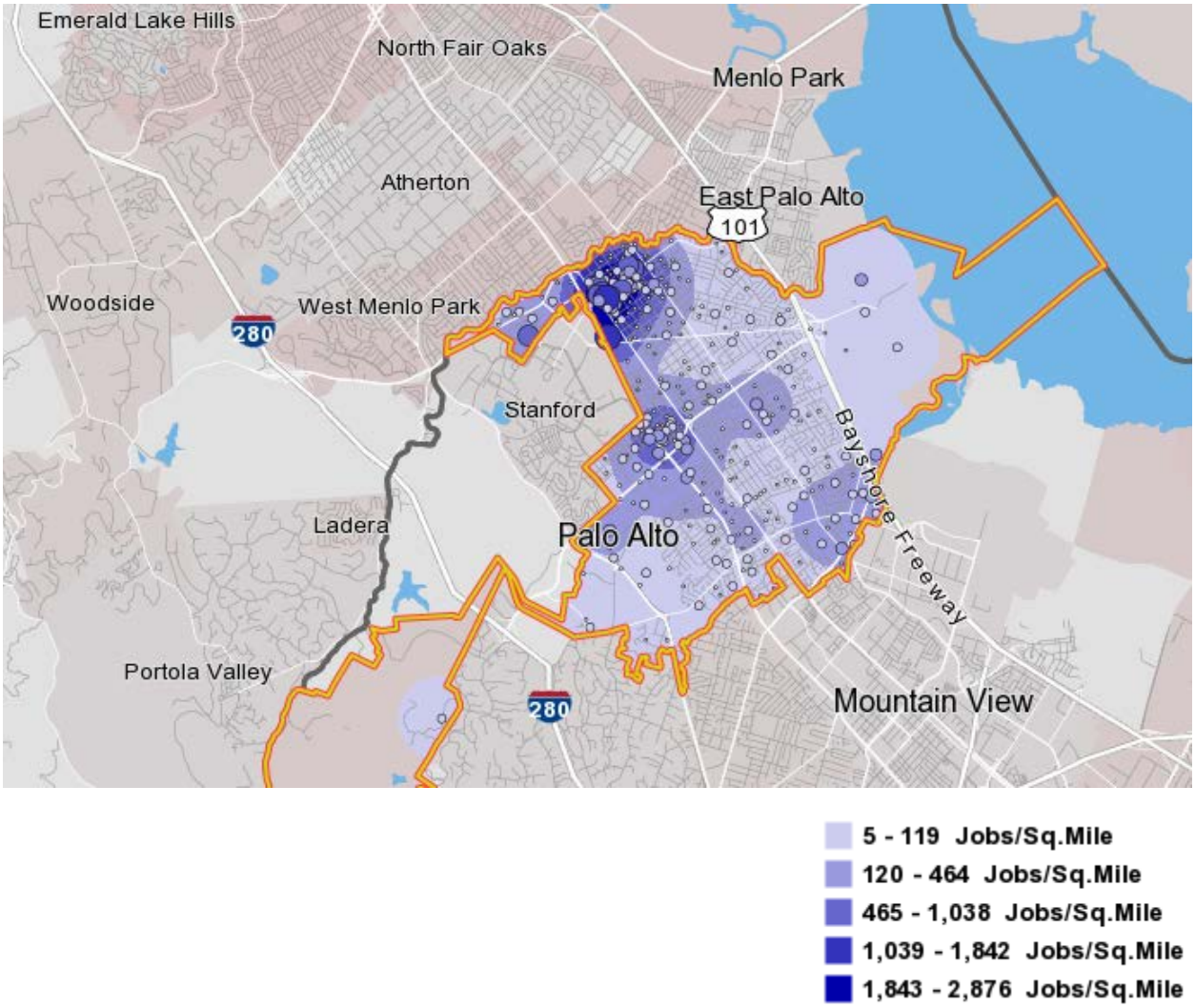
Moskowitz, T., and A Vissing-Jørgensen. 2002. The returns to entrepreneurial investment: A private equity premium puzzle? *American Economic Review* 92(4): 745-778.

Ouimet, P., and R. Zarutskie. 2014. Who works for startups? The relation between firm age, employee age, and growth. *Journal of Financial Economics* 112(3): 386-407.

Pugsley, B., and A. Sahin. 2014. Grown-up business cycles. *Federal Reserve Bank of New York Staff Report* No. 707.

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

Figure 1: Concentration of Start-up Employment near Stanford University and Palo Alto, CA



Notes: LEHD Origin-Destination Employment Statistics (LODES), 2013. Only employment in firms less than two years old is shown in map.

Figure 2a: Employment Shares by Firm Age

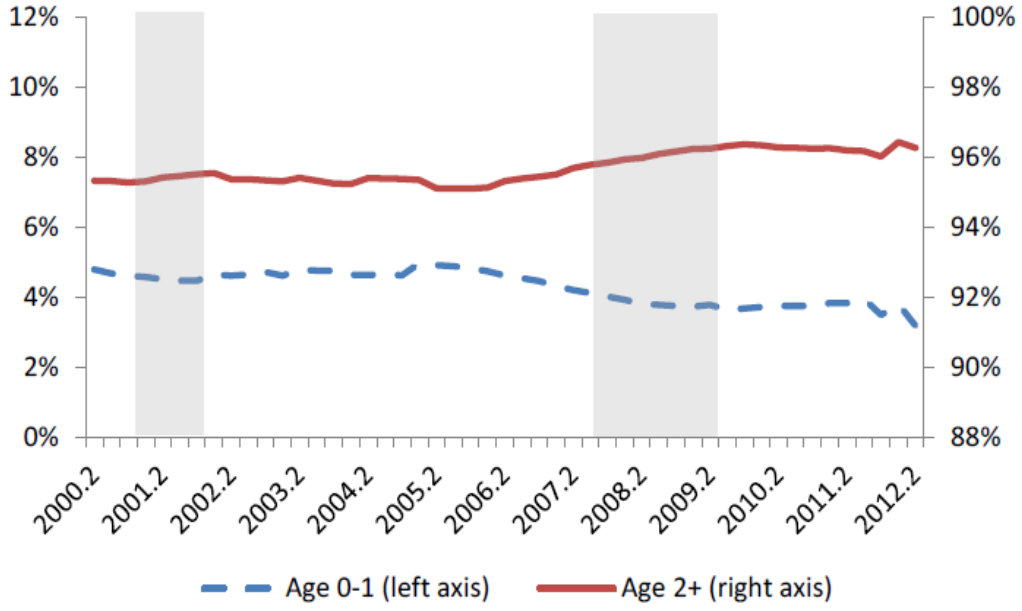
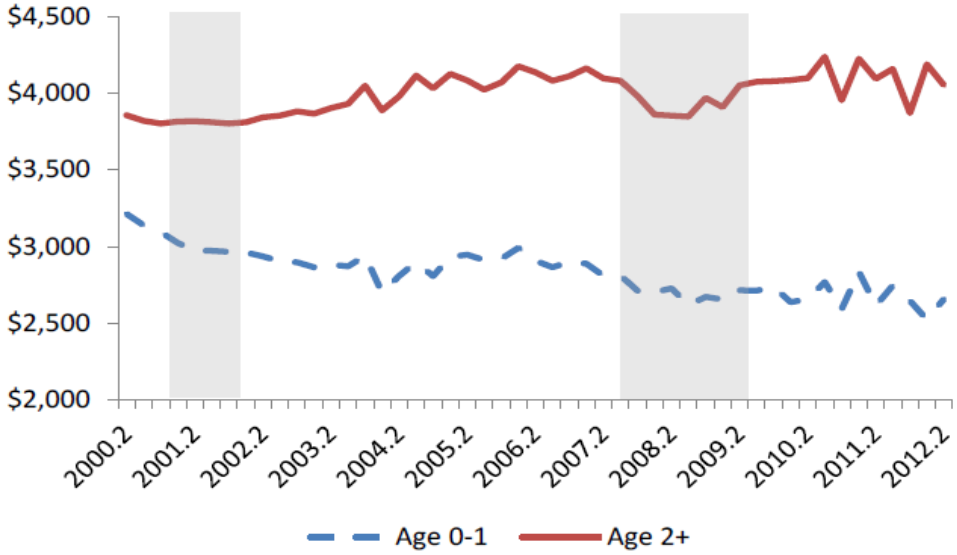


Figure 2b: Real Quarterly Earnings by Firm Age



Notes: Authors' calculation of the Quarterly Workforce Indicators. All data are seasonally adjusted.

Figure 3a: Hires and Separations at Young Firms (0-1 year old) 2000-2013

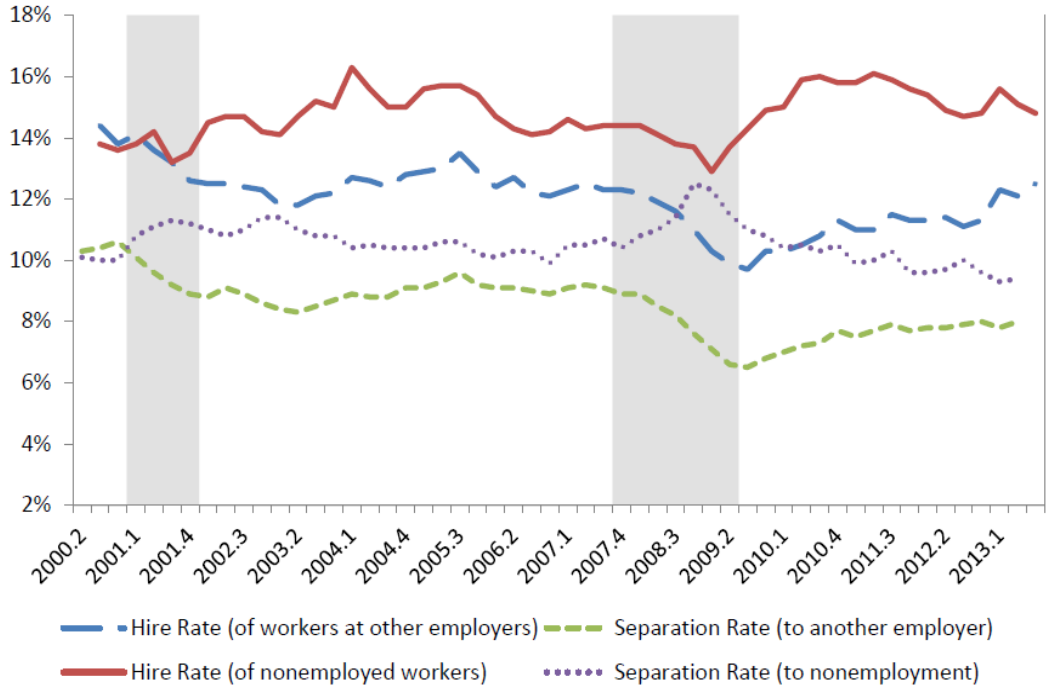


Figure 3b: Hires and Separations at Established Firms (11+ years old) 2000-2013

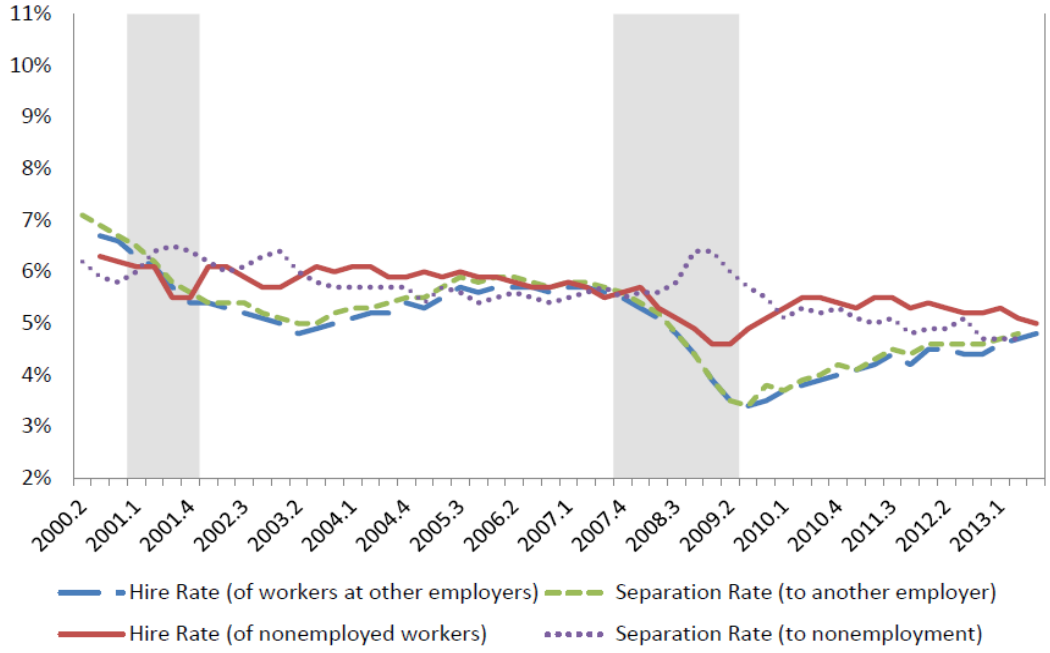
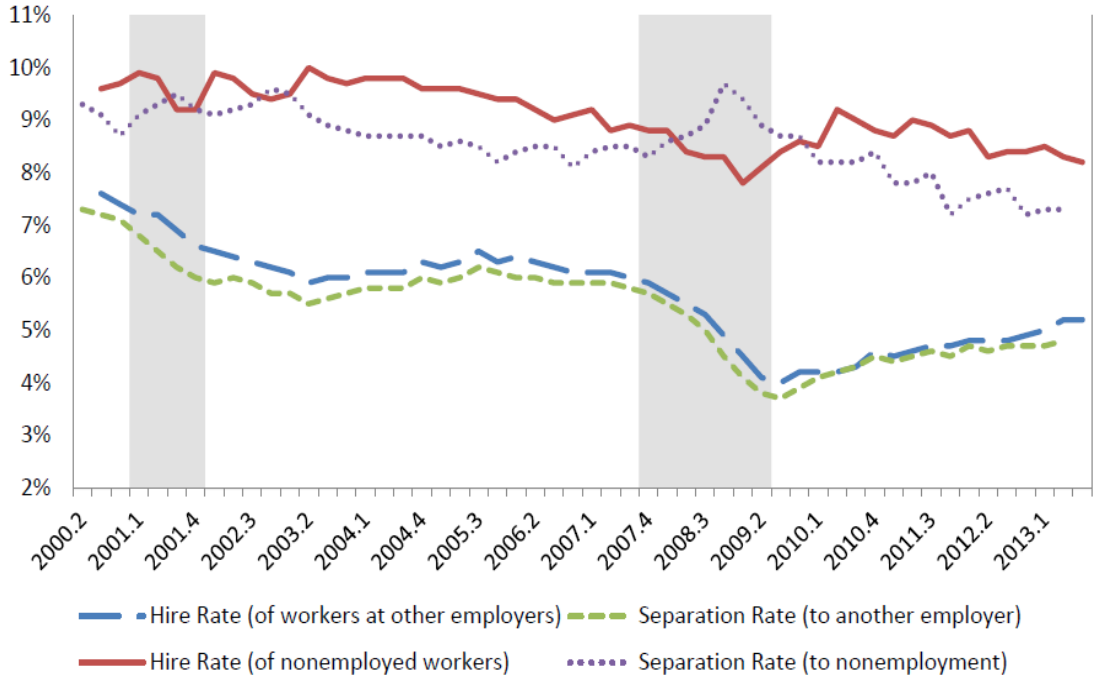


Figure 3c: Hires and Separations at Small Firms (<20 Employees) 2000-2013



Notes: Authors' calculations from national Job-to-Job Flows data, beta 2014Q1 release. All data are seasonally adjusted.

Table 1: Public use data to study firm dynamics and entrepreneurship

Dataset(s)	Sampling Unit or Frame	Key Variables	Frequency	Level of Detail	Strengths
Business Dynamics Statistics (BDS)	Establishment	Employment, job creation and destruction by firm age and size.	Annual: 1978-current. Two year lag.	Industry sector (SIC), National, state, and MSA.	Long time series on employment, job creation and destruction trends for young firms.
Business Employment Dynamics (BED)	Establishment	Job gains from new and expanding establishments and jobs lost from downsizing and closing establishments.	Quarterly: 1992-current. Nine month lag.	National and state by NAICS sector; 3-digit NAICS available nationally. Firm Age categories at state-level, firm size at national-level.	Quarterly frequency and relatively current data on start-ups and new establishments.
Quarterly Workforce Indicators (QWI)	Job (worker-establishment pair)	Employment, job creation and destruction, hires and separations, earnings and starting earnings by firm age or size.	Quarterly: 1990 (start year varies by state) – current. Nine month lag.	National,, state, CBSA, and county level data. Industry detail up to 4-digit NAICS. Worker age, sex, education, race/ethnicity.	Provides worker demographics, earnings, and turnover as well as job creation and destruction at young firms. Available at very detailed geography and industry. High frequency and relatively current.
Kauffman Firm Survey (KFS) Microdata, Dun & Bradstreet Market Identifier Data (D&B)	KFS: New firms in 2004 D&B: Around 50 million establishments since 1990	Business characteristics, with info on strategy, credit and financing. Kauffman includes demographics of the principals.	Annual. Kauffman survey stopped in 2011.	Firm or establishment level. Confidential version of KFS contains more industry and geographic detail.	Wealth of information on the firm-level, although samples are not representative of universe
Household surveys: Current Population Survey (CPS), National Longitudinal Surveys (NLS/NLSY), Survey of Income and Program Participation (SIPP), Panel Study of Income Dynamics (PSID), Survey of Consumer Finance (SCF)	Household	Detailed job and earnings histories of potential entrepreneurs, self-employment entry and exit.	Varies	Individual level. Confidential and restricted versions with more detail often available through application process.	Wide variety of information on potential entrepreneurs although samples are often small.
Census Business Register Statistics: County/Zip-Code Business Patterns, Nonemployer Statistics, Statistics of U.S. Businesses (SUSB)	Establishment	Establishment counts, employment and payroll by establishment and enterprise size class	Annually since the late-90s	Statistics for industry sectors generally available at the county-level and above	Establishment counts of small businesses at fine levels of geography, and ability to distinguish nonemployers
Survey of Business Owners (SBO)/Characteristics of Business Owners (CBO)	Business Owner	Owner demographics, geography, industry, firm receipts and employment size, detailed information on financing and revenues	Every 5 years since 2007	SBO: National, state, and county by NAICS 2- through 6-digit industry for selected geographies. CBO: National by industry	Rich set of variables describing the individual owners and their business finances.

Table 2: Questions in entrepreneurship research

Question	Selected Empirical Papers	Selected Data Sets Used	Potential Value-Added
What are the dynamics of new business formation and growth?	Birch (1979), Dunne, Roberts, and Samuelson (1989), Acs and Mueller (2008), Davis and Haltiwanger (2014)	Dun and Bradstreet microdata, Census of Manufactures microdata. Longitudinal Estab. and Ent. Microdata (LEEM), Quarterly Workforce Indicators	QWI and J2J: Ability to observe labor dynamics at firms 0-1 years old, stratified by a variety of observable characteristics
How does entrepreneurship interact with the business cycle?	Gertler and Gilchrist (1994), Carree (2002), Congregado et. al. (2010), Fort et. al. (2013)	Quarterly Financial Report--Manufacturing, County Business Patterns, Current Population Survey, Business Dynamics Statistics	QWI and J2J: Time-series measures of hiring, separations, and poaching at young firms vs. established firms
How does entrepreneurship depend on the available labor force?	Combes, Duranton, and Gobillon (2008), Doms, Lewis, and Robb (2011), Ouimet and Zarutski (2012), Figueiredo et. al. (2014)	French microdata, Kauffman Firm Survey, Decennial Census, LEHD microdata, Portuguese Administrative microdata	QWI and J2J: Observe demographics of the labor force at young firms, such as age, sex, race, and education
How and why are geographic and industrial clusters formed?	Ellison and Glaeser (1997), Rosenthal and Strange (2003), Glaeser, Kerr, and Ponzetto (2010), Kerr and Kominers (2010)	Census of Manufactures, Dun and Bradstreet Marketing Indicators, Longitudinal Business Database, U.S. Patent Office microdata	QWI and J2J: Statistics available at fine levels of geography and industry detail
How are spinoffs created?	Klepper (2001), Franco and Filson (2006), Chatterjee and Rossi-Hansberg (2012), Agarwal et al. (2013)	Business Employment Dynamics, Disk/Trend Industry DataStatistics of U.S. Businesses, LEHD Microdata	J2J:Ability to detect flows within an industry and geographic location
Where do entrepreneurs come from?	Evans and Leighton (1989), Hurst and Lusardi (2004), Lazear (2005), Hurst and Pugsley (2011)	Current Population Survey, NLSY, Panel Study of Income Dynamics, Data Set of Stanford Alumni, Statistics of U.S. Businesses	Future Sole-Prop Statistics: Info on previous employment status and industry experience of business owners
How do entrepreneurs fare in their outcomes?	Bates (1990), Holtz-Eakin, Joulfaian, and Rosen (1994), Hamilton (2000), Moskowitz and Vissing-Jørgensen (2002)	Characteristics of Business Owners, Internal Revenue Service microdata, Survey of Income and Program Participation, Survey of Consumer Finances	Future Sole-Prop Statistics: Ability to measure earnings and year-to-year survival of sole-proprietor businesses

Table 3: Newly available data on firm dynamics and entrepreneurship from LEHD

Variable	Available in	Frequency	Most granular geographic detail	Most granular industry detail	Worker demographics	Fills a gap in public use statistics by allowing researchers to	Also available in (most granular level of detail)
Employment by firm age	QWI	Quarterly	County	NAICS4	Age, Sex, Education, Race/Ethnicity	Examine demographics of workers at young firms and within detailed industries. Map detailed sub-state geographic industry clusters..	BDS (MSA-Year-Industry sector)
Employment by firm age	LODES	Annual	Census block	All industries	All demographic groups	Map clusters of young firms at very detailed geographies	BDS (MSA-Year-Industry sector)
Hires/separations/ by firm age	QWI	Quarterly	County	NAICS4	Age, Sex, Education, Race/Ethnicity	Examine churn at young firms within detailed industries/geographies.	None
Earnings and starting earnings by firm age	QWI	Quarterly	County	NAICS4	Age, Sex, Education, Race/Ethnicity	Examine earnings at young firms by worker demographics.	None
Job-to-job moves by firm age	J2J	Quarterly	State	Industry sector	All demographic groups	Examine where early start-up employees are coming from and going to after separating.	None
Hires/separations to nonemployment by firm age	J2J	Quarterly	State	Industry sector	All demographic groups	Decompose worker churn at young firms into workers moving to and from other jobs vs. moving in and out of nonemployment.	None

Table 4: Demographics of the Workforce at Young versus Established Firms

	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%
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%

Notes: Source is authors' calculations from Census Quarterly Workforce Statistics (QWI), using private sector employment counts in 2013:Q3 for all U.S. states (except Massachusetts) and the District of Columbia.

Table 5: Employment Composition on Differences in Employment and Earnings, 2000Q2 vs. 2012Q2

	Employment	Start-up 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%

Notes: Authors' 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 6: Change in Employment Dynamics due to Decline in Start-ups: 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%

Notes: Authors' calculations from the Quarterly Workforce Indicators. See text for formulas.

Table 7: 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	40.5%	34.8%	29.9%	36.3%	16.7%	24.4%	86,011
All Employers	21.0%	19.9%	20.6%	17.6%	14.9%	22.1%	721,807
New Nonemployers	68.3%	65.4%	62.3%	0.0%	100.0%	51.7%	6,158,104
All Nonemployers	53.9%	50.7%	50.2%	65.6%	100.0%	68.8%	17,912,997

Notes: Table reports percentages of sole-proprietor business owners in 2009 of a given type that also have positive income from wage and salary work and/or nonemployer activity in the years 2008-2010. Sample consists of 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 the 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.