

FINANCING, OWNERSHIP, AND PERFORMANCE:  
A NOVEL, LONGITUDINAL FIRM-LEVEL DATABASE

J. David Brown, U.S. Census Bureau  
Steven J. Davis, Hoover Institution, SIEPR and NBER  
Lucia Foster, U.S. Census Bureau  
John Haltiwanger, University of Maryland and NBER  
John Sabelhaus, Brookings Institution and New Light Technologies

March 2024

ABSTRACT

The Census Bureau’s Longitudinal Business Database (LBD) underpins many, many studies of firm-level behavior. It covers all employers in the nonfarm private sector but lacks information about business financing and owner characteristics. To address this shortcoming, we link firm-level observations in the LBD to micro data drawn from several large Census Bureau surveys. The resulting Longitudinal Employer, Owner, and Financing (LEOF) database contains more than two million observations at the firm-year level with information about start-up financing, current financing, owner demographics, ownership structure, profitability, owner aspirations, and more – all linked to annual firm-level employment data since the firm hired its first employee. We use the LEOF database to document trends in owner demographics and financing patterns and to investigate how these business characteristics relate to firm-level employment growth.

Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau, nor the views of the Brookings Institution or its funders. The Census Bureau Disclosure Review Board has cleared the statistics in this presentation for public dissemination (Data Management System: P-7514232, CBDRB-FY22-CES014-050). We thank John Eltinge, Nathan Goldschlag, Cristina Tello-Trillo, and participants at the AEA, FSRDC, and OUSEA-BEA-Census conferences for helpful comments on this paper or on early draft presentations of this work.

## **I. Introduction**

The Census Bureau's Longitudinal Business Database (LBD) has become a major empirical laboratory for the study of firm dynamics and performance. The LBD covers all employers in the nonfarm private sector back to 1976 but lacks information about business financing and owner characteristics. In this paper, we integrate rich micro data on business financing, owner characteristics, ownership structure, and more with firm-level data in LBD. Specifically, we draw on micro data from the Census Bureau's Annual Business Survey (ABS) in 2017 and its predecessors, the Annual Survey of Entrepreneurs (ASE) and the Survey of Business Owners (SBO), using data back to 2002.

Our resulting Longitudinal Employer, Owner, and Financing (LEOF) database contains more than two million observations at the firm-year level with information about business start-up financing, current financing, owner demographics, ownership structure, profitability, owner aspirations, and more – all linked to annual firm-level employment data (back to 1976) since the firm hired its first employee. Our work builds on previous studies that use ASE and SBO data on firm financing and owner characteristics linked to LBD firm-level outcome measures such as Jarmin and Krizan (2010), Brown et al. (2017, 2019, 2022), Azoulay et al. (2020), and Lee et al. (2023). We contribute by providing a longer time frame and consistent measures of financing over time.

To illustrate the usefulness of the LEOF database, we use it to document trends in owner demographics and financing patterns and to investigate how these business characteristics relate to firm-level employment growth. Among other results, we find that the share of firms with older owners rose from 2002 to 2017, coinciding with an aging of the U.S. workforce. The share with no male owners also rose over this period, as did the share with no non-Hispanic White owners. On the financing side, we see a fall in the use of bank loans over this period and a fall in reliance on home equity. When focusing on high-growth firms, we find that their owners tend to be younger and more educated. High-growth firms are also more likely to be male owned. Firms with bank loans or outside investor funds at start-up are much more likely to become high-growth firms. Not surprisingly, firms with more start-up financing are more likely to be in the upper quartile of the employment growth distribution of their birth cohort.

To build the LEOF database, we start from the annual LBD universe files described in Chow et al. (2021). LEOF uses the firm and establishment identifiers enabling longitudinal tracking of firms as well as the employment inputs from the LBD firm files and key firm characteristics including firm age, industry, and geography.<sup>1</sup> LEOF employment and firm age variables are conceptually consistent with those used in the public-domain statistics produced as part of the Business Dynamics Statistics (BDS).

The LEOF owner and firm characteristics are taken from three related Census Bureau business surveys. At a high-level, one can think of these three surveys as sequential collections with the common overarching goal to capture firm-level information on businesses and business owners for the non-agricultural private economy. We use information from the SBO for reference years 2002, 2007, and 2012; the ASE for reference years 2014, 2015, and 2016; and the ABS for reference year 2017.<sup>2</sup> Although the surveys have enough in common to create a core set of consistent variables over time, they differ in population coverage, sample size, and scope. Since these can also differ within a survey across different waves, we essentially have seven different collections to harmonize.

There are multiple ways to use LEOF. In this paper, we analyze the repeated cross-sections with linked longitudinal employment outcomes. We stack the seven linked LEOF files and group firms by age then compare outcomes by firm age group over time using the seven surveys. We distinguish young (less than 60 months since first paid employee) and mature firms in each survey year. We show the importance of firm age in a series of exercises that examine growth rates by young and mature firms. We start by confirming that there is a strong up or out dynamic among young firms, and that the young-firm shares of employment and firm numbers

---

<sup>1</sup> In multi-establishment firms the geographic and industry identifiers for the firm are based on the largest establishment in that first year. Alternative approaches are also feasible with the LEOF data infrastructure as with the LBD (e.g., shares of activity for a firm by geography and industry).

<sup>2</sup> The ASE was conducted by the Census Bureau in partnership with the Ewing Marion Kauffman Foundation and the Minority Business Development Agency (see Foster and Norman (2017)). The ABS is conducted by the Census Bureau in partnership with the National Center for Science and Engineering Statistics. We use data from the first reference year of the ABS; rounds with subsequent reference years are available, but not all include financial information. The ABS with reference year 2017 is sometimes referred to by its collection year (2018), but for consistency with the other surveys' use of the reference year in their naming conventions, we refer to the ABS reference year in this paper.

have declined sharply over time. See, for example, Decker et al. (2014). We then extend earlier analyses to consider owner demographics and financing.

The next section describes how we construct LEOF (with validation in the appendix). Section III uses LEOF to show trends in owner and firm characteristics for 2002 to 2017. Section IV shows how LEOF helps us understand the demographic and financing characteristics of high-growth firms. Section V sketches other potential applications of the LEOF database and concludes.

## **II. Combining Firm-Level Administrative and Survey Data**

The LEOF dataset brings together administrative and survey firm-level inputs. The administrative input is an LBD spine file that has one record for every unique LBD firm between 1982 and 2021. The survey inputs are from seven waves of periodic Census surveys conducted between 2002 and 2018, under the SBO, ASE, and ABS data collection programs. The firm-level surveys have evolved over time in terms of questionnaire design, sample size, sampling strategy, mode, and response rates. Thus, creating consistent repeated survey cross-sections involves first finding a common set of owner demographic, firm characteristic, and financing variables, and then reweighting to approximate LBD distributions by age, size, and industry. We validate the representativeness of LEOF across survey waves by comparing LEOF- and LBD-based estimates of employment growth by firm age.

### *LBD Spine File*

The administrative input is an LBD spine file that has one record for every unique LBD firm between 1982 and 2021.<sup>3</sup> The LBD is comprised of two interconnected cross-section data files, one for establishments, and one for firms (Chow et al., 2021). Every establishment is connected to a parent in the firm file using the identifier `lbdfid`, which also lets us link firms across years.<sup>4</sup> The interconnections are important because LEOF uses longitudinal employment

---

<sup>3</sup> We start in 1982 since we have constructed the LEOF database so that we can track the dynamics of young firms (age 6 or less) and mature firms. This implies we are using the LBD to its origin year of 1976 to define firm age.

<sup>4</sup> Although the `lbdfid` reflects the state of the art in LBD processing, it does have some limitations as a longitudinal identifier connecting firms over time. The primary role of `lbdfid` is to serve as a consistent mapping of "firmid" to "lbdfid". The only "improvement" currently baked into `lbdfid` is the treatment of recycled firmids. That is, if a firmid

from the firm file, but other key characteristics such as geography, industry, and the EINs needed to link to the survey files are only available in the establishment records.

The number of employer firms in the U.S. has grown from about 4 million in 1980 to about 6 million in 2021. Firms are born, and then they either survive, die, or their associated establishments are sold to another firm—what we refer to here as a “dissolver” firm. These dynamic processes lead to roughly fifty million *unique* firms in our LBD spine file over the period 1982 to 2021. For each of these unique firms, the LBD firm file contains the birth year, death year if they have gone out of business, and total employment in every year the firm exists.<sup>5</sup>

Most firms in the LBD are single establishments, so the incremental information we take from the establishment files (geography and industry) is directly associated with a unique firm. For multi-establishment firms LEOF uses the location and industry of the largest establishment in the first year during which the firm had positive employment. This configuration for LEOF — every firm has longitudinal employment along with fixed birth year, death year, geography, and industry—is very simple but is only one way to construct the file. The advantage of this simplifying rule is that it keeps the file size manageable.<sup>6</sup>

The remaining LBD-related characteristic of LEOF firms involves the idea of firm “dissolution.” A dissolver firm is one where the firm identifier disappears between two years (there are no associated establishments) but at least one of the underlying establishments that was

---

comes back into the data after not being seen for 7 or more years, then LBD assigns a new lbdfid. Linking multi-unit firms over time comes with a number of important caveats that are not applicable at the level, which, in the LBD are tracked using the “lbdnums” identifier. An establishment, so long as it is the same type of business (industry) in the same physical location will always (conceptually) get the same lbdnum. However,, lbdfid may change for any number of reasons. Even for a single unit firm that re-organizes will have the same lbdnum over time but its firmid might change. Another important structural break in the longitudinal properties of lbdfid occurs with single-unit (SU) to multi-unit (MU) transitions, where the firmid (and thus the lbdfid) necessarily change for some mixture of “true” and administrative reasons. In some cases the SU--MU transition occurs along with a change in ownership (resulting in what we would think of as a “true” change in lbdfid), and in others it changes because of how treats MUs vs SUs. To mitigate the uncertainty associated with using lbdfid, our analysis of high-growth firms limits the sample to include “non-dissolver” firms, meaning all establishments associated with the firm have ceased to exist or at least one of the establishments in the base year continues to exist. For further discussion of lbdfid, see the data appendix to Sterk, Sedlacek, and Pugsley (2021).

<sup>5</sup> All firm level variables are the “BDS” versions, meaning our statistics are consistent with public use Business Dynamic Statistics (BDS) tabulations. See <https://www.census.gov/programs-surveys/bds.html>.

<sup>6</sup> One exception to the fixed establishment variable rule is Employer Identification Numbers (EINs). EINs are the key to linking survey records, but (1) some firms have multiple EINs, and (2) EINs evolve over time. Linking survey records involves looping over all establishments for a given firm in the base year to find the matching EIN.

part of the firm in the base year continues to have positive employment under another unique firm id. Dissolution can mean a lot of things, but a simple tabulation of employment using all firms (including dissolvers) would suggest the dissolvers were firms that went out of business. Dissolution is relative to a point in time—here we focus on four-year growth horizons, starting in each possible base year. Firms that dissolve within the four-year horizon account for roughly 4.5 percent of (employment weighted) firms in the 1980s and 1990s, and roughly 3.5 percent in the period after the Great Recession. For our purposes, it is important to note that dropping dissolver firms does not affect survival, death, and firm growth rates.

### *Reconciling SBO, ASE, and ABS Data*

The LEOF survey inputs are from seven periodic Census surveys conducted between 2002 and 2018, under the Survey of Business Owner (SBO), Annual Survey of Entrepreneurs (ASE), and the (on-going) Annual Business Survey (ABS) data collection programs.<sup>7</sup> Our survey files cover the following years: SBO 2002, 2007, and 2012; ASE 2014, 2015, and 2016; and ABS for reference year 2017. The firm-level surveys have evolved over time in terms of questionnaire design, sample size, sampling strategy, mode, and response rates. Creating consistent cross-sections for LEOF requires first reconciling survey instruments, then adjusting for other differences in the LBD reweighting.

As shown in Table 1, the SBO, ASE, and ABS surveys collect information about business characteristics (panel A), owner demographics (panel B), and financing variables (panels C and D). The first block of variables is collected at the firm level, starting with an important qualifying question, which is whether any owner has a greater than 10 percent share.<sup>8</sup> The surveys in years after 2002 have questions about husband-and-wife ownership, including whether one or both spouses owns the business. The SBO and ASE surveys include questions about when the business was founded or acquired, and whether the business was operated out of the owner's

---

<sup>7</sup> Another survey program—the Characteristics of Business Owners (CBO)—was conducted every five years prior to the first SBO in 2002. The 1992 CBO micro file is available, and we have linked those records to the LBD. Unfortunately, that dataset is much smaller than the SBO and later files, and the inability to distinguish between subchapter C corporations and subchapter S corporations in the LBD legal form of organization information in years before 2002 makes it impossible to properly reweight the micro file.

<sup>8</sup> If the responding company does not have any owners with more than a 10 percent share, the respondent is directed to skip the owner section and move to the business section of the instrument. See for example, question 4 in SBO-1 2012. The ASE and ABS use a similar screener question for their business owner sections.

home. There are many other variables collected in some or all survey years that we do not use here, but are attached to the LEOF file and we expect will be made available for future research by qualified researchers on approved projects.

The SBO, ASE, and ABS surveys contain extensive information about owner demographics (panel B, Table 1) and how that owner is generally involved with the firm. The demographic information (along with ownership shares) is collected for the top four owners (top three owners in 2002) in the survey year. The first three variables for each owner cover owner function, average hours worked in the business, and whether the business is the primary source of income for that owner. Standard demographics also include owner veteran status (and service-related disability), sex, age, race/ethnicity, and education. The surveys also ask how each owner acquired ownership, and when, and whether that owner had any previous business ownership experience. The surveys ask about immigrant status in different ways, whether “born in the U.S.” or “born a citizen of the U.S.” The ABS for reference year 2017 fields two questions that cover the birth location and citizenship questions. Beginning in 2014, the surveys ask owners WHY they want to run a business. The list of possible reasons varies by survey, including some common answers such as “wanted to be my own boss” or “could not find other work.”

The final block of common survey variables involves business financing, including both retrospective questions about business start-up (panel C, Table 1) and on-going financing (panel D, Table 1). Common variables across all seven surveys include high-level sources of start-up financing such as bank loans, credit cards, home equity loans, and outside investors. Beginning in 2007, the surveys also collect owner reports of (nominal) start-up capital amounts. As with firm and owner characteristics, not all financial variables are collected in all LEOF years. For example, there is no information about start-up capital amounts in 2002, and home equity loans are not a separate start-up source in 2002. The only ABS reference year for which we currently have start-up finance measures is 2017. Selected later versions of the ABS include finance questions and could be incorporated into LEOF once the micro data are available.

In this paper we focus extensively on the correlation between start-up financing and firm growth, but as Table 1 shows, there is a wealth of other financing information across the LEOF survey years that can be used in future research. For example, the surveys also generally include

questions about on-going financing. In the SBO years, these are just sources of current financing. Beginning in 2014, however, the surveys include questions about sources and levels of capital used for on-going capital investment. The questions include whether the firm sought new forms of financing, and if not, whether it did not seek new funding because of fear of being rejected.

### *Linking Survey and LBD Files*

The SBO, ASE, and ABS survey micro files include the crucial EIN that lets us link a surveyed firm to its LBD records. However, the target populations, sample frames, survey mode, and other aspects of data collection have evolved substantially across the years. For example, one important change occurred between the SBO and ASE, when the surveys shifted from all firms to employer firms. The sample size of our micro data for each of these three surveys is shown in line 1 of Table 2. Also, the SBO was conducted using traditional paper mail-in responses, while the ASE and ABS are on-line. Another evolving factor is survey (item and unit) non-response rates.<sup>9</sup> These changes are reflected in a widely varying LEOF share of all BDS firms across years, ranging from 13.6 percent in the 2002 SBO to 2.9 percent in the 2016 ASE.

Our approach to building LEOF effectively controls for these evolving survey differences by using the LBD population to create a set of inverse probability weights for each survey after linking “usable” survey records to the LBD. The first step in that process is determining whether

---

<sup>9</sup> One challenge is what match year to use for linking LBD and survey year records. It would seem natural to match the reference year of the survey to the corresponding year of the LBD. We used such as an approach for the SBO. However, for the ASE and ABS we found some issues with using this approach in terms of sample sizes for very young firms. For this vintage of the LEOF infrastructure (i.e., the results reported in this paper), when matching to the ASE and ABS we use the LBD in the year prior to the survey reference year to yield larger samples of young firms. In considering these issues it is instructive to keep in mind the timing of Business Register (BR) operations at the Census Bureau. For a given reference year, the BR is closed out in the fall of the succeeding year. However, the BR for a given reference year undergoes considerable refinement in County Business Patterns (CBP) processing (which is not completed until spring of the second year relative to the reference year). Moreover, the LBD is not completed for a reference year until September of that second year. This timing is relevant, since surveys such as the ASE and ABS for a given reference year are drawn from a BR sample and in the field by fall of the year following the reference year. Importantly, our inverse propensity score matching implies that for whatever target year we use for the LBD, our matched sample replicates the firm size, age, and industry patterns for that year. The implication is that our matched sample in the target year is on an inverse probability-weighted basis for the representative year. Moreover, key questions in the ASE and ABS explicitly refer to firm financing conditions at start-up. There is a potential tradeoff between sample size and potential timing issues for some owner characteristics (e.g., the owner characteristics for a matched firm will reflect the owner characteristics for the survey reference year). We plan to document and investigate this tradeoff in future versions of the LEOF database and documentation. We anticipate this issue is not important for the basic facts reported below.



a survey micro record is “usable” for our purposes.<sup>10</sup> Our concept of a “usable” firm begins with the Census Bureau response indicator variable, but we also require that at least one owner have valid core demographics such as age, sex, and race/ethnicity, and that the firm is in scope for the target population (meaning the firm would be a part of the BDS population). The fraction of firms in each survey year eligible for inclusion in LEOF has declined from 67 percent in 2002 to 33 percent in 2017 (Table 2, row 2).

Identifying LEOF-eligible SBO, ASE, and ABS survey records is the first step in solving for a survey/LBD match rate. The second step is distinguishing between employer and non-employer firms (Table 2, row 3). In the ASE and ABS years, the sample is employer firms. In the SBO years, the sample included both employers and non-employers, and some (survey-based) non-employer firms show up with positive employment after linking to the LBD. However, some (survey-based) employer firms show up as non-employers in the LBD. In the ASE and ABS years, some firms that were employers when sampled show up as non-employer business after linking back to the LBD. Thus, rows 4 and 3 in Table 2 are slightly different. The bottom line in Table 2 is that nearly every survey firm eligible to be matched to the LBD is in fact matched. The match rates (Table 2, row 4) are generally 99 percent or higher. Some matched firms are subsequently dropped from the LEOF sample because of remaining categorical exclusions such as BDS ineligibility based on industry or other characteristics.

### *LEOF Sample Weights*

Knowing that virtually every eligible survey firm is linked to an LBD firm is important but does not in and of itself address the question of whether the linked sample is representative. As with any survey, standard practice involves reweighting (or post-stratifying) to account for differences in sampling and/or response rates by key characteristics. Our goal for LEOF is to represent the BDS population within the LBD universe in each survey year by firm age, firm size, and industry.

---

<sup>10</sup> Census Bureau SBO, ASE, and ABS micro files reflect the evolving differences described above. The survey files are tabulated independently for regular table production and include (for example) imputations.

The details about our LEOF inverse-probability weighting algorithm are presented in Appendix A. One important feature of LEOF comes out of the reweighting exercise. The linked survey records all pass the qualifying threshold of having one or more owner with at least a 10 percent ownership share. This of course limits the representation of large (especially publicly traded) corporations with large numbers of employees. Thus, the LEOF sample excludes firms with 10,000 or more employees.<sup>11</sup> That exclusion does not change the fundamental patterns in the LBD data that motivate our research (more on this below).

In Appendix A we also show that the inverse weighting procedures generate a representative linked sample in every base year across two important dimensions. We show results for two important separate sets of statistics—the fraction of firms that qualify as high growth and average growth rates for high-growth firms—across time and firm age groups. LEOF has sampling variability and a slight bias in employment outcomes because our inverse weighting does not control for firm survival probability, which is correlated with survey response. However, the comparison of LEOF and LBD employment dynamics reassures us that the inverse sampling procedure is accomplishing the desired outcome.

### *Importance of Firm Age*

In developing the LEOF data infrastructure, an important feature we can take advantage of is the high-quality measurement of firm age in the LBD. Analysis of firm dynamics by firm age using the LBD has yielded new insights and an extensive research literature exploring these insights. Young firms have a higher dispersion of growth rates exhibiting an “up or out” dynamic (see, e.g., Davis and Haltiwanger (2014) and Decker et al. (2014, 2016)). That is, young firms post-entry have higher exit rates than mature firms but, conditional on survival, higher average growth rates than mature firms. The latter is driven by the greater right skewness of growth rates of young firms post-entry relative to mature firms.

In innovation-intensive sectors, young firms exhibit these up or out growth dynamics and are more innovation-intensive (see, e.g., Acemoglu et al. (2018)). Relatedly, young firms are an

---

<sup>11</sup> Based on publicly available BDS statistics, in 2021 firms with 10,000 or more employees accounted for 0.02 percent of all employer firms, but roughly 30 percent employment at employer firms.

important part of the experimentation associated with the development and adoption of new technologies (see Foster et al. (2021), who use the LBD to explore the hypotheses of Gort and Klepper (1982)).

To help illustrate the importance of firm age, in Figure 1 we present the average (over time) growth rate distribution of young, mature, and all incumbent firms using the LBD. The growth rate distribution presented is based on employment-weighted growth rates over a four-year horizon for all firms that are at least one year old (so Figure 1 excludes firm entry but includes firm exit). Young firms are between ages of 1-5 years inclusive, and mature firms are 6 years or older.<sup>12</sup> Each type of firm is divided into 20 percentile groups of the employment growth rate distribution using the Davis, Haltiwanger, and Schuh (1996) (DHS) growth rates. We report the employment-weighted average growth rate for each of the percentile groups.

Young firms are much more likely to exit – Figure 1 shows that young firms with a DHS growth of -200 percent (the DHS growth rate for exiting firms, which is a -100 percent growth rate) account for between 15-19 percent of young firm employment. In contrast, the corresponding share for mature firms is less than 5 percent. In the upper end of the growth rate distribution, young firms are more likely to grow rapidly than mature firms. The average four-year growth rate for young firms in the 75-79 percentile group is 58 percent DHS change (an 82 percent growth rate) compared to 31 percent DHS change (a 37 percent growth rate) for mature firms. This gap becomes larger for the top growth rate percentile bins. Figure 1 thus illustrates lower survival rates for young firms but higher right tail growth rates (i.e., up or out dynamics) – these patterns underlie the greater dispersion of growth rates for young firms.

In addition to the substantial cross-sectional differences by firm age, the age distribution of employment and the shape of the growth rate distributions have been changing over the last few decades. The most dramatic change is that the shares of young firms and their employment have decreased substantially over time (see Figure 2).<sup>13</sup> This decline in young firm activity has

---

<sup>12</sup> In BDS age statistics, young firms are those with firm ages 1-5 and firm entrants have firm age 0. Firm age is defined based upon the age of the oldest establishment at the firm at the time of the firm's legal entity coming into being.

<sup>13</sup> In Figure 2, we focus on firms with less than 10,000 employees which the LEOF data infrastructure is restricted to given sample restrictions in core surveys used. Using the full LBD/BDS yields very similar patterns.

been the subject of a large literature (see, e.g., Alon et al. (2018), Davis et al. (2007), Davis and Haltiwanger (2014, 2024), Decker et al. (2014, 2016, 2020) and Karahan et al. (2021)) in terms of causes and consequences. There have also been notable changes in the shape of the growth rate distribution for existing firms by firm age. We show the distribution of four-year growth rates for 1981-2016 (and sub periods) for young firms (Figure 3) and mature firms (Figure 4). While the qualitative features of the growth rate distributions of firms at least one year old by firm age remain similar, exit rates of young and mature firms have declined and the average growth rate of high-growth firms (e.g., above the 75<sup>th</sup> percentile) has declined for both young and mature firms. These changes on the extensive and intensive margin underlie the decline in indicators of business dynamism that have also received much attention (see, e.g., the papers cited directly above and references therein).

The LEOF data infrastructure has the potential to help shed new light on the underpinnings of these cross sectional and changing patterns of firm dynamics by firm age. In what follows, we provide some basic facts that emerge from the LEOF data infrastructure about these owner and business characteristics by firm age in the cross section and over time. We also provide analysis of the pooled cross section and time series relationships between these characteristics and the propensity to be a high-growth firm by firm age. This analysis is intended to be illustrative of the variation that LEOF permits analyzing through exploiting the associations between firm dynamics by firm age and business owner and business characteristics in the LEOF data infrastructure.

### **III. Trends in Owner Demographics and Financing**

LEOF combines longitudinal administrative and cross-section survey datasets, which makes it possible to answer many questions about the interplay between employment dynamics, owner characteristics, and financing. Our starting point for exploring LEOF is to investigate what happened since 2002 in terms of the characteristics of firm owners and types of start-up financing. These trends are particularly interesting in light of the trend decline in young firm employment shares that occurred over the same period, amplified by a sharp drop in young firm employment during the Great Recession (Davis and Haltiwanger, 2024).

The trends we investigate are essentially tabulations of the LEOF survey variables with weight adjustments for consistency between the survey and LBD populations. We also make use of the administrative LBD component of LEOF in two additional ways. First, the LBD gives us a consistent method for determining the firm's age, which is the (BDS-consistent) year in which the firm hired its first paid employee (technically, the age of the oldest establishment at the birth of the firmid; thus, a firmid that shows up as the result of a merger or acquisition would have an age based on the oldest establishment). Second, LBD employment makes it possible to look at trends on either a firm- or employment-weighted basis. Here we focus on employment-weighted outcomes, using the same DHS weights we used in the LBD analysis above, which is the average of base year  $t$  employment and employment in year  $t+4$ .<sup>14</sup>

### *Owner Demographics*

The period for which we have processed LEOF surveys spans 2002 through 2017. During that period, we have three SBO waves (2002, 2007, 2012), three ASE waves (2014, 2015, and 2016), and one ABS wave (reference year 2017). Between 2002 and 2017 the U.S. experienced large demographic shifts because of aging of baby boomers and the cumulative impact of increased female labor force participation and immigration.<sup>15</sup> What happened to the demographic characteristics of business owners during this period?

The short answer is that business owners, much like the labor force more generally, have gotten older, more female, and less non-Hispanic White.<sup>16</sup> In our next set of figures, we show shares of employment-weighted firms for selected series on owner ages, education, race/ethnicity, and gender. In addition to the outcome variables being measured, one question is

---

<sup>14</sup> The trends shown here are generally robust to the choice of firm- or employment-weighting.

<sup>15</sup> Aging of baby boomers, see for example <https://www.census.gov/library/stories/2023/05/2020-census-united-states-older-population-grew.html>. Female labor force participation rate and female share of civilian labor market rise over time before flattening during our sample period (see <https://fred.stlouisfed.org/series/LNS11300002> and <https://www.dol.gov/agencies/wb/data/lfp/civilianlfbyssex>). The foreign-born share of the U.S. population rose from 11.1 percent in the 2000 Census to 13.7 percent in the 2017 American Community Survey (see [MPI-Data-Hub\\_Immigrants\\_N-Percent-US-Population\\_2022\\_0.xlsx](https://www.census.gov/data/tables/2022/ipeds/mpi-data-hub-immigrants-n-percent-us-population-2022-0.xlsx) (live.com)).

<sup>16</sup> People aged 55 and over made up 12.9 percent a larger share of the U.S. civilian labor force in 2017 (22.8 percent) than in 2000 (12.9 percent). Non-Hispanic Whites made up less of the labor force in 2017 (62.8 percent) than 2000 (73.1 percent). The female share of the labor force rose slightly from 46.6 percent of the labor force in 2000 to 46.9 percent in 2017. The 2000 numbers are from Toossi (2002), and the 2017 numbers are from Bureau of Labor Statistics (2018).

how to think about owner characteristics when there are multiple owners. One possibility is to weight the characteristics of all owners using ownership shares, but then (for example) a firm with one young owner and one old owner would show up in the same category as a single middle-aged owner. The selected series shown in Figures 5 and 6 therefore use either “any” or “all” measures.

Figure 5 shows the shares of employment-weighted firms with at least one owner aged 55 or older and at least one owner aged 35 or younger by the age of the firm (thus a firm could appear in both categories). The lack of markers on the lines for non-survey years is a reminder that we do not have ownership data in those years, and those values are interpolations between the survey years. The trends in owner ages clearly reflect the broader demographic aging experienced during this period, with the fraction of firms with any older owner rising substantially from 34 percent to 42 percent among young firms, and from 59 to 75 percent among mature firms. The shares of firms with any young owner are much lower and decline slightly over this same period.<sup>17</sup>

Figure 6 focuses on three other sets of owner demographics where we use the “all” designation in classifying a firm. Overall, college-educated owners tend to be at mature firms rather than young firms (the solid line is above the dotted line), while young firms have a higher propensity to have only female owners or no non-Hispanic White owners (the dotted lines for each are above their associated solid line). Firms with all college-educated owners have increased in prevalence by about 5 percentage points among young firms and 3 percentage points among mature firms. Firms in which no owners are non-Hispanic White surged from 10 percent of young firms in 2002 to 18 percent by 2017, and young firms with all female owners increased from 10 percent to 14 percent.<sup>18</sup> The changes in mature firm owner demographics are more muted, reflecting both a pipeline effect (those young firms will become mature firms over time)

---

<sup>17</sup> For comparison, the firm-weighted share of young firms with at least owner aged 55 or older increased from 26 percent to 37 percent over the same period. Thus, the employment-weighted shares in Figure 5 (34 percent in 2002, rising to 42 percent in 2017) are higher (young firms with older owners have more employees). However, both firm- and employment-weighted shares of firms with older owners are increasing substantially.

<sup>18</sup> The differences between firm- and employment-weighted outcomes are even more pronounced for all female- and no non-Hispanic White-owned firms because those have lower employment. For example, the firm-weighted share of young firms with all female owners increased from 15 to 21 percent, and the share of young firms in which no owners were non-Hispanic White from 12 percent to 22 percent.

and the differential in employment dynamics discussed above (those young firms must survive to become mature firms).

### *Start-up Financing Sources*

As shown in Table 1, the SBO, ASE, and ABS ask owners to check “all that apply” when presented with a list of the funding sources they used when they started the business. The list includes items such as “personal or family savings” and “personal or family assets” which are checked by most owners in all the surveys. The list also contains more focused sources of funding, which we aggregate into several broader types. Figure 7 shows the reported incidence of bank loans, credit cards, home equity loans, and outside investors for young firms across the seven LEOF survey years.<sup>19</sup>

The trends in employment-weighted start-up financing sources reported by young firms tell a compelling story. As shown in Figure 7, the incidence of bank loans fell from about 35 percent in 2002 to near 20 percent by 2012, where it has remained. Outside investors were involved in just over 10 percent of start-ups in 2002, and that share has fallen to about 7 percent in recent years. And the incidence of home equity loans—unfortunately not captured in 2002—fell from about 13 percent in 2007 to 5 percent by 2017. The only relatively stable source is credit cards, reported by just over 10 percent of start-ups in every period.

These observed declines in the incidence of larger-scale financing sources (bank loans, home equity loans, and outside investors) are consistent with the general decline in start-up activity discussed in the previous section, and with the causal mechanisms described in Davis and Haltiwanger (2024). But these statistics only reflect outcomes for firms that *came into existence* during this period. If the availability of financing deteriorates, that is likely to also cause a decline in start-up activity generally, but we never see data on the firms that were never

---

<sup>19</sup> The start-up financing questions are most salient (and relevant) for young firms, because asking the owner of a mature firm what sources they used to start the firm is likely to be more subject to recall error and less relevant for the impact on firm growth (which we show in the next section).

founded. In the next section we explore the second possible financing channel, which is the correlation between financing and growth outcomes for LEOF firms.<sup>20</sup>

#### **IV. Correlations Between Owner Characteristics, Financing, and Employment Growth.**

In addition to looking at trends in owner characteristics and financing, LEOF is designed to study the relationships between longitudinal employment outcomes and the additional information available on the SBO, ASE, and ABS surveys. What types of owners are most likely to start firms that survive? What types of owners start firms that grow rapidly? What is the correlation between start-up financing and survival or growth, and how do those interactions depend on owner demographics?

This section provides examples of simple (but new) correlations that can be gleaned from the LEOF infrastructure. We make no attempt at an identification strategy that relies on plausibly exogenous financing, for example, as in Davis and Haltiwanger (2024). The correlations we identify are generally consistent with the idea that financing constraints have played a role in the decreased “dynamism” observed in LBD and BDS data, however, because well-financed firms are more likely to grow faster.

The primary measure of success we study here is whether a firm is in the top 25 percent of the (employment-weighted) four-year growth rate distribution for firms that are least age one.<sup>21</sup> As shown in Figures 1, 3, and 4, the 75<sup>th</sup> percentile of the growth distribution is around the cutoff where modest growth generally is observed. Our DHS employment growth cutoff for the top quartile of growth is fixed at the average for the sample period, because crossing the 75<sup>th</sup> percentile of growth in (say) the period 2007-2011 is very different than crossing the 75<sup>th</sup> percentile under more normal employment growth periods.

---

<sup>20</sup> The strong correlation we find between financing and employment growth does have implications for the first channel, even if we do not have data on the firms that did not come into existence because financial conditions prevented the start-up. We show that (unsurprisingly) successful firms are more likely to have large-scale financing, and if we interpret that to mean that potential success depends on the ability to get financing, then the fact that potential owners refrain from starting firms when financing is constrained becomes a key part of the overall story.

<sup>21</sup> An interesting use of the LEOF data infrastructure is to study the impact of financing conditions on startups (age=0 firms). We leave that for future work.



We investigate correlations between the “high-growth” employment outcomes, owner demographics, and start-up financing using a linear probability model. Specifically,

$$(1) \quad \text{Pr}(\text{high growth}_{i,t}) = \sum_t I_t \lambda_t + X_{it}' \beta + Z_{it}' \gamma + \varepsilon_{it}$$

Where  $I$  is a vector of year dummies,  $X$  are owner demographic characteristics, and  $Z$  are firm financing variables. The year dummies are the starting point for estimating correlations because we want to remove (for example) the dramatic differences between the incidence of the high-growth outcome in 2007-2011 versus other years. We run a series of regressions, incrementally adding owner demographics and then start-up financing variables from the surveys, then further controlling for employment size and age class and two-digit industry from the LBD.

Complete regression results for young and mature firms are shown in Appendix Tables B.1 through B.4. There are two sets of regressions each for young and mature firms, where the first set includes all LEOF years for which we have the key start-up sources, and the second set has years 2007 and later, for which we have both start-up sources and start-up amounts. One of the key meta-findings in the regression results is that the very large time dummies—on the order of 7 to 9 percentage points relative to the average probability of success equal to 25 percentage points—are relatively invariant to the inclusion of LEOF owner characteristics and start-up financing variables. The omitted year is 2007, when the probability of crossing the 75<sup>th</sup> percentile cutoff is much lower, and thus the dummies are large for every other year. But note that the time dummies generally increase slightly as we move across columns in any set of regressions, meaning that more firms crossed the success threshold despite changes in owner demographics and financing trends.

Our primary focus here is on the correlations between success (as captured by being above the high-growth threshold) and the LEOF owner and financing characteristics. These correlations are summarized in Figures 8, 9, and 10. Remembering that each of these dummies is relative to the average success probability of 25 percent, the estimated correlations between

employment growth and demographics, financing sources, and start-up amounts are generally economically (as well as statistically) significant.<sup>22</sup>

Figure 8 shows that having at least one owner aged 35 or younger raises the probability of being high growth by more than 5 percentage points, while having at least one owner aged 55 or older lowers the probability of success by a similar amount. College-educated owners are more likely to run high-growth firms, while all female firms—and especially all-female owned young firms—are less likely to cross the growth threshold. The differential effects for firms with no non-Hispanic White owners are generally smaller and go in opposite directions for young and mature firms. These results hold even after controlling for industry, so (for example) it is not just a matter of young owners choosing high growth sectors.

Figure 8 captures just one measure of success (four-year DHS growth) and there is no attempt at formal identification for the separate channels. For example, using LEOF we can see (not shown here) that firms with older owners are more likely to survive over the four-year period, which suggests that older owners may be more focused on stability than growth (which is itself a question that can be answered using other survey variables available using LEOF years in which those motive questions are asked). Firms with older owners are also generally larger in the base year, which may suggest they have less room for “growth” as measured by the DHS growth rate. The bottom line is that Figure 8 should not be interpreted as showing that firms with young firms are generally more successful.

Variation in the incidence of start-up financing sources (Figure 9) and start-up financing amounts (Figure 10) are also both statistically significant and economically meaningful. The correlations between the (generally) larger start-up sources (especially bank loans and outside investors) and reported start-up amounts are much more muted for mature firms, but again, those involve both salience (start-up may have happened many years ago) and relevance (a bank loan ten years ago has little correlation with current growth). And again, the same sorts of cautions about causal interpretations with owner demographics apply here as well.

---

<sup>22</sup> In the next iteration of the charts, we will add 95 percent confidence intervals to these bar charts. In addition, we will make a few changes to color/line types to other charts to increase their readability.

Indeed, one possible confounding factor in the correlation between start-up financing and growth is that certain types of owners may be more likely to have certain types of financing. This is indeed the case, which we capture using regressions of the incidence of bank loans (Table B.5 and Figure 11) and outside investors (Table B.6 and Figure 12) on time dummies and owner characteristics. The estimated correlations should be viewed relative to the incidence of the funding source, so, for example, the fact that young owners are roughly 2 percentage points more likely to have outside investors is notable given an overall incidence of 7 to 10 percentage points over the sample period.

Despite the possibly confounding interconnections, there is a lot of evidence that the correlations in even this simple exercise are robust. For example, the coefficients on owner demographics and financing variables do not change much as we move across columns in the regression tables, and that involves adding detailed firm age, size, and industry controls. Thus, although the estimated relationships may not be causal per se, they are likely to be an important part of any analysis of differences in firm employment dynamics.

## **V. Concluding Remarks**

We describe the construction of the new LEOF linked administrative/survey file, containing over two million 2002-2017 survey-year firm observations on owner demographics and firm start-up financing sources and amounts, as well as firm age and employment for all years in which the firm has positive employment. The data exhibit several clear trends over the 15-year period. Firms with older owners, only female owners, no non-Hispanic White owners, and college-educated owners have become more common. Fewer entrepreneurs use bank loan or home equity financing to start their businesses in 2017 as compared to 2002.

We find that the propensity to be a high-growth firm varies significantly with the owner and firm characteristics contained. Firms with young, more educated owners, those with at least one male owner, and young firms with a bank loan or outside investor at start-up are more likely to be in the upper quartile of the employment growth distribution.

Future research could further explore these findings. For example, the trends in owner and firm characteristics over time could be the result of changes within industries. Alternatively,

if the characteristics are more similar within than across industries, changes in industry composition could produce these trends. Researchers can distinguish these effects with the LEOF dataset. With the LEOF data, one can extract all firms of a given age in any calendar year. Thus, a five-year-old firm interviewed in 2017 would be in the same age cohort as a four-year-old firm interviewed in 2016. LEOF makes it possible to discern separate age, time, and cohort effects. Researchers could also study how associations between owner/firm characteristics and firm outcomes vary with local or industry-specific economic conditions and with policy changes. Additional outcome measures could be examined as well.<sup>23</sup>

The LEOF data could potentially contribute to the BDS program, which is studying how financial administrative and private data from sources such as Securities and Exchange Commission filings and Compustat could be used to produce business dynamic statistics by type of financing. The SBO/ASE/ABS collect financing data on both the types of firms found in these financial administrative and private data, as well as young and small firms that infrequently appear in such sources. The LEOF could thus supplement and validate these other data sources. In addition, the LEOF exercises that we conduct could provide useful information to another planned BDS experimental product tracking high-growth firms.<sup>24</sup>

We hope that the LEOF and this paper will encourage researchers to explore other uses and questions that we have not considered. It would be possible to update the LEOF as the ABS continues to collect business and owner characteristics on a yearly basis and financing over select years. The user base would help determine whether this would be useful. Researcher input is a critical component of the feedback loop that enables the Census Bureau's Center for Economic Studies to continuously refine and improve its research datasets.

---

<sup>23</sup> For example, Lee et al. (2023) link LBD revenue and U.S. Patent and Trademark Office patent data to the ASE to study associations between owner nativity and innovation and productivity.

<sup>24</sup> The full set of existing and planned BDS experimental products are described in: <https://www.census.gov/programs-surveys/ces/data/public-use-data/experimental-bds.html>.

## References

- Acemoglu, Daron, Ufuk Akcigit, Nicholas Bloom, and William. R. Kerr (2018) "Innovation, Reallocation and Growth," *American Economic Review*, 108, 3450-3491.
- Alon, Titan, David Berger, R. Dent, and Benjamin Pugsley, "Older and slower: The startup deficit's lasting effects on aggregate productivity growth," *Journal of Monetary Economics*, 2018, 93.
- Azoulay, Pierre, Benjamin F. Jones, J. Daniel Kim, and Javier Miranda. 2020. "Age and High-Growth Entrepreneurship," *AER Insights*, 2(1): 65-82.
- Brown, J. David, John S. Earle, Mee Jung Kim, Kyung Min Lee, and Jared Wold. 2022. "Black Owned Firms, Financial Constraints, and the Firm Size Gap," *AEA Papers and Proceedings*, 112:282-286.
- Brown, J. David, John S. Earle, Mee Jung Kim, and Kyung Min Lee. 2019. "Start-ups, Job Creation, and Founder Characteristics," *Industrial and Corporate Change*, 1–36. doi: 10.1093/icc/dtz030.
- Brown, J. David, John S. Earle, and Yana Morgulis. 2017. "Job Creation, Small versus Large versus Young, and the SBA," in *Measuring Entrepreneurial Businesses*, John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar eds., University of Chicago Press.
- Bureau of Labor Statistics. 2018. "Labor force characteristics by race and ethnicity, 2017," BLS Reports 1076, August.
- Chow, Melissa C., Teresa C. Fort, Christopher Goetz, Nathan Goldschlag, James Lawrence, Elisabeth Ruth Perlman, Martha Stinson, and T. Kirk White. 2021. "Redesigning the Longitudinal Business Database," NBER Working paper 28839.
- Davis, Steven J., John Haltiwanger, and Scott Schuh, 1996. *Job Creation and Destruction*. Cambridge, Massachusetts: MIT Press.
- Davis, Steven J., and John Haltiwanger, 2014. "Labor Market Fluidity and Economic Performance," in *Re-Evaluating Labor Market Dynamics*, Federal Reserve Bank of Kansas City. Also available at NBER Working Paper No. 20479.
- Davis, Steven J., and John C. Haltiwanger. 2024. "Dynamism Diminished: The Role of Housing Markets and Credit Constraints," *American Economic Journal: Macroeconomics* (forthcoming).
- Davis, Steven J., John Haltiwanger, Ron Jarmin, and Javier Miranda, 2007. "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms." *NBER Macroeconomics Annual*, 2006.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2016. “Where has all the skewness gone? The decline in high-growth (young) firms in the U.S.” *European Economic Review*, 86:4–23.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. “The Role of Entrepreneurship in U.S. Job Creation and Economic Dynamism,” *Journal of Economic Perspectives*, 28(3): 3-24. (Summer)

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda (2020) “Changing Business Dynamism and Productivity: Shocks vs. Responsiveness,” *American Economic Review*, 110, 3952-3990.

Foster, Lucia, and Patrice Norman. 2017. “The Annual Survey of Entrepreneurs,” *Journal of Economic and Social Measurement*, (2017) 42:199–224. DOI 10.3233/JEM-180446.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. (2021) “[Innovation, Productivity Dispersion, and Productivity Growth](#).” In *Measuring and Accounting for Innovation in the 21<sup>st</sup> Century*, ed. Carol Corrado, Jonathan Haskel, Javier Miranda, and Daniel Sichel. Chicago: University of Chicago Press.

Gort, Michael and Steven Klepper. 1982. “Time Paths in the Diffusion of Product Innovations,” *The Economic Journal*, Vol. 92, no. 367 (Sept.), 630–53.

Jarmin, Ron, and C.J. Krizan. 2010. “Past Experience and Future Success: New Evidence on Owner Characteristics and Firm Performance.” U.S. Census Bureau, Center for Economic Studies Working Paper CES-10-24.

Karahan, Fatih, Benjamin Pugsley, and Aysegul Sahin (2021) “The Demographic Origins of the Startup Deficit,” mimeo.

Lee, Kyung Min, Mee Jung Kim, J. David Brown, John S. Earle, and Zhen Liu. 2023. “Are Immigrants More Innovative? Evidence from Entrepreneurs,” U.S. Census Bureau, Center for Economic Studies Working Paper CES 23-56.

Sterk, Vincent, Petr Sedlacek, and Benjamin Pugsley. 2021. “The Nature of Firm Growth,” *American Economic Review*, 111(2): 547-579.

Toossi, Mitra. 2002. “A century of change: the U.S. labor force 1950-2050,” *Monthly Labor Review*, May, 15-28.

**Table 1. LEOF Business, Owner, and Financing Variables**

<i>Panel A: Business Characteristics</i>	<b>SBO</b>			<b>ASE</b>			<b>ABS</b>
	<b>2002</b>	<b>2007</b>	<b>2012</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>
Does any individual own more than 10%?	X	X	X	X	X	X	X
Husband and wife owned (and equal or not)?		X	X	X	X	X	X
Year business established	X	X	X	X	X	X	
Home-based business?	X	X	X	X	X	X	
Family owned?	X	X	X	X	X	X	X
Franchise?	X	X	X	X	X	X	
Total number of owners		X	X	X	X	X	X
Types of workers (FT, PT, day labor, contractors)	X	X	X	X	X	X	X
Types/sales shares of customers (government, private, ...)	X	X	X	X	X	X	X
Types of benefits (health, retirement, paid leave, ...)		X	X	X	X	X	X
Currently operating? If not, reasons...		X	X	X	X	X	X
Domestic versus international sales		X	X	X	X	X	X
Business profitability and factors affecting profitability				X	X	X	X
Operations outside United States, primary language for business		X	X	X	X	X	
Business aspirations (larger, smaller, same size)				X	X	X	
Website and e-commerce details		X	X	X	X	X	

**Table 1. LEOF Business, Owner, and Financing Variables (Continued)**

<i>Panel B: Owner Characteristics (Top 4 owners)</i>	<b>SBO</b>			<b>ASE</b>			<b>ABS</b>
	<b>2002</b>	<b>2007</b>	<b>2012</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>
Ownership share top four owners (0 to 100%)	X	X	X	X	X	X	X
Owner function (production, management, finance, none)	X	X	X	X	X	X	X
Average hours	X	X	X	X	X	X	X
Primary source of income for this owner?	X	X	X	X	X	X	X
Veteran status and service-related disability	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X
Age	X	X	X	X	X	X	X
Hispanic	X	X	X	X	X	X	X
Race	X	X	X	X	X	X	X
Education	X	X	X	X	X	X	X
Field of highest degree						X	X
How acquired (founded, purchased, inherited, gifted)		X	X	X	X	X	X
Year ownership acquired?	X	X	X	X	X	X	X
Previous business ownership/self-employment?		X	X	X	X	X	X
Born in the United States?		X					X
Born a citizen of the United States?			X	X	X	X	
Currently a citizen of United States?							X
Reasons for being a business owner (10-12 reasons, varies)				X	X	X	X



**Table 1. LEOF Business, Owner, and Financing Variables (Continued)**

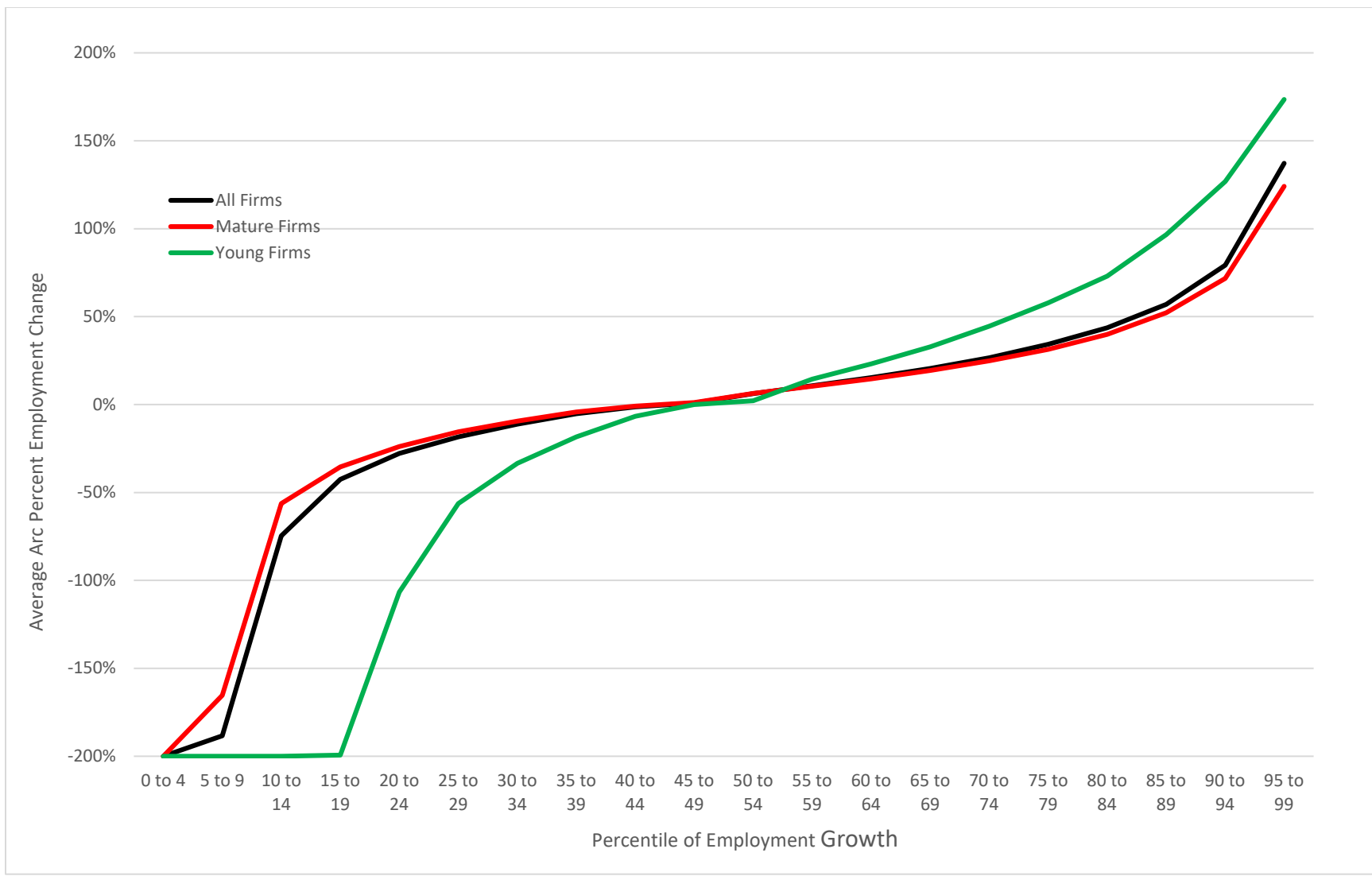
<i>Panel C: Start-up Financing</i>	<b>SBO</b>			<b>ASE</b>			<b>ABS</b>
	<b>2002</b>	<b>2007</b>	<b>2012</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>
Personal/family savings	X	X	X	X	X	X	X
Personal/family assets other than savings	X	X	X	X	X	X	X
Personal/family home equity loan		X	X	X	X	X	X
Personal/business credit card	X	X	X	X	X	X	X
Government (direct or guaranteed) loan	X	X	X	X	X	X	X
Government guaranteed loan	X	X	X	X	X	X	X
Bank loan	X	X	X	X	X	X	X
Business loan from family/friends		X	X	X	X	X	X
Outside investor/venture capitalist	X	X	X	X	X	X	X
Grants		X	X	X	X	X	X
Don't know or other		X	X	X	X	X	X
None needed	X	X	X	X	X	X	X
Total start-up capital (dollar ranges)		X	X	X	X	X	X
<i>Panel D: Ongoing Financing</i>							
Sources of funds for expansion (list varies)	X	X	X	X	X	X	X
Total capital used for expansion				X	X	X	X
New funding, received or not; if avoided applying, why?				X	X	X	X

**Table 2. LEOF Linked Sample Statistics**

	SBO			ASE			ABS
	2002	2007	2012	2014	2015	2016	2017
1. Survey Micro File Sample Size	2,247,000	2,245,000	1,045,000	420,000	430,000	435,000	1,483,000
2. Survey Firms with Valid Owner Information	1,507,000	1,226,000	663,000	202,000	185,000	178,000	482,000
<i>Usable survey record rate = Row 2/Row 1</i>	<i>67.1%</i>	<i>54.6%</i>	<i>63.4%</i>	<i>48.1%</i>	<i>43.0%</i>	<i>40.9%</i>	<i>32.5%</i>
3. Survey Employer Firms with Valid Owner Information	822,000	673,000	300,000	202,000	185,000	178,000	482,000
<i>Employer firms as a share of survey firms = Row 3/Row 2</i>	<i>54.5%</i>	<i>54.9%</i>	<i>45.2%</i>	<i>100.0%</i>	<i>100.0%</i>	<i>100.0%</i>	<i>100.0%</i>
4. Survey Firms with Valid Information Matched to LBD	817,000	667,000	302,000	200,000	183,000	176,000	479,000
<i>Survey/administrative match rate = Row 4/Row 3</i>	<i>99.4%</i>	<i>99.1%</i>	<i>100.7%</i>	<i>99.0%</i>	<i>98.9%</i>	<i>98.9%</i>	<i>99.4%</i>
5. Less: Categorical Exclusions	108,000	64,500	30,000	24,500	21,500	21,000	52,000
<i>Categorical exclusion rate = Row 5/Row 4</i>	<i>13.2%</i>	<i>9.7%</i>	<i>9.9%</i>	<i>12.3%</i>	<i>11.7%</i>	<i>11.9%</i>	<i>10.9%</i>
6. Equals: LEOF Final Sample Size	672,000	577,000	259,000	169,000	156,000	150,000	417,000
LEOF Final Sample Percent of BDS firms	13.6%	10.9%	5.1%	3.3%	3.1%	2.9%	8.0%

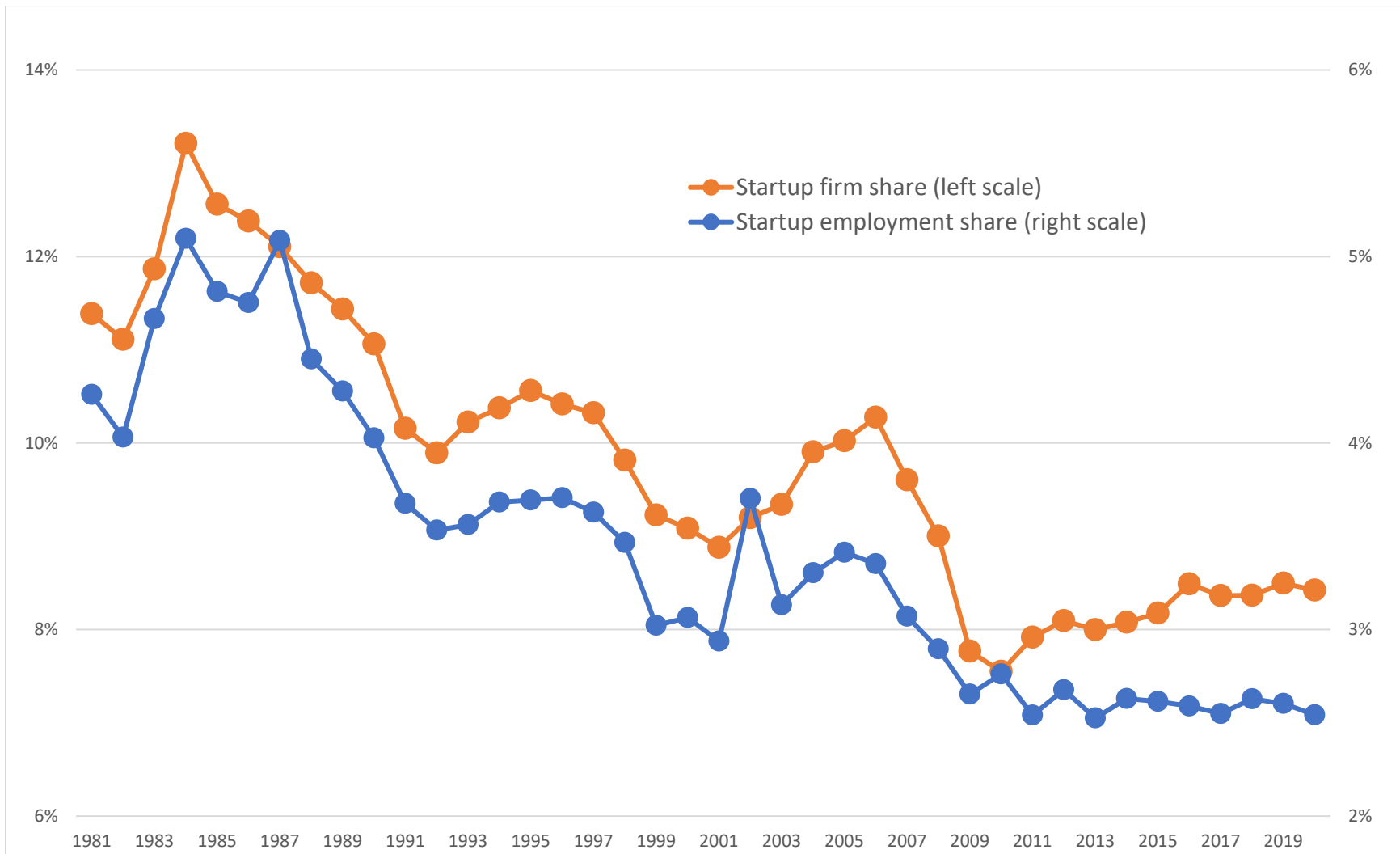
Note: Sample frame is all firms in SBO years, employer firms only in ASE and ABS years. Survey/administrative match rate can exceed 100 percent because some SBO non-employer firms in survey are identified as employer firms in the LBD. Categorical exclusions occur because of BDS inconsistencies based on industry, key missing variables.

**Figure 1. Distribution of Firm Four-Year Growth Rates, 1981 to 2016 (Employment Weighted)**



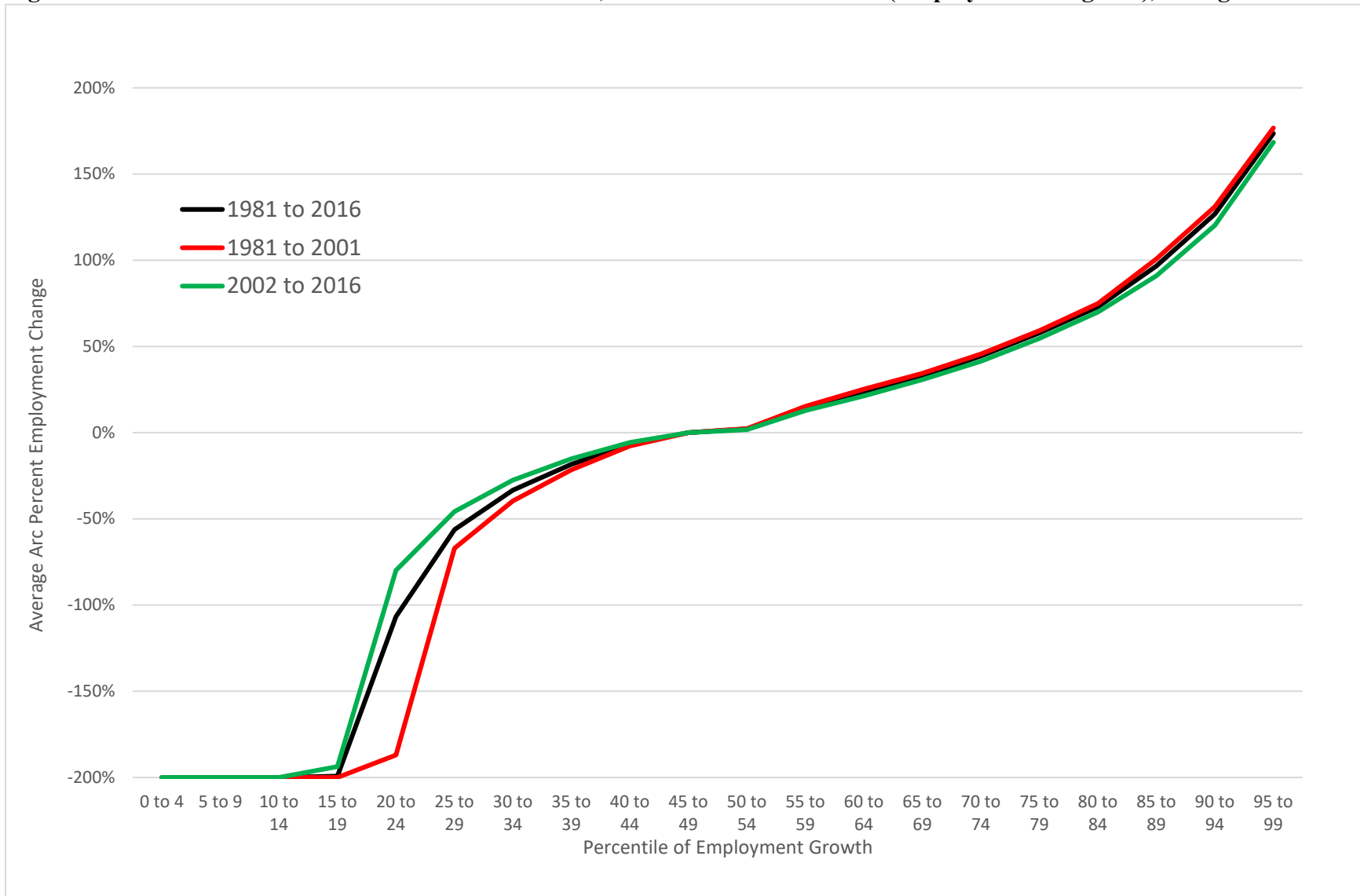
Notes: Tabulations from authors using the LBD. Young firms are firms age 1 to 5, mature firms are age 6+.

**Figure 2. Start Up Firm and Employment Shares (Fewer than 10,000 Employees)**



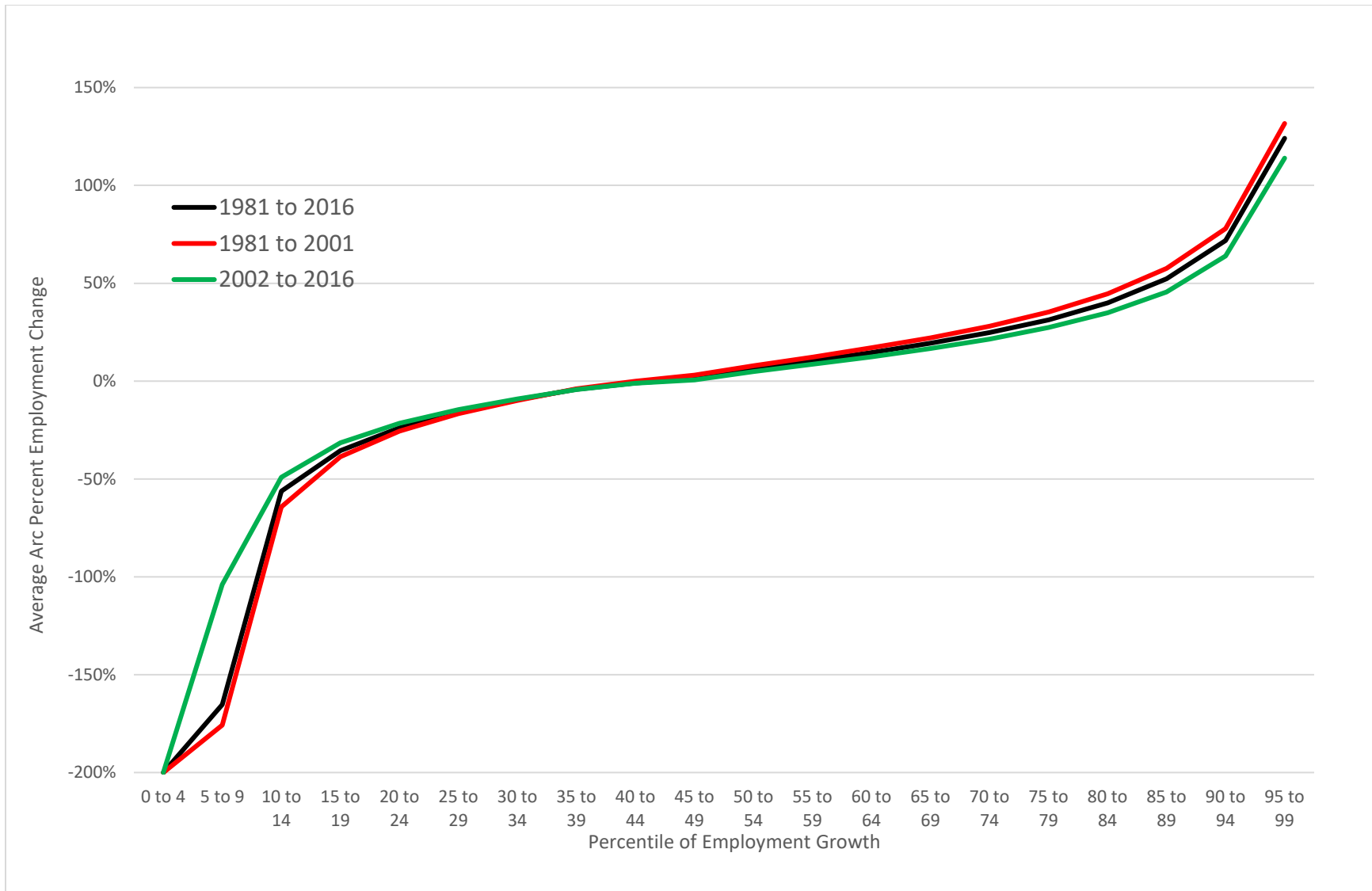
Notes: Tabulations by authors using the BDS. Startups are age=0 firms in the BDS.

**Figure 3. Distribution of Firm Four-Year Growth Rates, Sub Periods 1981 to 2016 (Employment Weighted), Young Firms**



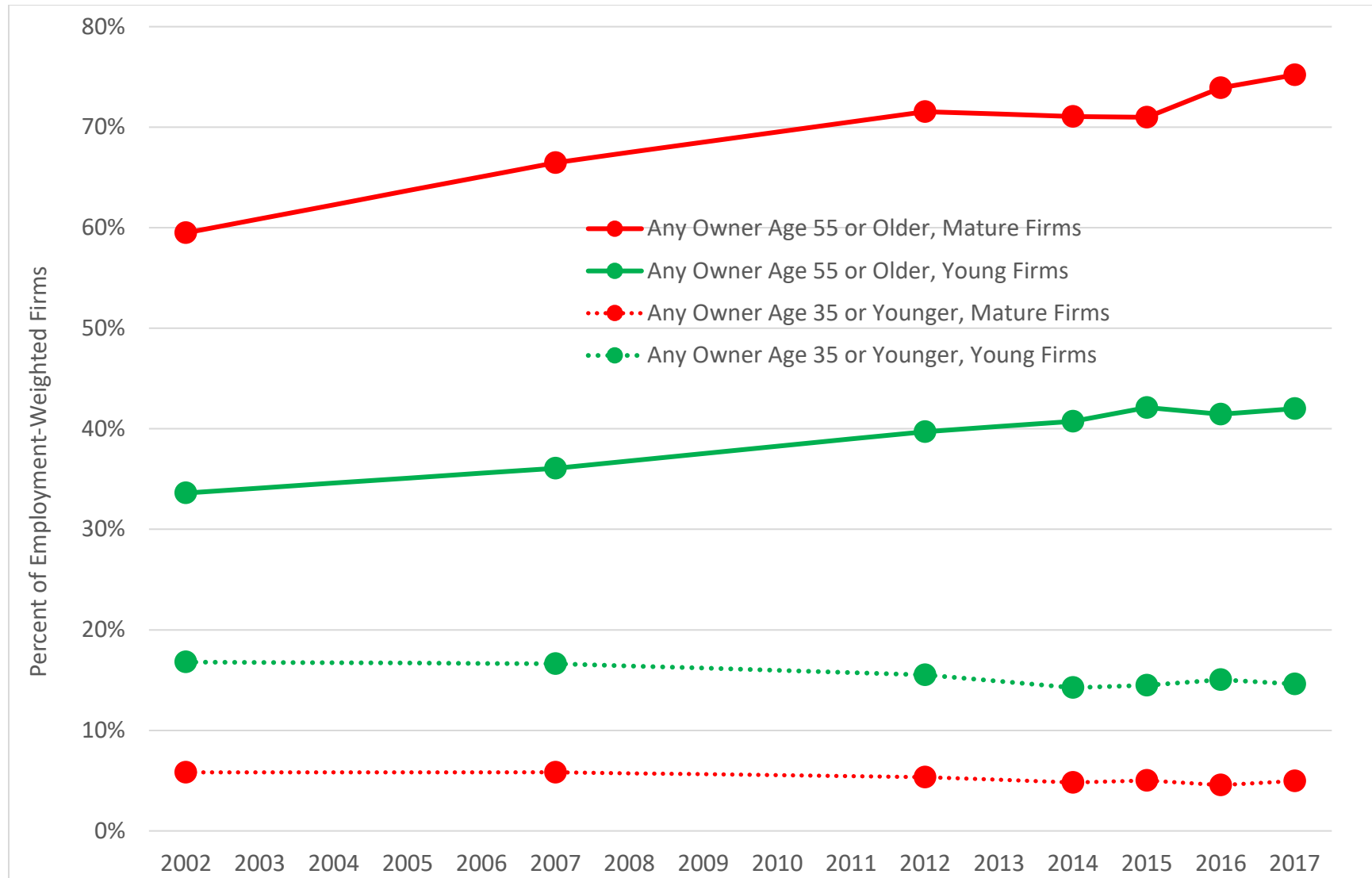
Notes: See notes to Figure 1.

**Figure 4. Distribution of Firm Four-Year Growth Rates, Sub Periods 1981 to 2016 (Employment Weighted), Mature Firms**



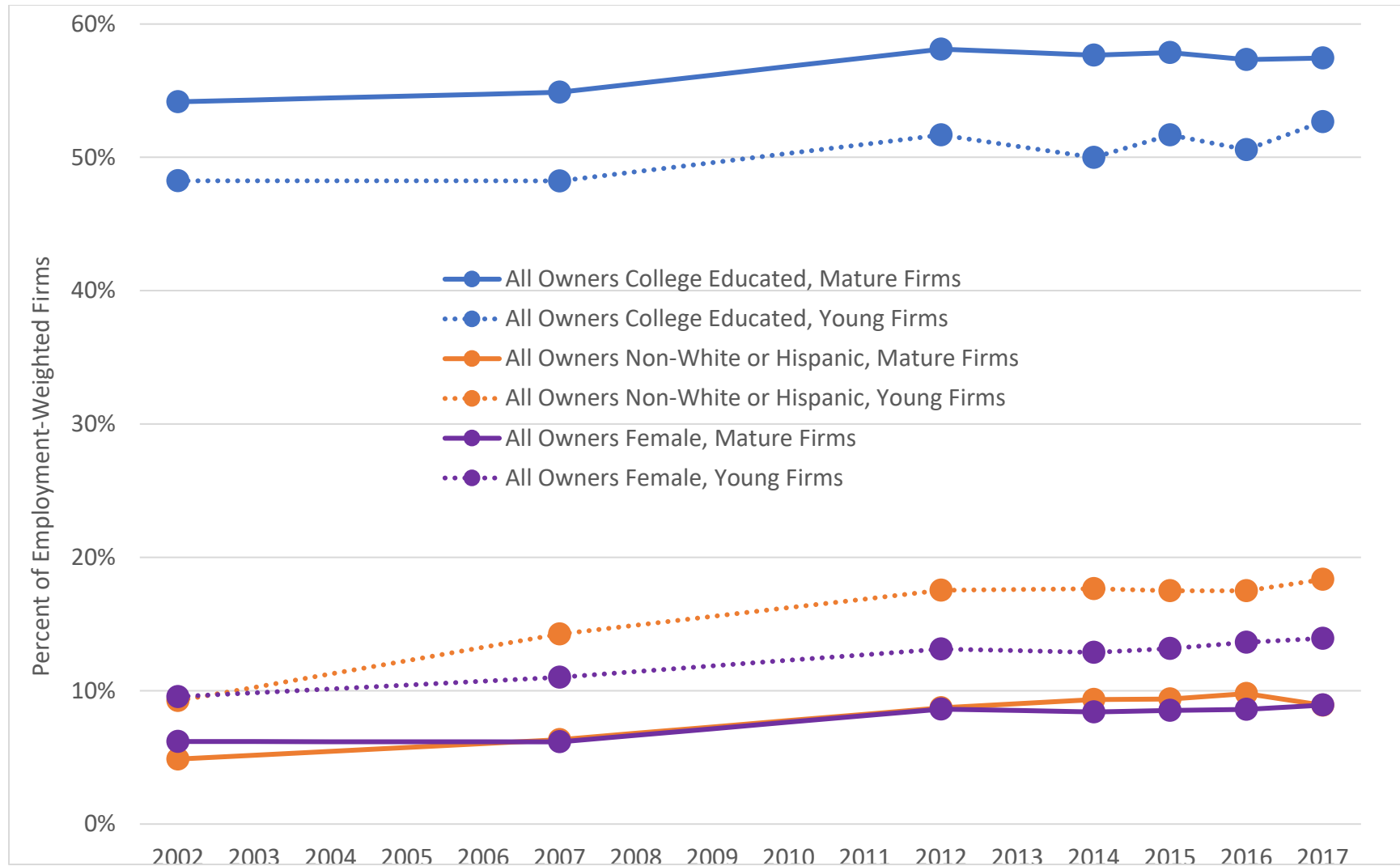
Notes: See notes to Figure 1.

**Figure 5. Trends in Owner Age Composition (Employment Weighted)**



Notes: Tabulations from authors from LEOF data infrastructure. Young firms are firm ages 1 to 5. Mature firms age 6+.

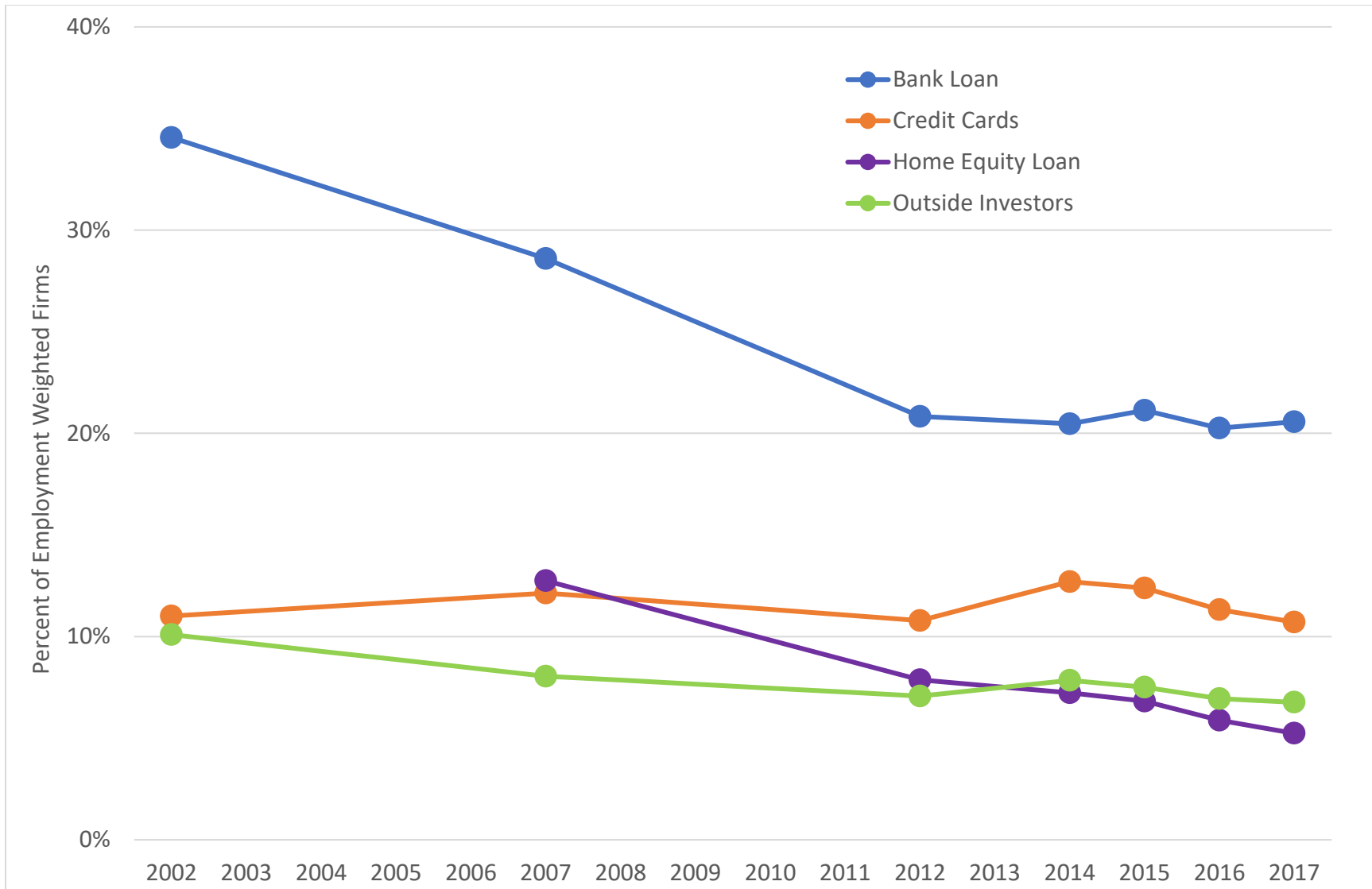
**Figure 6. Trends in Other Owner Demographics (Employment Weighted)**



Notes: See notes to Figure 6.



**Figure 7. Trends in Owner-Reported Start-up Financing Sources (Employment Weighted, Young Firms Only)**



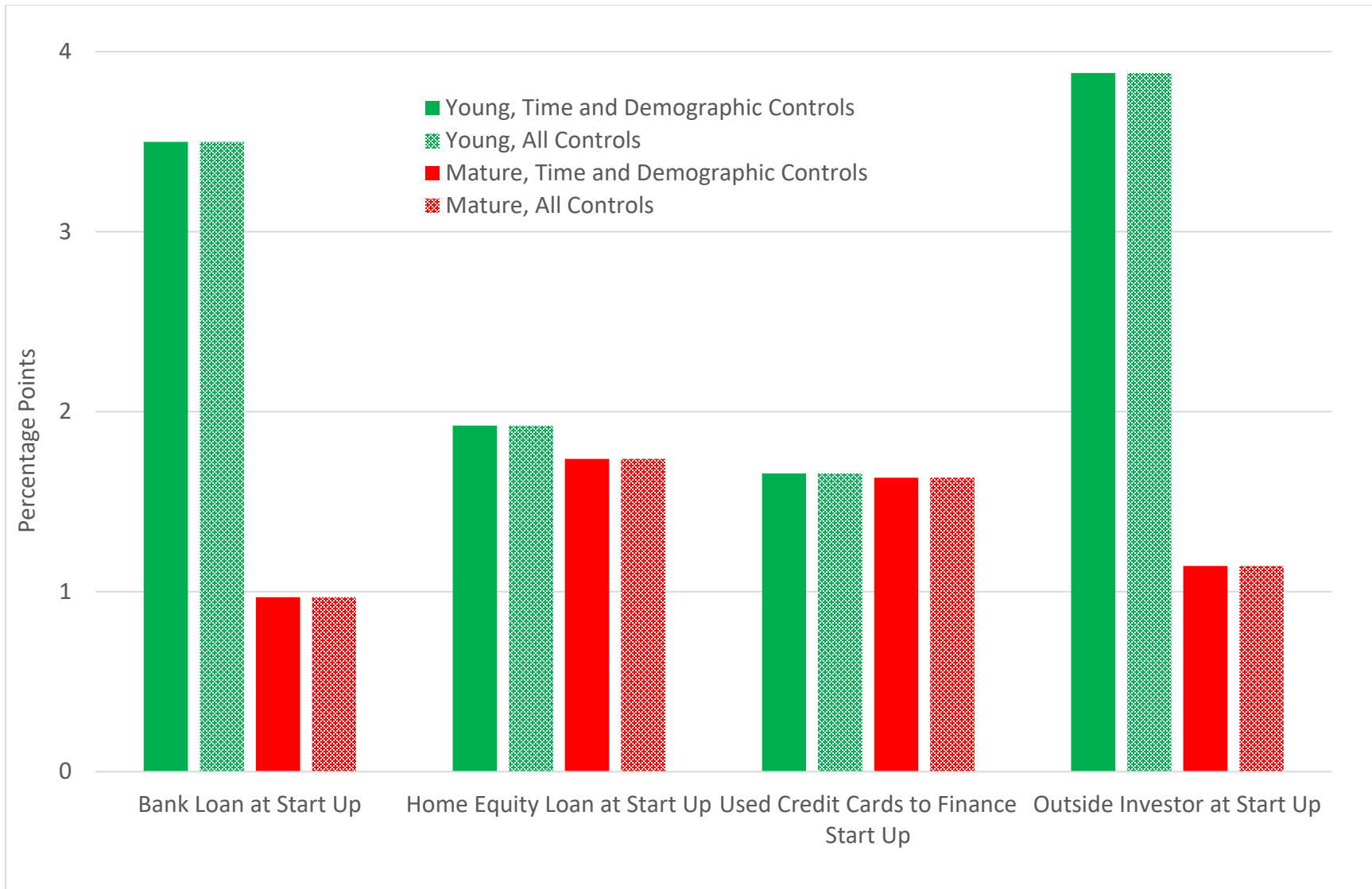
Notes: Source: LBD. Young firms are firms age 1 to 5, mature firms are age 6+.

**Figure 8. Probability of Being a High-Growth Firm, Owner Characteristic Coefficients**



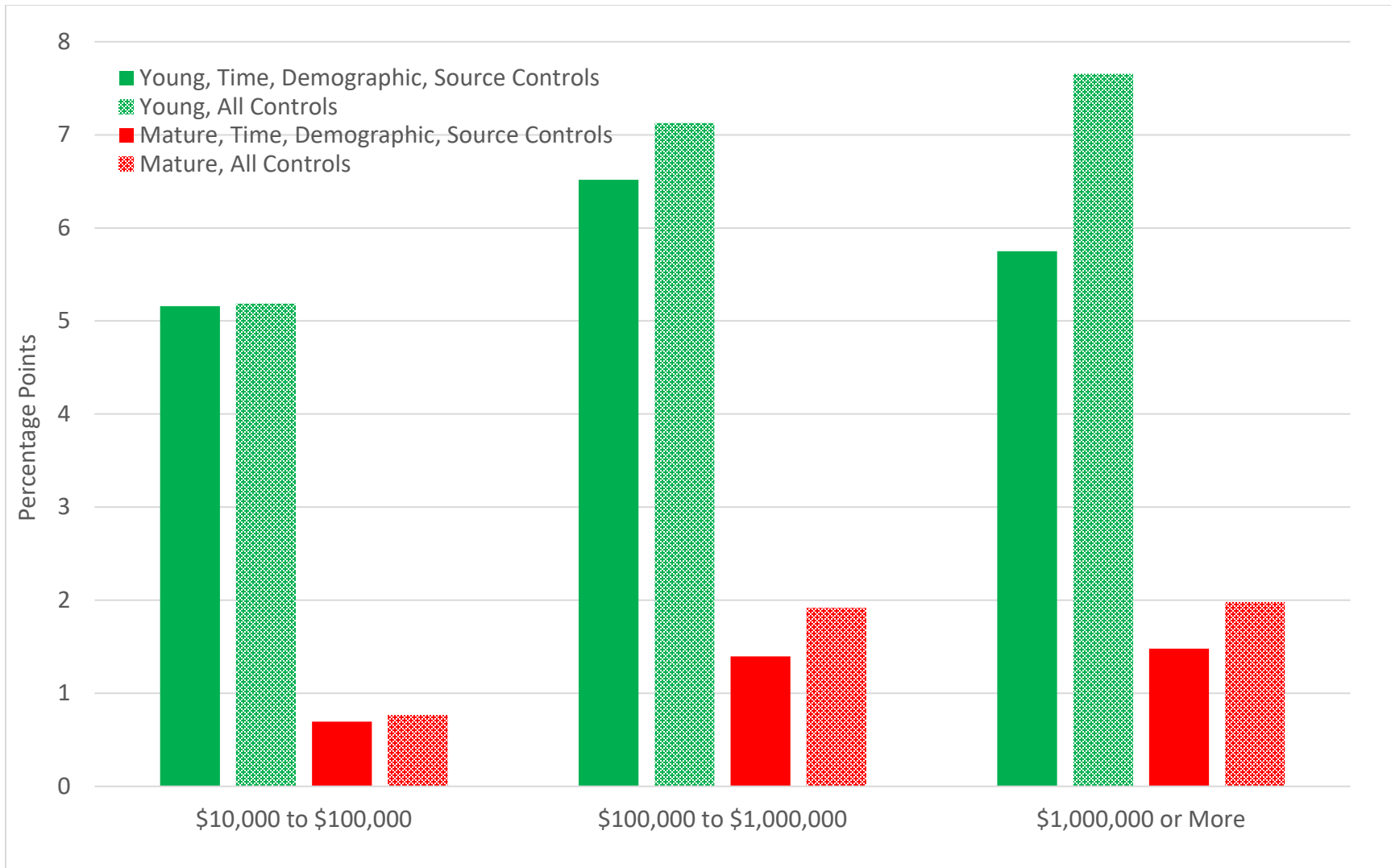
Notes: Tabulations by authors from LEOF data infrastructure using regression specification (1). High growth firms are in the top 25 percent of the employment-weighted four year growth distribution pooled over time. Young firms are age 1-5, mature firms 6+.

**Figure 9. Probability of Being a High-Growth Firm, Start-up Financing Source Coefficients**



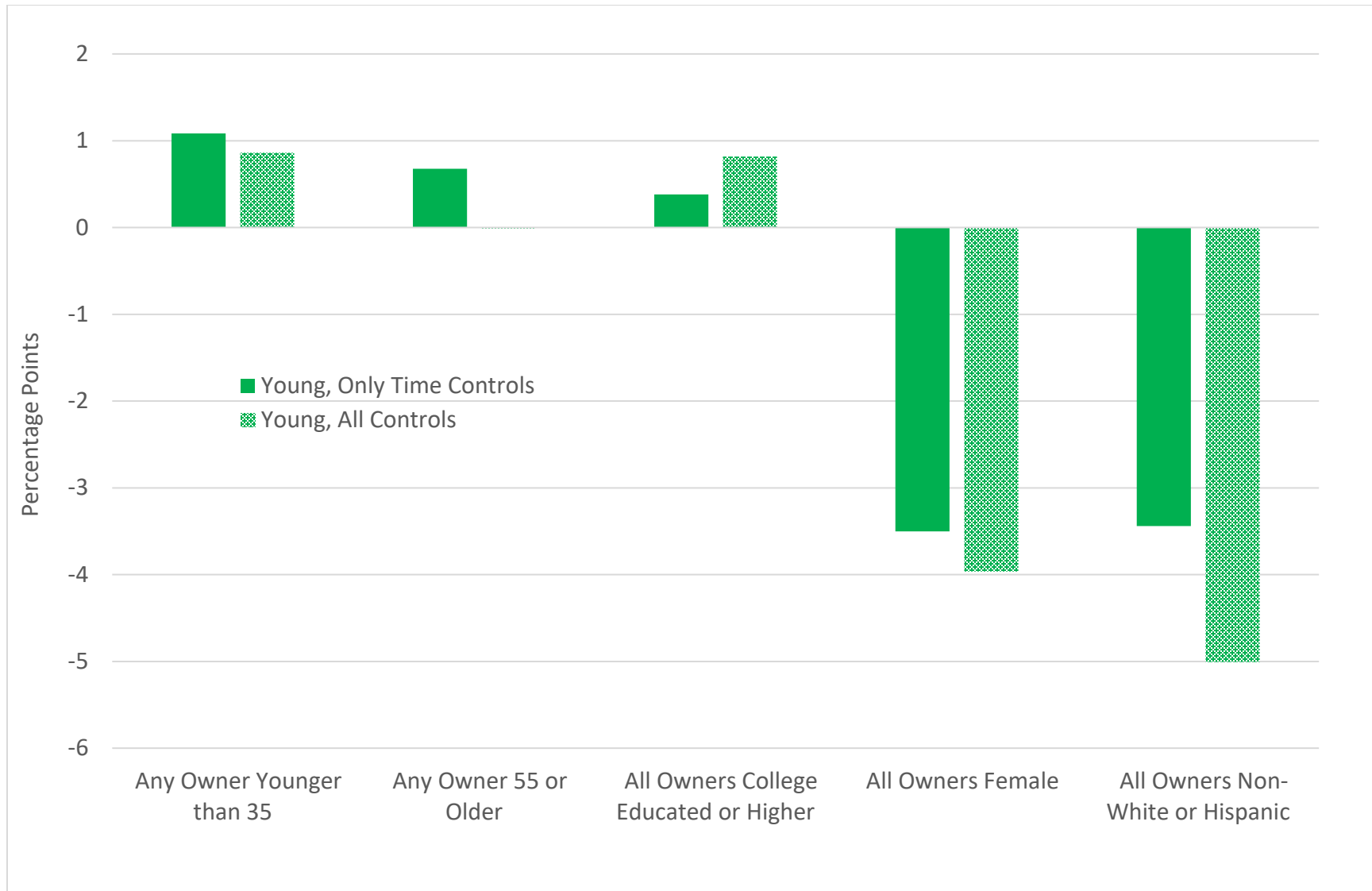
Notes: See notes to Figure 8.

**Figure 10. Probability of Being a High-Growth Firm, Start-up Financing Amount Coefficients**



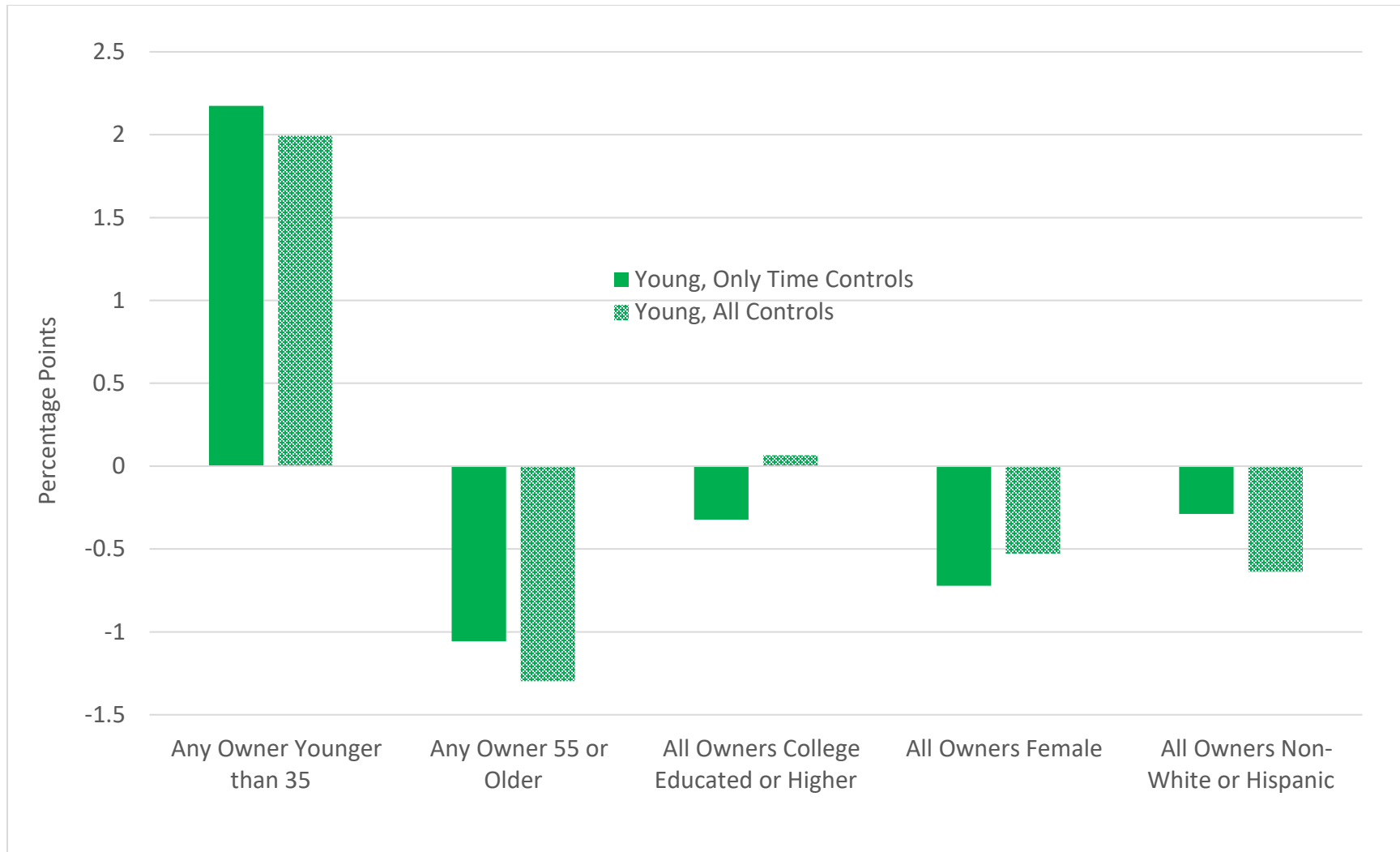
Notes: See notes to Figure 8.

**Figure 11. Probability of Having a Bank Loan at Start-up, Owner Characteristic Coefficients**



Notes: Tabulations by authors from LEOF data infrastructure. Young firms are firms age 1 to 5.

**Figure 12. Probability of Having an Outside Investor at Start-up, Owner Characteristic Coefficients**



Notes: See notes to Figure 11.

## Appendix A. Constructing and Validating LEOF

The Longitudinal Employer, Owner, and Financing (LEOF) data set combines Longitudinal Business Database (LBD) records with periodic Census business surveys conducted between (reference years) 2002 and 2017.<sup>25</sup> The linked LEOF records currently have (1) longitudinal firm employment (from the LBD) for all years 1982 through 2021, (2) cross section geographic and industry identifiers for the firm in the first year of positive employment, (3) cross-section owner demographics and other firm characteristics for the year in which the firm participated in a periodic survey, and (4) look-back start-up financing source and financing amount variables for the year in which the firm was founded. The LEOF universe for each of the periodic surveys is employer firms with positive employment in the year for which the survey sample is drawn, and LEOF records are weighted to match the corresponding LBD population in that year.

In addition to the differences in survey variables (Table 1) the periodic surveys also differ in terms of sampling strategy, data collection, and sample size, thus each linked survey file is reweighted to match the LBD universe by firm age, firm size, and two-digit NAICS. The largest exclusion is dropping firms with 10,000 or more employees from the LBD universe before reweighting, because the periodic finance surveys have very few firms in that size class. Most of these very large firms in the LBD would likely fail the “at least one owner with 10 percent ownership” rule at the beginning of each survey. From the other perspective, there are virtually no survey firms that do not have an LBD counterpart (see row 4 of Table 1).

The LEOF inverse weights are estimated using a logistic regression that varies by year. The outcome variable is 1 if the LBD firm has a match to a survey firm and is thus included in the LEOF sample. The control variables in the weighting regression include a dummy for firm age (young versus mature), an initial employment size class dummy (1, 2 to 4, 5 to 9, 10 to 19, 20 to 49, 50 to 99, and 100 or more), a secondary employment size variable that splits the largest

---

<sup>25</sup> The emphasis here is on “reference years” when describing the surveys, but it is important to note that the reference year and sampling or base year can be different. Due to the timing of updates to the business register and survey collections, it seems most appropriate to link the ASE and ABS to the LBD for the year prior to the survey’s reference year. Thus, the ASE base years are 2013, 2014, and 2015, while the base year is 2016 for the ABS with reference year 2017.

(100 or more group) into subgroup dummies (100-249, 250-499, 500-999, 1000-2499, 2500-4999, 5000-9999) and also interacts values within those groups with linear splines, NAICS 2 industry dummies. The specifications have fully interacted industry\*first size class dummies, fully interacted young\*second size class dummies, and the spline variables. We check that the counts of weighted firms and employment by industry and size class are in generally in line with the LBD universe.

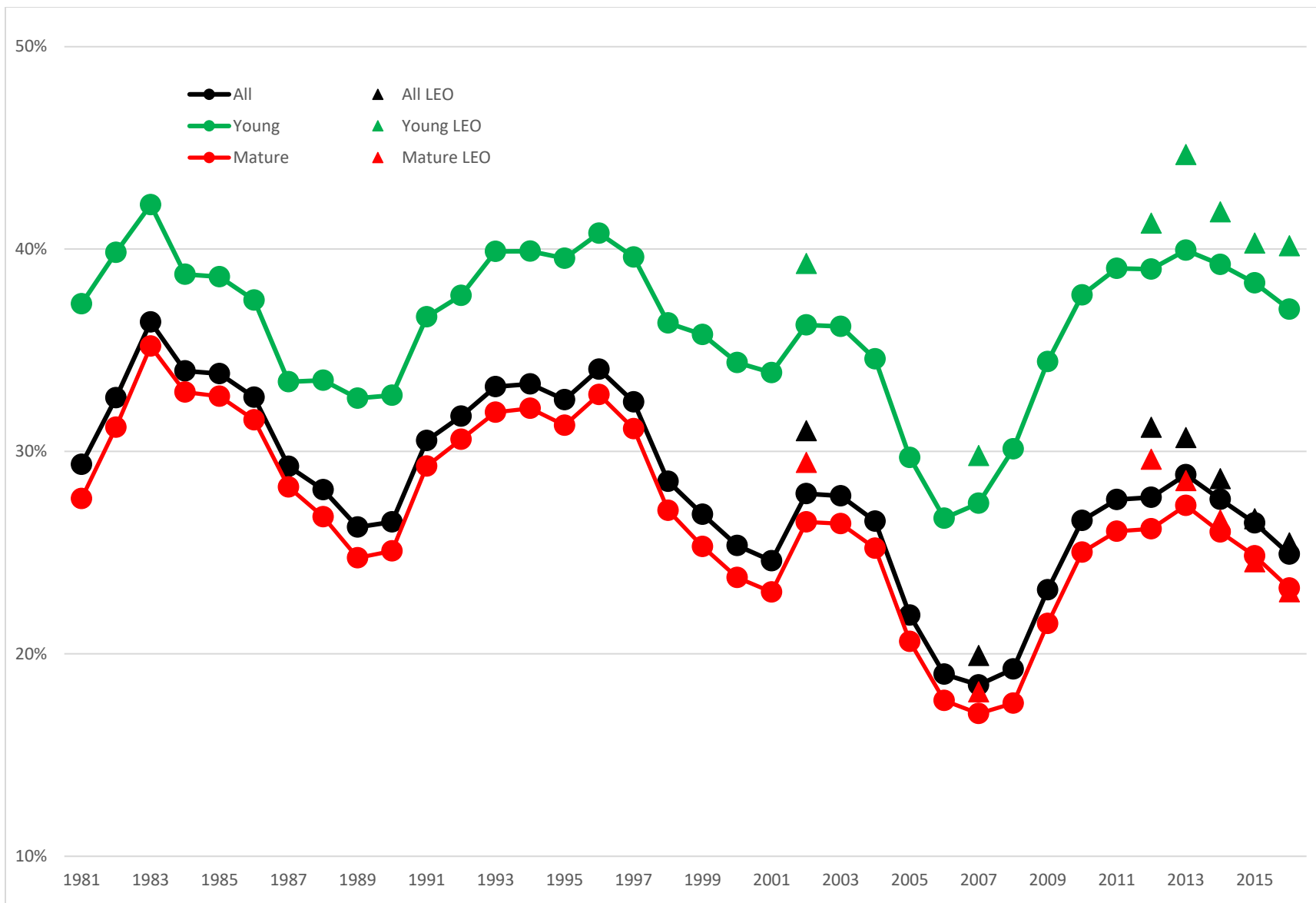
LEOF inverse weights are generated to represent the LBD universe in the base year for each survey sample, but there is some variability around the LBD targets along the various dimensions because the samples are limited in size. In addition, the outcomes of interest in this paper generally involve employment weighted outcomes, where the employment weights are DHS averages of base year employment and employment in year  $t+4$ . As such, there are multiple reasons why weighted LEOF sample outcomes may diverge from estimates generated using the LBD universe.

Given the sampling and weighting concerns, it is reassuring to see a close alignment between weighted LEOF and LBD universe outcomes along two key dimensions. The first is the fraction of employment-weighted firms in the top 25 percent “high-growth” category (Figure A1), and the second is the average growth rate of high-growth firms (Figure A2). Both charts have lines with round markers that capture the LBD universe outcomes for the entire sample period, base years 1981 through 2016.

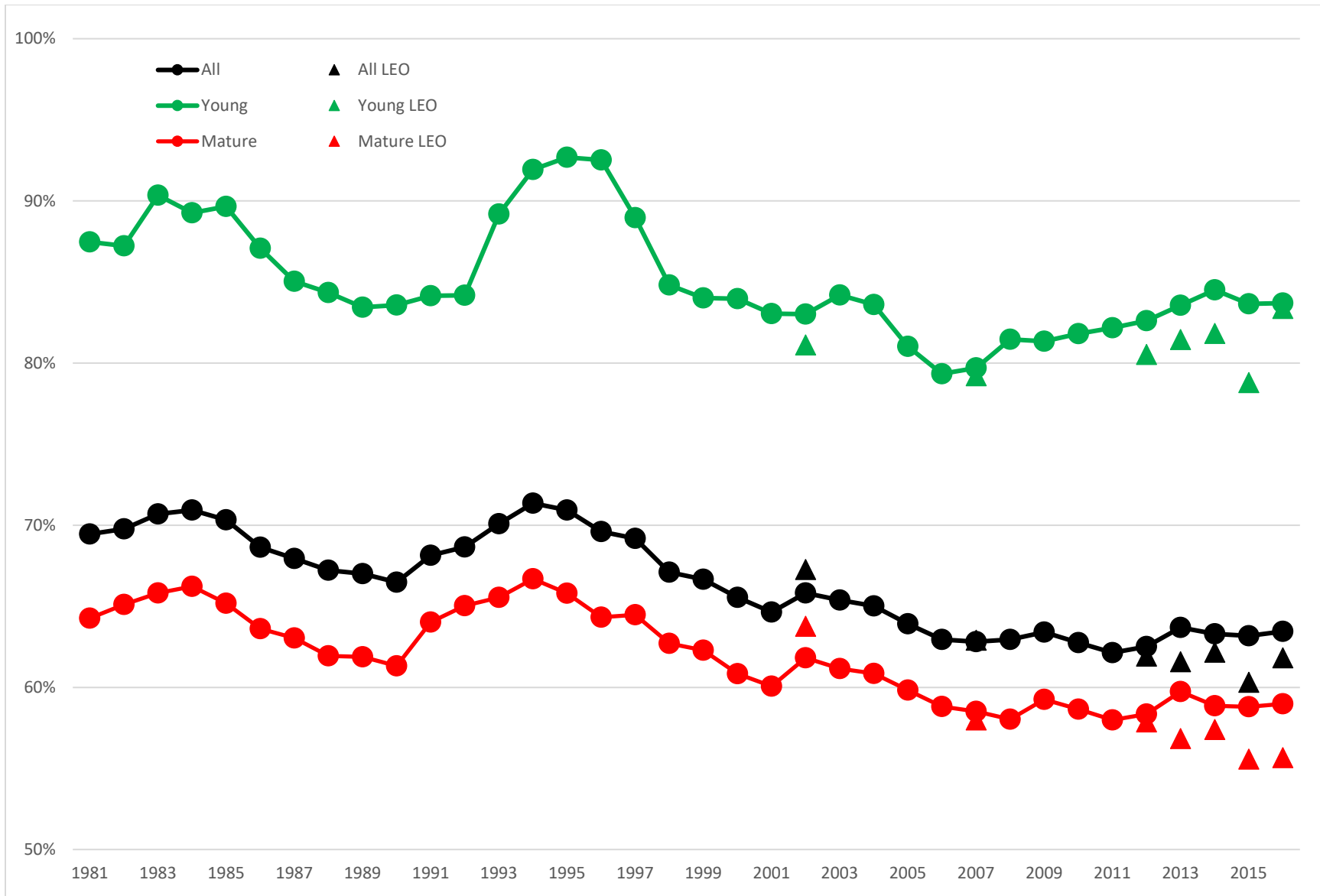
Both charts also show additional markers (triangles) for the LEOF samples in the respective base years (2002, 2007, 2012, 2013, 2014, 2015, and 2016). The charts show a slight bias (more LEOF firms are high growth) that is to be expected because the inverse weighting does not control for survival over the four years, and LEOF firms are slightly more likely to survive after controlling for the factors that were used in the inverse weights (firms that go out of business are less likely to answer the survey). However, the bias is clearly small relative to differences between young and mature firms at any point in time, and LEOF also captures the time series patterns over time.



**Figure A1. Fraction of Firms by Age that Are High Growth, LBD and LEOF (Employment Weighted)**



**Figure A2. Average Growth Rates for High Growth Firms, LBD and LEOF (Employment Weighted)**



## Appendix B. Detailed Regression Results

This appendix reports the detailed regression results underlying Figures 5 through 9 in the main text. The regressions are all linear probability models, with observations weighted using the DHS denominator, which is the average of employment in time  $t$  and time  $t+4$ . Tables B1 through B4 show regression results for the probability of being a high-growth firm, while B5 shows regression results for the probability of having bank loans and outside investors in reported start-up financing sources.

Tables B1 through B4 are structured the same, incrementally expanding the list of potential correlates from just time dummies (column 1) to adding owner demographic characteristics (column 2), then start-up financing sources and amounts (column 3), then LBD-based employment size and firm age dummies (column 4), and finally LBD-based two-digit industry dummies (column 5). Tables B1 and B3 are for young firms, while tables B2 and B4 are for mature firms. The only difference between Tables B1/B2 and B3/B4 is the inclusion of start-up financing amounts in the later pair, which requires restricting the sample to 2007 and later.

Table B5 shows three regressions each for the probability of having a bank loan at start-up (columns 1 to 3) and an outside investor (columns 4 to 6). The progression across columns is like the high-growth regressions, with the first column in each time dummies only, the second column adding owner characteristics, and the third column adding LBD-based employment size, firm age dummies, and two-digit industry.

**Table B1. Probability of Being in Top 25 Percent of Firm Growth Distribution, 2002-2016, Young Firms**

Notes: Young firms only, regressions are firm-weighted, year=2007 is omitted category.

	(1)	(2)	(3)	(4)	(5)
<b>Independent Variables</b>					
Year = 2002	5.903*** (0.142)	5.725*** (0.142)	5.914*** (0.146)	5.984*** (0.146)	5.945*** (0.146)
Year = 2012	7.201*** (0.212)	7.668*** (0.213)	7.965*** (0.213)	8.080*** (0.213)	8.088*** (0.213)
Year = 2013	8.203*** (0.235)	8.866*** (0.235)	9.070*** (0.236)	9.090*** (0.235)	9.097*** (0.235)
Year = 2014	8.619*** (0.245)	9.341*** (0.245)	9.579*** (0.245)	9.606*** (0.245)	9.624*** (0.244)
Year = 2015	7.498*** (0.245)	8.309*** (0.245)	8.660*** (0.246)	8.734*** (0.246)	8.795*** (0.245)
Year = 2016	8.446*** (0.179)	9.212*** (0.179)	9.606*** (0.180)	9.691*** (0.180)	9.778*** (0.180)
Any Owner Younger than 35		5.928*** (0.188)	5.814*** (0.188)	5.479*** (0.188)	5.444*** (0.188)
Any Owner 55 or Older		-5.500*** (0.139)	-5.401*** (0.140)	-5.178*** (0.140)	-4.980*** (0.140)
All Owners College Educated or Higher		1.212*** (0.131)	1.298*** (0.131)	1.340*** (0.131)	1.487*** (0.145)
All Owners Female		-3.296*** (0.171)	-3.173*** (0.171)	-3.356*** (0.172)	-3.507*** (0.175)
All Owners Non-White or Hispanic		-0.988*** (0.172)	-0.851*** (0.171)	-0.923*** (0.171)	-1.112*** (0.174)
Firm Had Bank Loan at Start Up			3.462*** (0.177)	3.798*** (0.180)	3.498*** (0.182)
Firm Had Home Equity Loan at Start Up			1.862*** (0.287)	1.997*** (0.287)	1.922*** (0.287)
Firm Used Credit Cards to Finance Start Up			1.969*** (0.195)	1.800*** (0.195)	1.657*** (0.195)
Firm Had Outside Investor at Start Up			3.757*** (0.309)	3.878*** (0.309)	3.881*** (0.309)
<b>Additional Controls</b>					
Interacted Employment Size and Age Class Dummies				Yes	Yes
Two Digit NAICS Industry Dummies					Yes
<b>Observations</b>	657,000	657,000	657,000	657,000	657,000
<b>R-Squared</b>	0.005	0.013	0.015	0.017	0.020

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

CBDRB-FY23-CES014-050

**Table B2. Probability of Being in Top 25 Percent of Firm Growth Distribution, 2002-2016, Mature Firms**

Notes: Mature firms, regressions are firm-weighted, year=2007 is omitted category.

	(1)	(2)	(3)	(4)	(5)
<b>Independent Variables</b>					
Year = 2002	5.624*** (0.09)	5.179*** (0.09)	5.317*** (0.09)	5.249*** (0.09)	5.297*** (0.09)
Year = 2012	5.518*** (0.12)	6.011*** (0.12)	5.983*** (0.12)	5.904*** (0.12)	5.892*** (0.12)
Year = 2013	5.969*** (0.13)	6.598*** (0.13)	6.474*** (0.13)	6.330*** (0.13)	6.340*** (0.13)
Year = 2014	5.183*** (0.13)	5.952*** (0.13)	5.828*** (0.13)	5.700*** (0.13)	5.698*** (0.13)
Year = 2015	4.215*** (0.13)	5.074*** (0.13)	4.992*** (0.13)	4.896*** (0.13)	4.878*** (0.13)
Year = 2016	3.786*** (0.09)	4.695*** (0.09)	4.633*** (0.09)	4.625*** (0.09)	4.592*** (0.09)
Any Owner Younger than 35		5.631*** (0.20)	5.620*** (0.20)	5.414*** (0.20)	5.297*** (0.20)
Any Owner 55 or Older		-6.348*** (0.08)	-6.212*** (0.08)	-5.640*** (0.08)	-5.471*** (0.08)
All Owners College Educated or Higher		-0.936*** (0.07)	-0.866*** (0.07)	-0.840*** (0.07)	0.338*** (0.08)
All Owners Female		-2.229*** (0.11)	-2.250*** (0.11)	-2.571*** (0.11)	-1.956*** (0.11)
All Owners Non-White or Hispanic		0.624*** (0.12)	0.574*** (0.12)	0.19 (0.12)	0.363*** (0.12)
Firm Had Bank Loan at Start Up			0.174* (0.09)	0.571*** (0.09)	0.969*** (0.09)
Firm Had Home Equity Loan at Start Up			1.739*** (0.18)	1.700*** (0.18)	1.737*** (0.18)
Firm Used Credit Cards to Finance Start Up			2.238*** (0.14)	1.784*** (0.14)	1.633*** (0.14)
Firm Had Outside Investor at Start Up			0.655*** (0.17)	1.021*** (0.17)	1.143*** (0.17)
<b>Additional Controls</b>					
Interacted Employment Size and Age Class Dummies				Yes	Yes
Two Digit NAICS Industry Dummies					Yes
<b>Observations</b>	1,742,000	1,742,000	1,742,000	1,742,000	1,742,000
<b>R-Squared</b>	0.003	0.011	0.012	0.015	0.018

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

CBDRB-FY23-CES014-050

**Table B3. Probability of Being in Top 25 Percent of Firm Growth Distribution, 2007-2016, Young Firms**

Notes: Young firms only, regressions are firm-weighted, year=2007 is omitted category.

	(1)	(2)	(3)	(4)	(5)
<b>Independent Variables</b>					
Year = 2012	7.201*** (0.21)	7.672*** (0.21)	8.039*** (0.21)	8.168*** (0.21)	8.206*** (0.21)
Year = 2013	8.203*** (0.24)	8.877*** (0.24)	8.984*** (0.24)	9.010*** (0.24)	9.047*** (0.24)
Year = 2014	8.619*** (0.25)	9.354*** (0.25)	9.469*** (0.25)	9.501*** (0.25)	9.551*** (0.24)
Year = 2015	7.498*** (0.25)	8.323*** (0.25)	8.590*** (0.25)	8.673*** (0.25)	8.765*** (0.25)
Year = 2016	8.446*** (0.18)	9.228*** (0.18)	9.772*** (0.18)	9.877*** (0.18)	9.996*** (0.18)
Any Owner Younger than 35		6.131*** (0.22)	5.984*** (0.22)	5.662*** (0.22)	5.652*** (0.22)
Any Owner 55 or Older		-5.628*** (0.16)	-5.748*** (0.16)	-5.537*** (0.16)	-5.324*** (0.16)
All Owners College Educated or Higher		1.128*** (0.15)	1.136*** (0.15)	1.145*** (0.15)	1.152*** (0.16)
All Owners Female		-3.115*** (0.19)	-2.592*** (0.19)	-2.794*** (0.19)	-2.932*** (0.20)
All Owners Non-White or Hispanic		-1.079*** (0.19)	-1.281*** (0.19)	-1.359*** (0.19)	-1.428*** (0.19)
Firm Had Bank Loan at Start Up			1.734*** (0.22)	2.066*** (0.22)	1.806*** (0.22)
Firm Had Home Equity Loan at Start Up			0.11 (0.30)	0.20 (0.29)	0.18 (0.29)
Firm Used Credit Cards to Finance Start Up			1.411*** (0.22)	1.235*** (0.22)	1.065*** (0.22)
Firm Had Outside Investor at Start Up			2.267*** (0.35)	2.359*** (0.35)	2.406*** (0.35)
Firm Start Up Capital Between \$10,000 and \$100,000			5.158*** (0.18)	5.134*** (0.18)	5.186*** (0.18)
Firm Start Up Capital Between \$100,000 and \$1,000,000			6.518*** (0.22)	6.951*** (0.23)	7.126*** (0.23)
Firm Start Up Capital \$1,000,000 or More			5.749*** (0.46)	6.972*** (0.47)	7.655*** (0.47)
<b>Additional Controls</b>					
Interacted Employment Size and Age Class Dummies				Yes	Yes
Two Digit NAICS Industry Dummies					Yes
<b>Observations</b>	459,000	459,000	459,000	459,000	459,000
<b>R-Squared</b>	0.006	0.014	0.02	0.022	0.026

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

CBDRB-FY23-CES014-050

**Table B4. Probability of Being in Top 25 Percent of Firm Growth Distribution, 2007-2016, Mature Firms**

Notes: Mature firms, regressions are firm-weighted, year=2007 is omitted category.

	(1)	(2)	(3)	(4)	(5)
<b>Independent Variables</b>					
Year = 2012	5.518*** (0.116)	6.028*** (0.116)	6.011*** (0.116)	5.927*** (0.116)	5.919*** (0.116)
Year = 2013	5.969*** (0.129)	6.620*** (0.129)	6.471*** (0.129)	6.322*** (0.129)	6.325*** (0.129)
Year = 2014	5.183*** (0.132)	5.979*** (0.132)	5.822*** (0.132)	5.688*** (0.132)	5.677*** (0.132)
Year = 2015	4.215*** (0.132)	5.103*** (0.132)	4.990*** (0.132)	4.889*** (0.132)	4.861*** (0.132)
Year = 2016	3.786*** (0.0935)	4.726*** (0.0944)	4.666*** (0.0946)	4.651*** (0.0944)	4.615*** (0.0944)
Any Owner Younger than 35		5.749*** (0.237)	5.709*** (0.237)	5.520*** (0.237)	5.414*** (0.236)
Any Owner 55 or Older		-6.557*** (0.0889)	-6.385*** (0.0892)	-5.792*** (0.0912)	-5.588*** (0.0914)
All Owners College Educated or Higher		-1.045*** (0.0813)	-1.020*** (0.0814)	-1.006*** (0.0812)	0.198** (0.0946)
All Owners Female		-2.243*** (0.120)	-2.218*** (0.120)	-2.531*** (0.121)	-1.879*** (0.122)
All Owners Non-White or Hispanic		0.597*** (0.130)	0.447*** (0.130)	0.0608 (0.130)	0.214 (0.132)
Firm Had Bank Loan at Start Up			-0.316*** (0.108)	0.0485 (0.108)	0.373*** (0.109)
Firm Had Home Equity Loan at Start Up			1.379*** (0.181)	1.319*** (0.181)	1.272*** (0.181)
Firm Used Credit Cards to Finance Start Up			2.237*** (0.156)	1.806*** (0.156)	1.619*** (0.156)
Firm Had Outside Investor at Start Up			0.319* (0.193)	0.695*** (0.193)	0.745*** (0.193)
Firm Start Up Capital Between \$10,000 and \$100,000			0.694*** (0.0974)	0.569*** (0.0975)	0.767*** (0.0977)
Firm Start Up Capital Between \$100,000 and \$1,000,000			1.395*** (0.126)	1.566*** (0.127)	1.916*** (0.129)
Firm Start Up Capital \$1,000,000 or More			1.478*** (0.277)	1.728*** (0.281)	1.980*** (0.282)
<b>Additional Controls</b>					
Interacted Employment Size and Age Class Dummies				Yes	Yes
Two Digit NAICS Industry Dummies					Yes
<b>Observations</b>	1,269,000	1,269,000	1,269,000	1,269,000	1,269,000
<b>R-Squared</b>	0.003	0.012	0.013	0.016	0.019

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B5. Probability of Having a Given Start Up Finance Source, 2002 to 2016, Young Firms**

Notes: Young firms only, regressions are firm-weighted, year=2007 is omitted category.

Independent Variables	Bank Loan			Outside Investor		
	(1)	(2)	(3)	(1)	(2)	(3)
Year = 2002	3.714*** (0.145)	3.573*** (0.145)	3.580*** (0.141)	-0.843*** (0.0798)	-0.889*** (0.0800)	-0.958*** (0.0798)
Year = 2012	-4.893*** (0.186)	-4.479*** (0.187)	-4.714*** (0.182)	-0.306*** (0.117)	-0.171 (0.117)	-0.237** (0.117)
Year = 2013	-4.009*** (0.203)	-3.740*** (0.203)	-3.858*** (0.199)	0.260** (0.132)	0.444*** (0.132)	0.394*** (0.131)
Year = 2014	-4.281*** (0.208)	-4.020*** (0.209)	-4.114*** (0.204)	-0.0344 (0.134)	0.169 (0.134)	0.110 (0.134)
Year = 2015	-5.429*** (0.205)	-5.118*** (0.206)	-5.210*** (0.201)	-0.554*** (0.130)	-0.330** (0.131)	-0.379*** (0.131)
Year = 2016	-5.612*** (0.157)	-5.299*** (0.157)	-5.429*** (0.154)	-0.986*** (0.0942)	-0.764*** (0.0948)	-0.826*** (0.0945)
Any Owner Younger than 35		1.085*** (0.151)	0.860*** (0.149)		2.174*** (0.108)	1.994*** (0.108)
Any Owner 55 or Older		0.677*** (0.121)	-0.0112 (0.119)		-1.056*** (0.0731)	-1.297*** (0.0736)
All Owners College Educated or Higher		0.382*** (0.110)	0.817*** (0.117)		-0.322*** (0.0697)	0.0650 (0.0783)
All Owners Female		-3.501*** (0.137)	-3.964*** (0.139)		-0.722*** (0.0912)	-0.528*** (0.0933)
All Owners Non-White or Hispanic		-3.440*** (0.134)	-5.007*** (0.135)		-0.288*** (0.0917)	-0.637*** (0.0929)
<b>Additional Controls</b>						
Interacted Employment Size and Age Class Dummies			Yes			Yes
Two Digit NAICS Industry Dummies			Yes			Yes
<b>Observations</b>	657,000	657,000	657,000	657,000	657,000	657,000
<b>R-Squared</b>	0.008	0.011	0.060	0.000	0.003	0.012
Robust standard errors in parentheses , colle						

CBDRB-FY23-CES014-050