High-Growth Entrepreneurship

J. David Brown[a,c]
John S. Earle[b,c]
Mee Jung Kim[a,b]
Kyung Min Lee[b]

December 2017

Abstract
Who founds high-growth firms? We study the patterns and determinants of job creation for a large cohort of start-ups. Data for all U.S. employers show strong persistence in employment size from firm birth to age seven, with a small fraction of firms accounting for most employment at both ages, patterns that are little explained by either finely disaggregated industry controls or amount of finance. We link these data to characteristics of 55,800 founders of 37,100 start-ups, and define "high growth" as the top 5% of firms in the size distribution at age zero and seven. Firms founded only by females are 34% less likely to be high-growth entrepreneurships at both ages. A similar gap for African-Americans at start-up disappears by age seven. Differences with respect to other races, ethnicity, and nativity are modest. Founder age is positively associated with high growth probability at birth, but the profile flattens after seven years and even becomes slightly negative. The education profile is initially concave, with graduate degree recipients no more likely to found high growth firms than high school graduates, but the former catch up to those with bachelor's degrees by firm age seven, while the latter do not. Most other relationships of high growth with founder characteristics are highly persistent over time. Prior business ownership is strongly positively associated, and veteran experience negatively associated, with high growth. A larger founding team raises the probability of high growth, while, controlling for team size, diversity (by gender, age, race/ethnicity, or nativity) either lowers the probability or has little effect. Controlling for start-up capital raises the high-growth probability of firms founded by women, minorities, immigrants, veterans, sole proprietors, and novice, younger, and less educated entrepreneurs. Perhaps surprisingly, female, minority, and less-educated entrepreneurs tend to choose high-growth industries, but fewer of them achieve high growth relative to their industry peers.

Acknowledgements and Disclaimer: We thank the National Science Foundation for support (Grants 1262269 and 1719201 to George Mason University), and Zoltan Acs, Nathan Goldschlag, Javier Miranda, and Martha Stinson for helpful comments. Any opinions and conclusions expressed herein are ours only and do not necessarily reflect the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information on individual firms is disclosed.

[a]Center for Economic Studies, U.S. Census Bureau
[b]Schar School for Policy and Government, George Mason University
[c]IZA – Institute for the Study of Labor

1. Introduction

Recent research exploiting the availability of large firm-level datasets has made great strides in understanding patterns of job creation by firm size and age in the U.S. Conclusions about the role of small versus large firms dating back to Birch (1979, 1981, 1987) have been amended to recognize the predominance of entrants and young firms in the job creation process (Haltiwanger, Jarmin, and Miranda 2013; Decker et al. 2014). At the same time, there is increasing recognition that most firms enter at a small size and remain small afterward (Shane 2008, Hurst and Pugsley 2011, 2017). While these empirical regularities may seem mutually inconsistent, they can be reconciled if firm growth, like firm size, is positively skewed, so that a small fraction of all entrants account for most employment growth, as shown by Cabral and Mata (2003), Decker et al. (2016), and others.[1]

The importance of this high-growth entrepreneurship is widely recognized, yet many open questions remain. In this paper, we study two sets of such questions. The first set concerns the basic patterns of size at entry and subsequent growth. Do high-growth entrepreneurships begin operations already at an unusually large size, or are they initially indistinguishable from other entrants and only become large after several years of rapid growth? To what extent is heterogeneity in start-up size and growth accounted for by the industries in which firms operate, and to what extent by the availability of finance? Much of the literature has focused on differences across industries and on financial access, but even after controlling for narrow industries and financial access, we find substantial heterogeneity in firm size both at start-up and subsequently.

Motivated by these findings, we consider a second set of questions about the characteristics of the owner-founders of high-growth entrants. Do the founders of these firms differ from others by demographic characteristics such as age, gender, race/ethnicity, and citizen/immigrant status? Does the human capital of high-growth entrepreneurs differ in terms of education, general labor market experience, veteran status, and prior entrepreneurial experience, compared to owners of low-growth firms? Are the founding teams of high-growth entrepreneurships larger, and to what extent do they involve family members versus unrelated individuals or with more diverse founding teams, defined by age, gender, and race/ethnicity? Do these patterns vary if the amount of start-up finance and specific industry choice are taken into account? Finally, how persistent are the impacts of start-up characteristics on the probability of high growth as the firm ages?

We address these questions following a large cohort of firms from their initial entry and analyzing a data set that is larger, richer, and more representative compared with those studied in previous research. Rather than study incumbent firms that have already attained some size, as is common in previous research, our approach to measuring high growth entrepreneurship is to track entrants from the first quarter in which they hire an employee and analyze the determinants of the top five percent in employment size at age zero (their entry quarter) and at age seven (28 quarters later). Our analysis in each case thus compares firms at exactly the same age, focusing on the start-up period through age seven. We avoid any conditioning on prior growth, and at age 7 we treat early growth and later "catch-up" equivalently: all jobs created by firms from their initial entry are counted, rather than excluding those created at start-up or through some later age.

[1] There are extensive literatures on all these issues. A short list of additional references would include Evans (1987), Brown, Hamilton, and Medoff (1990), Dunne and Hughes (1994), Geroski (1995), Davis, Haltiwanger, and Schuh (1996), Neumark, Wall, and Zhang (2011), and Acemoglu et al. (2013).

The data we analyze include the Business Register (BR) and Longitudinal Business Database (LBD), covering the universe of U.S. private employers, for analyzing the patterns of entry and growth. In order to incorporate founder characteristics we focus on a particular entry cohort and link to the 2007 Survey of Business Owners (SBO), resulting in about 37,100 observations on start-up firms and about 55,800 on founders. The rich set of founder characteristics in these data goes beyond the basics of age and years of schooling considered in most studies, such as Cabral and Mata (2003). We add gender, detailed race/ethnicity, type of schooling, and other aspects of human capital: veteran status, citizen/immigrant, and previous entrepreneurship. Exploiting detailed information on up to four owners of each firm, we also study the size and composition of founding teams of entrepreneurs, including the extent to which diversity is correlated with high growth. Linking to the BR and LBD permits us to track this 2007 entry cohort until age 7 in 2014, the last year available in the data. To check whether results are sensitive to the macroeconomic environment, we also carry out an analysis of the 2012 entry cohort, as one of many robustness checks.

Our paper builds on several strands of previous research. Haltiwanger, Jarmin, and Miranda (2013) report that the only age group with substantial positive net job creation is the age zero entrants.[2] We add to this finding in several ways. Unlike their examination of mean differences by age and size, we focus on the right tail of the employment (growth) distribution, we consider the same set of firms at different ages and compare job creation among them, and we investigate whether entrants in the largest size category tend to remain large and the extent to which they continue to grow. In addition to looking beyond the mean and in tracking an entry cohort, we also analyze the association of high job creation with a rich set of founder characteristics.

Our focus in this paper is job creation during firms' initial, entrepreneurial period, but the analysis is related to previous research on "high-growth firms." Most of this research is essentially cross-sectional in comparing firms without regard to age or stage of life cycle. Many studies follow the "Eurostat-OECD" definition, which examines a short panel of incumbent firms, conditions on an initially observed employment (or sales) level, and then measures growth for a few years thereafter (e.g., Eurostat-OECD 2008; OECD 2010). As documented by Acs, Parsons, and Tracy (2008), Holzl (2014), and Daunfeld and Halvarsson (2015), however, within-firm growth is highly volatile, and a "high-growth" period by this definition is frequently preceded and followed by periods of low or negative growth. The typical practice of excluding all firms with initial size below some threshold (usually, 10 employees, as in the Eurostat-OECD definition) effectively ignores most firms in any economy. Our approach also avoids the common dilemma in these studies of whether to measure growth in absolute terms (which favors firms that are large at the beginning of the growth period) or relative terms (which favors firms that are initially smaller), because we measure from the entry date.[3]

[2] Employment among entrants in their 2005 data is 3.5 million, all of which is job creation by definition. The only other age group with positive net job creation is over 25 years old. They report net job creation for this latter group at 400,000, which can be compared to their total employment of 6.9 million that same year (U.S. Census Bureau, 2016).
[3] Decker et al. (2014) use a high-growth definition as employment increase over 25 percent, Stangler (2010) examines the top one and five percent, and Storey (1994) the top four percent in a particular year and without regard to age. The "Birch Index" attempts to resolve the absolute versus relative dilemma by taking the product of the two (Birch, 1987).

Concerning the determinants of high growth, a long literature in management and related disciplines has examined some characteristics (e.g., Kalleberg and Leicht, 1991). Typically these are cross-sectional analyses of small samples, however, and in many cases they study incumbents and take no account of firm age. Within economics, most studies of firm growth focus on the mean, as in Neumark, Wall, and Zhang (2011) and Haltiwanger, Jarmin, and Miranda (2013). Fairlie and Robb (2007, 2009) and Robb and Fairlie (2009) analyze the impact of race, gender, and family history using the 1992 Characteristics of Business Owners (CBO), the predecessor of the SBO, but they do not observe employment level or growth in these data. Their analysis estimates cross-sectional differences in mean sales, survival, and the probability of hiring at least one employee, without distinguishing by firm age, all of which differ from our focus on high-growth entrepreneurship among entrants.[4] The economics literature also tends to focus on industry characteristics rather than firms and founders.

Previous research studying a cohort of entrants includes Cabral and Mata's (2003) analysis of Portuguese firms and studies of the Kauffman Firm Survey (KFS) in the U.S. Cabral and Mata (2003) study entrants at age 0 and 7, as do we, but they do not focus on the high-growth group. They condition on survival to age 7, so that firms exiting before age 7 are not in their age 0 analysis. Their sample is restricted to manufacturing, the sample size is 515 firms, and the only founder characteristics in their data are age and education. The KFS is a cohort and it includes rich information on founders, but the samples are very different in size and composition from ours: the KFS includes fewer than 5,000 entrants, and the sample is drawn from a list of Dun and Bradstreet firms, which is more likely to include firms that already had some credit history, unlike our data, where inclusion is based on reporting payroll employment to the Internal Revenue Service. The Dun and Bradstreet data do not distinguish employers from nonemployers, because owners are included in the employment count whether or not they are on the payroll, and some of the KFS firms could thus be nonemployers. The KFS sample also includes purchases of existing businesses, and purchases of franchises, which we exclude.[5] In general, previous research on high growth has been unable to study entry cohorts because of sample sizes that are too small to permit reliable estimation for the few firms of the cohort experiencing high growth.

Similar to our research in data sources are Jarmin and Krizan (2010) and Jarmin, Krizan, and Luque (2014) who link data on firm characteristics from the 2002 SBO to a longitudinal data source on employment (LBD).[6] They focus on mean, not high, growth determinants, and they analyze the cross-section of all firms, rather than an entry cohort. The data do not permit them to study several important issues including immigrant status, husband-wife ownership, prior business ownership, and amount of start-up capital, which we are able to address with the 2007 SBO.

[4] Bates (1990) examines firm survival using the 1982 CBO data.

[5] In a study of mean employment growth in immigrant-owned firms, Kerr and Kerr (2017) also follow cohorts of entrants, but they exclude age 0 job creation from their measure of employment growth (and their regression estimates control for age 0 employment). In another type of study, Garnsey, Stam, and Heffernan (2006) examine growth patterns in cohorts including about 400 firms, and Brown, Earle, and Lup (2006) examine growth determinants in a similarly sized sample; in both cases the samples are non-randomly selected.

[6] Fairlie and Robb (2007, 2009) and Robb and Fairlie (2009) use the CBO, but they do not link it to employment data, so they study only if a firm is an employer at the time of the survey. By contrast, Jarmin and Krizan (2010), Jarmin, Krizan, and Luque (2014), and our paper use data on the number of employees.

Despite these substantial differences between our approach and the previous research, we discuss some of the key results from this literature to provide context when we report our findings below. In one way, however, our aim in this paper is similar to the previous research in establishing important empirical regularities that may be useful for theory and policy, but not to test an explicit model. The results are related to some theories, however. Our finding of large heterogeneity in firm size at entry, even within narrowly defined industries, is inconsistent with standard models of industry dynamics going back to Jovanovic (1982) and Hopenhayn (1992) and extending to Melitz (2003) etc., which have all entrants choosing the same optimal size (Frank 1988 is an exception). Our result that entrant size heterogeneity declines when start-up capital is taken into account is suggestive that varying financial constraints may account for some of the size heterogeneity. On the other hand, our finding of high persistence of size from age zero to age seven suggests some strong underlying heterogeneity in firm potential and possibly in founder motivations that deserves further research. We find the "up-or-out" dynamic, a productivity-enhancing mechanism frequently posited in discussions of industry dynamics (e.g., Decker et al. 2014), is strongest among the largest entrants, and is weaker in smaller start-ups.

Our results also relate to labor market theories of human capital, discrimination, occupational choice, and complementarities. The finding that more education is not uniformly valuable in raising firm success (defined by employment size) challenges single-factor models of human capital, and suggests instead that multiple dimensions of skill are relevant. The finding of lower prevalence of women and minorities among high-growth entrepreneurs could be consistent with theories of either discrimination or self-selection into occupations, but the result that these differences are diminished when start-up capital is taken into account suggests possible discrimination in financial markets. A further result that the gaps are larger within narrow industries implies, contrary to the possibility that women and minorities choose unambitious fields in which to open businesses, that in fact they choose high-growth sectors, but their performance is worse (in the sense of firm size) within sector.

Finally, the research in this paper is relevant to theories of complementarity and diversity within teams. Are larger teams more likely to found high-growth firms (Ruef, Aldrich, and Carter 2003)? What kinds of skills and characteristics combine to promote growth? Lazear (2005) has posited the desirability of "balanced skills" for individual entrepreneurs, but perhaps the balance can be achieved with a diverse team. Again, these issues have previously been studied at the mean for a cross-section of incumbents, while our focus in this paper is on high growth among start-ups.

The remainder of the paper is organized as follows. The next section describes our data and measurement approach. The following section contains results. First, we describe the empirical regularities of size at entry and at age 7, and the transitions between size categories with age. Then we provide estimates of the impact of the founder characteristics, start-up finance, and narrow industry on the probability of a firm being in the top five percent of the employment size distribution. The concluding section contains a summary and draws out some further implications of the findings.

2. Data and Measurement

2.1 Data

In order to measure the precise level of employment at start-up, track firm employment over time, and incorporate characteristics of business founders, we link together several data sources. We start with the U.S. Census Bureau's Business Register (BR), which includes all nonfarm businesses filing Internal Revenue Service tax forms as individual proprietorship, partnership, or any type of corporation, and with receipts of $1,000 or more. The BR is available quarterly, and employment is the number of employees in the payroll period including March 12 for quarter 1, June 12 (quarter 2), September 12 (quarter 3), and December 12 (quarter 4), as reported to the Internal Revenue Service at the Employer Identification Number (EIN) level. Different units within a firm may file under separate EINs each quarter, and we aggregate such cases to obtain firm-level employment.

We define entry as hiring a first employee, so entrants in a particular quarter are firms with positive employment for the first time. We also combine information from the the Longitudinal Business Database (LBD), an annual database containing all non-farm businesses with positive payroll in the year, and we require that an entrant first appear in the LBD the same year and that none of an entrant's establishments that year had positive employment beforehand in the LBD. We restrict the sample to firms found in the LBD because we use the LBD's longitudinal establishment links across annual BR files to track firms and their reorganizations over time. We take employment from the BR rather than the LBD, because the LBD only contains employment in the pay period including March 12. Age 0 is the firm's first quarter with positive employment.[7]

For detailed information on characteristics of firms and owners, we focus on the four quarterly entry cohorts in 2007 in order to link these data to the Census Bureau's 2007 Survey of Business Owners (SBO).[8] The SBO uses the BR as the sampling frame, stratified by state, industry, owner demographic group, and whether the firm has employees or not.[9] The largest companies in each stratum are selected with certainty, and the remainder of the sample is randomly selected. The SBO has been carried out every five years, and we use the 2007 SBO rather than the 2002 or 2012 data because the 2002 SBO lacks information on several of the factors we wish to study, and the 2012 SBO permits observation only on a short time span after start-up.[10] Motivated by concerns about whether results differ for firms founded in 2007, just before the Great Recession, compared to firms founded in other years, we have also estimated all the Age 0 relationships with

---

[7] As with any longitudinal data, there is a possibility of broken links, and the data do contain a small number of implausibly large entrants. But while such outliers may have large leverage on estimated effects in standard employment regressions, our approach of estimating the probability of being in a high-growth group gives no extra leverage to these firms and is therefore more robust to such measurement errors. We also find similar results when we exclude all observations over 100 employees, as discussed in the robustness subsection below.

[8] Choosing the most recent start-up cohort reduces but does not completely eliminate survival bias, as the survey was conducted in 2008 and 2009 for the 2007 reference year. This problem is larger for studies of the 1992 CBO, which was carried out in 1996 for reference year 1992.

[9] The SBO does not collect ownership information on firms without an individual owner of at least 10 percent or that are majority owned by another company or organization, Employee Stock Ownership Plan, members in a cooperative or club, an estate or trust, an Alaska Native Regional or Village Corporation, or an American Indian tribal entity.

[10] The 2002 SBO does not contain information on whether the owner was born in the U.S. or not, husband-wife ownership, prior business ownership, and amount of start-up capital.

the 2012 data. The results from this analysis, discussed in the robustness sub-section and provided in the Appendix, are very similar to those from 2007.

We are interested in studying the determinants of job creation over a longer period, not only at start-up. For this purpose, we use age 7 employment, defined as the firm's employment in the same quarter of 2014 as its start-up quarter in 2007. We choose age 7 for measuring the firm's longer-run net job creation because it is the oldest age we can currently observe for the 2007 start-up cohorts (2014 is the most recent available year for the BR), and because some researchers (e.g., Nightingale and Coad 2014) define "entrepreneurial firms" as those under age 7. Thus, age 7 employment is the net job creation over the entrepreneurial period so defined. Even if available, using employment at a later age would have drawbacks: the older the firm, the more difficult it is to attribute its performance to a single origin and founding team, because of firm boundary and ownership changes taking place over time.

Tracking a firm longitudinally involves measurement problems that are much more prevalent among older and larger firms.[11] In particular, the firm identifier can change due to a re-registration, change in legal form, or switch from single- to multi-establishment. The LBD is designed to track such changes, but we implement additional procedures to restore broken firm links from identifier switches.[12] Approximately 6.3 percent of firms in the top 5 percent of employment at age 7 in the Table 2 and 3 sample below undergo a firm identifier switch between age 0 and age 7, vs. 2.9 percent of firms in the bottom 95 percent. This is consistent with growing firms being more likely to change legal form and/or become a multi-establishment firm.

Our focus is on the firm's organic growth, rather than growth through acquisition, so we adjust age 7 employment to remove the effects of establishment acquisitions and divestments. By definition of entry in our analysis, age 0 employment excludes acquisition of existing establishments. We use the LBD to track subsequent boundary changes with procedures similar to Haltiwanger, Jarmin, and Miranda (2013) and Brown and Earle (2017). If the firm acquires a pre-existing establishment, the establishment's employment in the year prior to acquisition is subtracted from the firm's age 7 employment.[13] If an establishment is sold or spun off and continues to operate in subsequent years, the establishment's employment in the year prior to divestment is added to the firm's age 7 employment. The reasoning behind these boundary change adjustment procedures is that the firm is responsible only for those establishment employment changes that occur while the establishment is under its control. In practice for the sample analyzed below these adjustments are inconsequential, because while a higher share of firms in the top 5 percent of employment have an employment adjustment due to boundary changes relative to the

[11] McKelvie and Wiklund (2010) discuss challenges of tracking firms across time when measuring growth.

[12] We link firm identifiers A and B if identifier A last appears in year t and identifier B first appears in t+1, at least one establishment is in A in t and B in t+1, the establishments in A in t and B in t+1 (denoted A-B) have more total employment in t than any other establishment groups in A in t switching to some other firm identifier in t+1 (A-C, A-D, etc.), and A-B has more employment in t than any other establishment groups switching to B in t+1 from another firm identifier (E-B, F-B, etc.).

[13] The BR may sometimes misclassify new establishment openings by pre-existing firms as new firms. The Census Bureau learns about such establishment-firm linkages during the quinquennial economic census, but it does not know when the firm first owned the establishment. In such cases we misclassify new establishment openings as acquisitions and undercount the firm's organic growth. It is unlikely that any of the firms in our SBO sample are new establishments owned by pre-existing firms, because businesses majority owned by other businesses are not asked for owner information.

bottom 95 percent, the share is well below 1 percent in both groups. If a firm disappears from the LBD prior to 2014, and none of its establishments continue to operate subsequently, we treat it as an exit, imputing zero for age 7 employment.[14] In contrast, if at least one of its establishments in the firm's last year in the LBD continues to operate in subsequent years, we impute the firm's employment (or boundary-adjusted employment if it had boundary changes) in its last year in the LBD as age 7 employment.[15]

The sample used for the analysis of characteristics consists of all owner observations for firms in the four quarterly 2007 LBD start-up cohorts in the 2007 SBO that have non-missing values for all the characteristics. The sample size is 55,800 owners of 37,100 firms, which is about 7.0 percent of all firm start-ups that year (U.S. Census Bureau, 2016). To make the analysis firm-level, we weight each owner by the ownership equity shares (so they sum to one for each firm). To reflect the industry-size composition of the LBD, we also weight by the inverse of the sample-population ratio (the share of firms in the two-digit NAICS industry–employment category in the 2007 quarterly start-up cohorts in the LBD divided by the sample's share of firms in the two-digit NAICS industry–employment category). The size categories for these weights are 1, 2-4, 5-9, 10-19, and 20 and more employees. We use LBD-based weights for the 2007 start-up cohorts rather than SBO survey weights, because the SBO survey weights do not take nonresponse or firm age (a crucial variable for our analysis) into account.[16]

All the independent variables are measured for the year 2007, the start-up year for the firms in the main sample and the reference year in the SBO. Founder characteristics from the SBO include basic demographics (age, gender, race, ethnicity, and immigrant/native), human capital (type of education, veteran, and prior entrepreneurial experience), and the size of the founding team and relationships among multiple founders (family/unrelated and diversity by demographics and education). We use the firm's 6-digit industry from the BR, and categories of the amount of start-up finance from the SBO.

Details of the construction of founder characteristics from the raw data are as follows. Among races, we distinguish whites, African Americans, and Asians, and we group native Hawaiians, Guamanian or Chamorro, Samoan, and other Pacific Islanders and some other race as "other minorities." Immigrant indicates the owner was not born in the United States. For firms with multiple owners, the gender, race, ethnicity, and immigrant variables are defined to indicate whether all the firm's owners are in that category or not, and thus include a label "all," in order to permit us to measure the impact of diversity, as discussed below. For instance, for analysis of gender diversity, we define "all female" and "all male" variables to indicate firm with owners only from one gender or the other (including single owners).

Among human capital variables, the educational categories are self-explanatory. Veteran indicates whether the owner is a veteran of any branch of the U.S. military service, including the Coast Guard. Prior business indicates that the owner previously owned a different business prior to owning the current business.

[14] Just as firms first appearing with pre-existing establishments are not classified as entrants, firms that disappear without closing down all their establishments are not classified as exits.

[15] In the sample for Tables 2 and 3, 7.3 percent of firms from the age 7 top 5 percent exit before age 7 with subsequently continuing establishments, compared to 0.2 percent of firms in the bottom 95 percent.

[16] Foster, Grim, and Haltiwanger (2016) apply similar LBD weights when using the Annual Survey of Manufactures.

The data permit us to construct detailed measures of the size and composition, including family relationships, of founding teams. We define diversity variables as follows: Gender diversity indicates that the business is jointly owned by at least one owner of each gender (except when husband and wife, for which we provide a separate category), ownership ethnic diversity indicates that the business is jointly owned by at least two individuals with different race or ethnicity from one another, ownership immigrant diversity is a dummy equal to 1 when the business is jointly owned by individuals who are immigrants and U.S.-born, and multi-generation indicates that at least one owner is 20 or more years older than another. By controlling for all-female, the variables for different types of husband-wife ownership and gender diversity for non-couples measure whether gender ownership effects vary depending on who else co-owns the firm. Similarly, including variables for all of one ethnicity or race, all immigrant, and ethnic/racial and immigrant diversity in the regression allow us to examine whether race/ethnicity and immigrant effects differ with homophily or diversity among founder teams along those dimensions.

2.2 Measuring High Growth

Most studies of high-growth firms typically examine existing firms, incumbents, with growth above a certain threshold over a period of one or a few years. Some use the top 1, 5, or 10 percent of the growth rate distribution, which results in a high-growth group dominated by initially very small firms, because the same absolute employment gain measured as a rate is magnified for smaller firms, by construction. This implicitly favors undersized start-ups that are catching up. Some studies examine firm growth in a particular year, although as noted below, growth is highly volatile over time so that a particular year may not reflect longer term job creation. To avoid these problems, Eurostat-OECD's (2008) definition focuses on firms that have at least 10 employees at the beginning of the period and an average of at least 20 percent annual growth over the next three years. As Daunfeldt, Johansson, and Halvarsson (2015) point out, however, a 10-employee initial size restriction excludes the vast majority of firms, which is also unattractive.[17] As another way to address the problem that tiny firms can more easily have high growth rates, Acs, Parsons, and Tracy (2008) define "high-impact firms" as those at least doubling sales over a four-year period and with a product of absolute and percent change in employment (sometimes called the "Birch index") of at least two during the same period.

The definition that focuses on incumbents with a high growth spurt over some time period has multiple drawbacks. Firm growth has been found to be extremely volatile even with respect to multi-year periods (e.g., Acs, Parsons, and Tracy 2008; McKelvie and Wiklund 2010). Daunfeldt and Halvarsson (2015), for example, show that Swedish firms with a three-year period of high-growth tend to have declining growth in the previous three-year period, and the probability that they repeat their high growth performance in either of the next two three-year periods is very low. These high growth definitions also exclude job creation from entrants, which Decker et al. (2014) report to account for about 20 percent of gross job creation in the U.S., and which constitute the only age group to show substantial net job creation (Haltiwanger, Jarmin, and Miranda 2013).

[17] The definition excludes almost 95 percent of surviving firms in their Swedish sample.

Other studies focus on start-up size, and a subset of those follow the same cohort of firms for several years from start-up.[18] Coad et al. (2014) suggest that the best way to ensure a firm reaches a large size at a particular age is to be large at start-up. As we show below, start-up size is a powerful indicator of size at age 7. Examining employment at age 7 size places uniform weight on job creation throughout the entrepreneurial period, including from start-up. By this measure, firms can be high growth either by creating many jobs at start-up or by catching up later.

In this paper, we define high-growth entrepreneurship as the subset of entrants with the highest net job creation. Given that entrants have zero employment prior to entry, their net job creation is simply their size. We measure high growth by this definition at ages 0 and 7. In most of the results reported below, the high-growth group is defined as the top 5 percent, distinguishing them from the bottom 95 percent of the employment distribution within the sample. Average employment among those in the top 5 percent is much larger than in the bottom 95 percent: for the sample studied below, average employment in the top group is 57 and 67, and for the bottom 95 percent it is 3 and 2, at age 0 and at age 7, respectively. The top 5 percent account for 52.0 percent of the sample's employment at age 0 and 66.6 percent at age 7.[19] At age 0 the top 5 percent have 17 or more employees, and at age 7 the threshold is 19.

The regression specifications are variants of the following equation:

$$\Pr(HG_{ijt}) = X_{ij}\beta + \theta_j\delta + \rho_j + K_j + S_j + u_{ijt}, \quad (1)$$

where HG_{ijt} is a dummy equal to 1 if founder *i* owns firm *j* with employment in the top 5 percent of the employment distribution among firms at age t, and t = 0 or 7 in alternative specifications. X_{ij} contains characteristics for founder i of firm j, θ_j contains firm-level characteristics, ρ_j is a vector of dummies denoting the quarter of 2007 in which firm *j* first has positive employment, K_j is a vector of start-up capital amount categories, S_j is a vector of 6-digit NAICS industry dummies, and u_{ijt} is an idiosyncratic error term. As noted above, the regressions are weighted by owner shares (so that each firm rather than each owner receives equal weight) and by LBD weights (so that results reflect the full population).

For each firm age (0 and 7), we estimate a base specification including only factors that are predetermined at the time of start-up, excluding start-up capital K_j and industry dummies S_j. We exclude these variables from the base specification because they may be at least partly choice variables of the entrepreneur in the start-up process. The firm's growth potential may influence the amount of financing through, for example, the quality of the business plan presented to investors.

[18] See, for example, Cabral and Mata (2003) and Garnsey, Stam, and Heffernan (2006). The latter study argues that following a cohort over the same time period reduces survival bias and thus can increase the consistency in the measurement of impacts of firm growth factors. Focusing on mean employment growth in firms owned by immigrants, Kerr and Kerr (2017) follow entry cohorts, but their outcome variables are 3- and 6-year growth relative to age 0 employment, so they exclude age 0 job creation from their analysis (and their regression estimates control for age 0 employment).

[19] As a robustness check, we have also re-run the analysis using the top 2 percent (36 employees or more at age 0 and 40 employees or more at age 7) and top 10 percent (10 employees or more for both ages) thresholds. The top 2 percent make up 39.4 percent of sample employment at age 0 and 50.8 percent at age 7, while the top 10 percent account for 63.1 percent of employment at age 0 and 79.7 percent at age 7. Results using these alternative thresholds, available upon request, are qualitatively similar to those using 5 percent.

And entrepreneurs desiring to create high-growth firms may choose sectors where large, fast-growing firms are more common.[20]

Though the factors we examine other than industry and finance are predetermined at start-up, some of them could be jointly determined with the high growth outcome through unobserved channels, including the founders' motivations and the quality of the entrepreneurial idea. For example, it is possible that some human capital investment decisions are driven by the intention to start-up a high-growth business. It is also likely to be easier to recruit additional founding team members when the business idea has greater potential, which could be reflected in a larger coefficient on multiple owners.

To see if start-up capital and sectoral choice are channels through which predetermined characteristics influence high growth, we estimate an additional specification adding K_j and a second specification with both K_j and S_j. If, say, female entrepreneurs systematically access less financing, and K_j is positively associated with high growth, then including K_j will raise the coefficient on female owner.[21] If a coefficient rises (falls) after controlling for S_j, that suggests that the particular type of entrepreneur systematically selects sectors with a lower (higher) share of high-growth entrepreneurship. Coefficients after controlling for sector also show the performance of particular types of entrepreneurs relative to their competitors, which is relevant for their long-run viability. We test for the statistical significance of such differences in coefficients across specifications by jointly estimating the equations.

We also test for differences in coefficients across age (age 7 versus age 0) by pooling the data for the two ages, allowing all coefficients to vary by age, and testing for equality of the coefficients for the same variable at the two ages. This permits us to assess the degree to which the predictive power of a coefficient for firm size at age 0 persists or diminishes at age 7.

3. Results

3.1 Entry Size and Growth

The first questions we examine concern the basic patterns of heterogeneity in the size of firms upon entry and in their subsequent growth. We measure the extent to which firms that start large continue to be large at the end of the entrepreneurial phase (age 7) and the degree of concentration of employment among large firms. The data construction follows the sample procedures described above, containing all firms in the LBD that first have positive employment in the BR in one of the quarters of 2007. We use a transition matrix across size categories for these

[20] Hurst and Pugsley (2011) show that entrepreneurs with nonpecuniary motives for owning the business tend to choose sectors with a higher share of small firms.

[21] Systematic differences in the amount of start-up capital could be due to individual choice or external financial constraints. Some types of entrepreneurs may be more reluctant to use their own resources or take on debt than others, their creditworthiness may be systematically different, or investors may discriminate against some types of borrowers. We are unable to test among these alternatives here, although results including the industry dummies provide evidence on the degree to which individual choice is reflected in the sector in which the business operates. Fairlie, Robb, and Robinson (2016) and Coleman and Robb (2014) report that African-American, Hispanic, and female entrepreneurs systematically use less start-up capital. Blanchflower, Levine, and Zimmerman (2003), Blanchard, Zhao, and Yinger (2008), and Fairlie, Robb, and Robinson (2016) provide evidence of discrimination against African-American and Hispanic entrepreneurs in the small business credit market. Coleman and Robb (2014) find that the loan denial rate does not vary significantly by gender, but female entrepreneurs are less likely to apply for credit due to fear their loan application will be denied, even when controlling for measured creditworthiness.

entrants from start-up in 2007 (age 0) to age 7 in 2014, with categories defined as 1, 2-4, 5-9, 10-19, and 20 or more employees.[22]

Tables 1.1 and 1.2 display the results, with row percentages in the former and column percentages in the latter. Most firms start very small: 78 percent have initial employment less than 5. Only 4.1 percent of the entrants have 20 or more employees, but they account for over half (54.1 percent) of all age 0 employment. Employment concentration in the largest category is even higher at age 7, when the largest category at age 7 is 3.3 percent of the number of initial start-ups, but accounts for 60.1 percent of all age 7 employment. Of those starting with 20 or more employees, 31.4 percent remain in that category at age 7, while most of the rest exit. Those remaining in the largest category make up 38.2 percent of that category at age 7 despite being just 4.1 percent of the start-ups, so firms starting large have a much higher propensity to be large at age 7 than firms starting smaller: the probability for firms starting with 20 or more employees is 3 times higher than for firms starting with 10-19 employees, and about 40 times higher than for firms starting with one employee.

Not only do large entrants tend to stay large, but they also tend to grow faster than smaller entrants. Table 1.3 shows the average employment changes by start-up size category and separately for exiting, declining, unchanging, and growing firms to age 7. The average job loss among exiting and declining firms increases in initial size, which may not be surprising because the larger entrants have the most to lose, and their exit rate is not much lower than for smaller firms. But the average employment growth among growing firms is also increasing in start-up size. Gross job creation (per firm) is highest among the largest entrants, and the big future job creators are more likely to be found among the largest entrants. Thus, the "up" dynamic of fast growth referred to by Decker et al. (2014) is strongest for the largest entrants, further motivating our analysis of the large entrants at age 0.

These results demonstrate the importance of understanding the determinants of starting size. Note, however, that, although the probability of growing large (20 or more employees) after starting small (fewer than 20) is much smaller than the probability of remaining large, the former nevertheless outnumber the latter in an absolute sense: 61.8 percent vs. 38.2 percent of the large category, respectively. The age 7 employment shares of firms by age 0 employment categories are also more evenly distributed than at age 0. It is thus possible that the factors explaining large size at birth and at age 7 could differ significantly. We examine this below.

Standard models of industry dynamics imply that all entrants should choose the same, optimal size, but our results and others imply substantial heterogeneity. To measure the extent to which start-up size and growth can be explained by industry, we estimate a set of regressions of age 0 and age 7 employment on highly disaggregated (6-digit NAICS) industry dummies (measured at age 0). To assess the role of capital access in accounting for size variation, we estimate similar regressions with category dummies for the amount of start-up capital. We also estimate a set with both industry dummies and start-up capital. The R^2 with industry dummies is 0.086 at age 0 and 0.034 at age 7, it is 0.020 at age 0 and 0.006 at age 7 with start-up capital, and it is 0.100 at age 0 and 0.039 at age 7 with both industry dummies and start-up capital. It is

[22] Kerr and Kerr (2017) provide a transition matrix for start-up cohorts in the LBD, with a focus on the share of immigrants in each cell, but they do not describe the size distribution at age 0 (the marginal distribution).

noteworthy that all the R^2s fall with age, and in particular that within-industry size variation relative to cross-industry variation increases with age.

Given our focus on high-growth entrepreneurship, we also calculate the R^2, replacing the dependent variable with a dummy for being in the top 5 percent of the employment distribution at the particular age. Using the top 5 percent dummy, the R^2 with industry dummies is 0.128 at age 0 and 0.075 at age 7, it is 0.072 at age 0 and 0.050 at age 7 with start-up capital, and it is 0.172 at age 0 and 0.122 at age 7 with both industry dummies and start-up capital. In all cases, the R^2s fall with age, although less so than for the employment regressions, and they are higher for these top 5 percent regressions than for employment regressions. Thus, the detailed industry and start-up capital variables do better at distinguishing the high-growth group, and their effects are also more persistent for the top 5 percent. But these factors explain a small part of the heterogeneity for either dependent variable, further motivating our examination of the effects of founder characteristics in the next subsection.

3.2 Determinants of High Growth at Entry and Age Seven

Table 2 displays descriptive statistics for founder and firm characteristics, organized into sub-tables by groups of related variables. Firms are divided into the top 5 percent and bottom 95 percent of the sample employment distribution at ages 0 and 7.[23] The first four columns show the share of the employment-age category that has the particular owner or firm characteristic, and the last two columns are odds ratios. We report the results from six LPM regression specifications in Table 3. The dependent variable is a dummy for being in the top 5 percent of the sample employment distribution at age 0 or 7, in alternative equations. The first two specifications include gender, owner age, ethnicity, race, citizenship, education, veteran status, prior business experience, and founding team characteristics. We then control for start-up finance amount categories and industry dummies. To see how finance affects the coefficients separately from industry, specifications 3 and 4 add start-up finance amount categories, then specifications 5 and 6 add 6-digit NAICS industry dummies.[24] We conduct tests for equality of coefficients at age 0 and 7 and across specifications after adding start-up finance and industry dummies (i.e., the specification in column 3 versus column 1, 4 versus 2, 5 versus 3, and 6 versus 4).

Beginning with differences by gender of founder, numerous studies have found that women-owned businesses are both less common and tend to grow more slowly on average than those owned by men.[25] Despite the differences in our approach from previous research, our results

---

[23] As discussed in the data section, the top 5 percent employment thresholds at age 0 and age 7 correspond to 17 and 19 employees, respectively, so results are similar to the 20+ employees category in the transition matrices above. But we find it more natural there to use absolute employment, while here it is simpler to keep the fraction in the top group constant in order to interpret the comparison of results at age 0 and age 7.

[24] This procedure is somewhat similar to the studies of the probability of having at least one employee with 1992 CBO data by Fairlie and Robb (2007, 2009) and Robb and Fairlie (2009), although they present results only with both start-up capital and industry controls in the same specification.

[25] For example, Jarmin and Krizan (2010) find that women-owned businesses have lower average employment growth rates in the 2002 SBO linked to the LBD. Using the CBO, Fairlie and Robb (2009) report women have a lower probability of hiring employees among other measures of business success, while Kalleberg and Leicht (1991) find small, statistically insignificant disadvantages of women in survival and earnings growth by gender of owner in a survey they conducted of 411 firms, 99 of them owned by women. These studies examine cross-sections, not

in Table 2.1 are qualitatively similar: we find that all female-owned businesses account for a much smaller fraction (16 percent) compared with all male-owned (51.4 percent). Even among this smaller set of female-owned businesses, the propensity to be high-growth, as shown by the odds ratios, is low, close to 0.5, while men are correspondingly over-represented in high-growth entrepreneurship. All regression specifications in Table 3.1 show negative and significant coefficients. The magnitude of the estimated effects are not significantly different at age 7 compared to age 0 and for the base specification imply a 1.7 percentage point lower probability of being in the top 5 percent, or a lower probability (relative to the baseline 5 percentage points) of 34 percent.

This large estimated gender gap is significantly diminished, to 20 percent, when controlling for start-up finance. This finding is consistent with the presence of greater financial constraints for women as well as with the possibility of more non-pecuniary motives for women founding businesses. Table 3.1 shows that the propensity to be a high-growth firm is positively associated with the amount of start-up capital, implying that if a coefficient on an owner characteristic increases (decreases) after controlling for start-up finance, that indicates that the characteristic is associated with less (more) start-up capital. The results here thus indicate that women-owned businesses use less start-up capital, consistent with Fairlie and Robb's (2009) finding that less start-up capital helps to explain why women-owned businesses have lower average sales. The age 7 female coefficient drops almost as much after controlling for start-up finance as the age 0 coefficient drops, suggesting that using less start-up capital has long-lasting effects on growth.

The estimated gender gap increases substantially in the specification controlling for industry. The coefficients around -2 imply a 40 percent lower probability of being in the top 5 percent. This result suggests that women tend to enter industries with higher shares of high-growth entrepreneurship, and it appears inconsistent with women choosing to start businesses for non-pecuniary, non-growth-related reasons.[26] It also contrasts with Fairlie and Robb's (2009) report that female businesses have an unfavorable industry distribution for average sales, although again their methods are quite different from ours.[27]

The age of the founder(s) may be associated with business success as a result of human capital accumulation (labor market experience, increasing in age), financial constraints (likely to decrease with age), and time horizon (decreasing in age). In all cases, the effect may be non-linear. We find, as shown in Table 3.1, that firms with older owners are much more likely to be large at start-up, with a roughly concave shape, but the owner age effect disappears by firm age 7. However, we also find that controlling for finance reduces the owner age effect similarly for the age 0 and 7 coefficients. In this specification at age 7 (shown in column 4), the age profile actually becomes significantly negatively sloped, so that entrepreneurs under 35 years old are much more likely to start top 5 percent companies than those 45 and older. This implies that not only do younger entrepreneurs use less start-up capital, but the negative effect of lower capital is persistent through

distinguishing firm size or growth by age, while we follow an entry cohort, and they report average differences, while our focus is the right tail of the distribution, so our approach differs significantly.

[26] Hurst and Pugsley (2011) report a positive correlation between non-pecuniary motivations for founding the business and the share of small firms in the industry.

[27] Fairlie and Robb's (2009) female coefficient changes from -0.69 to -0.57 when controlling for both start-up capital and industry.

this early phase of business life. It also suggests that older entrepreneurs may have a skill disadvantage when it comes to starting up high-growth entrepreneurships, so that traits such as flexibility and ability to develop new ideas may dominate labor market experience.[28] This result is also inconsistent with Cabral and Mata's (2003) interpretation that a diminished founder age effect as the firm ages means that liquidity constraints lessen over time.[29] Finally, we find that the owner age effects are not sensitive to industry selection.

Differences in performance by race, ethnicity, and citizenship could potentially be due to discrimination in financial markets or by customers, as well as correlated skills or preferences of individuals selecting into entrepreneurship. Previous research reports significant differences in average business size and growth along these dimensions, but does not analyze the probability of high growth or large size.[30] Our results for unconditional odds ratios in Table 2.1 for Hispanic, African-American, and Asian owners are well below 1, the immigrant odds ratios are just below 1, and those for other minorities are above 1. The coefficients for these categories are insignificant in the base regressions, however, with the exception of African Americans at firm age 0, who have a coefficient of -1.4, implying a 28 percent lower probability of operating a high-growth entrant, as shown in Table 3.2. The African-American coefficient becomes positive, but statistically insignificant at firm age 7. The inclusion of start-up finance moves the Hispanic and African-American coefficients in a positive direction, consistent with these groups using less start-up capital. The Asian and immigrant coefficients are lower (though insignificantly so for Asian at age 0), suggesting they use more start-up capital. The coefficients for all these ethnicity and race categories move sharply in the negative direction with industry controls (though insignificantly so for other minority race at age 0), implying that minorities and immigrants are more prevalent in industries with higher shares of high-growth entrepreneurship.

Turning to measureable skills, formal education may increase an entrepreneur's ability to make decisions about business development. It may also be associated with better social networks and higher earnings prior to starting the business, increasing access to start-up capital.[31] Cabral and Mata (2003) find that years of education is positively associated with employment at both age 0 and 7, but more strongly so at age 7. Our results, in Table 2.2, show that bachelor's degree has the highest odds ratios of being in the top 5 percent. Graduate degree, however, is little different from lower educational attainment at age 0, suggesting that the effect of education is concave. The impact of graduate degree is much higher at age 7 than age 0, though, implying that a graduate degree is associated with high growth after start-up. These results hold up when controlling for

[28] Cabral and Mata (2003) find firm size correlated with owner age at firm age 0 but not 7, which they interpret as more consistent with liquidity constraints, which they reason should diminish over time, while skills should persist. Their data lack information on the amount of finance, however.

[29] In a footnote Cabral and Mata provide an alternative explanation for the diminished owner age effect over time that is more consistent with our findings, namely that firm-specific experience eventually overtakes previous owner labor market experience in importance. The equalization could also simply reflect the aging of the owners over seven years (many of them would be in the next higher age category if measured at firm age 7).

[30] Fairlie and Robb (2007, 2009) and Robb and Fairlie (2009) find that Native American-owned and Asian-owned businesses have higher average sales than White-owned businesses, while those of AfricanAmerican-owned businesses are lower. Kerr and Kerr (2017) find that immigrant-owned firms in the LBD start with lower average employment, not controlling for other owner or firm characteristics. Jarmin and Krizan (2010) find that Hispanic, African American, and other minorities (except Asian) have lower employment growth rates.

[31] See Baptista, Karaoz, and Mendonca (2014) for a discussion.

other owner characteristics, shown in Table 3.3. Controlling for start-up finance lowers the bachelor and graduate degree coefficients, while industry controls sharply raise them for graduate degree, and also for bachelor's degree at age 0. This means more highly-educated owners use more start-up capital and choose industries with a lower share of high-growth entrepreneurship. The latter result is consistent with Hurst and Pugsley's (2017) observation that skilled professions (e.g., dentists, doctors, lawyers, accountants, and insurance agents) are industries dominated by small businesses both when firms are young and old.

Military service is another type of human capital that could influence entrepreneurial performance, but it has been little studied. Using the 2002 SBO, Headd and Saade (2008) find that the size and industry distributions and of veteran-owned and non-veteran-owned firms are similar, without controlling for other factors, and they show that veterans and non-veterans have similar propensities to use different start-up financing sources. We find that veterans have a lower propensity to own firms in the high-growth entrepreneurship group at both age 0 and 7, however, with or without controls. Less negative coefficients with start-up finance controls indicate that veterans use less start-up capital.

Past entrepreneurial experience may give owners better managerial and technical skills, a more developed business network, and greater knowledge about business opportunities (e.g., Baptista, Karaoz, and Mendonca 2014; Shaw and Sorenson 2017). It could also increase start-up capital via personal wealth accumulation, credit and entrepreneurial performance history, and an investor network. Shaw and Sorensen (2017) find that firms owned by serial entrepreneurs in Danish data have higher employment than those with novice entrepreneurs, but this result reverses once controlling for other owner and firm characteristics. As shown in Table 2.2, we find that entrepreneurs with prior business ownership experience account for just over half of the founders in the sample, and they are twice as likely to be classified as high growth at age 0 and 63 percent more likely at age 7, compared to those with no prior ownership experience. With the baseline demographic, human capital, and founding team controls, these differentialsdecline but they are still substantial, as shown in Table 3.3: 40 percent at age 0 and 30 percent at age 7, both statistically significant. Controlling for the amount of start-up capital reduces both estimates to about 20 percent, consistent with serial entrepreneurs using more start-up capital, which Shaw and Sorensen (2017) also find.

About half the firms in our sample are founded by teams rather than single entrepreneurs, and the data permit us to investigate a number of interesting questions about the size and composition of the teams. A larger founding team can involve more diverse skill sets, providing a "jack of all trades" in a group that may be hard to find in an individual entrepreneur. More team members may also provide greater resources and networks for start-up capital.[32] The data in Table 2.3 show that nearly half the firms have multiple owners, with most of these (nearly 90 percent) being two-owner businesses. Start-ups are frequently family-owned, and Table 2.3 implies that more than 70 percent of two-owner businesses are founded by related individuals, most of them

[32] See Ruef, Aldrich, and Carter (2003) for a discussion of the literature about founding teams. Baptista, Karaoz, and Mendonca (2014) find that firms with multiple owners have higher survival rates than single-owner firms.

married couples. Clearly, resources and interpersonal dynamics may differ in family and non-family teams.[33]

Our calculations of odds ratios in Table 2.3 show that at age 0 firms owned by three or more owners have the highest propensity to be in the group of high-growth entrepreneurship, followed by family businesses not owned by a husband and wife, two unrelated owners, husband and wife-run business, businesses primarily run by the husband or by the wife, and single-founder businesses. After controlling for other founder characteristics, as shown in Table 3.4, having three or more owners is still the ownership type with the strongest association with high growth, again followed by family businesses not owned by a husband and wife and two unrelated owners, while the differences among the other types of owners are generally insignificant, with the exception that equal operation by a couple is more positive than the others at age 7. Controlling for start-up finance reduces the three or more owners, two unrelated owners, equal operation by a couple, primarily operated by the husband, and primarily operated by the wife (at age 7 only) coefficients, indicating that they raise more capital. Industry controls further reduce the three or more owners, equally operated by a couple, primarily operated by the wife, and two unrelated owners (only for age 7) coefficients. These groups have a greater propensity to locate in industries with more high-growth entrepreneurship.

Multi-generational ownership could potentially combine experience with new ideas. The relationships among owners may be less equal when age gaps are large, though, which could result in conflicts. The odds ratios in Table 2.3 are over 2 for multi-generational ownership, but once other founder characteristics are controlled, the differences are insignificant, as shown in Table 3.4.

A related issue is the impact of diversity versus similarity, or "homophily," in founding teams. Founders of the same gender, race, or ethnicity may have easier communication, coordination, and trust-building. On the other hand, gender and ethnic diversity can bring together varied skill sets and knowledge, leading to greater creativity and innovation, and they can combine desirable traits in a team that are not present in single individuals, thus providing another way to a "jack of all trades" (Lazear 2004, 2005). Hoogendoorn and van Praag (2012) report that business performance decreases with increasing ethnic diversity below a certain share of minorities on the founding, team, but it becomes positive above a certain threshold. Hoogendoorn, Oosterbeek, and van Praag (2013) find that equally balanced male-female founding teams achieve higher profits than male-dominated teams.

Our data in Table 2.3 show odds ratios greater than 1 for gender, ethnic, and immigrant diversity, but the regression results in Table 3.4 are very different: the gender diversity coefficient is negative and significant at age 7, the age 0 ethnic diversity coefficient is negative and significant, and the immigrant diversity coefficients are insignificant. The gender and ethnic diversity coefficients are still negative but smaller when controlling for start-up finance, while the

[33] Brannon, Wiklund, and Haynie (2013) suggest that trust and familiarity are more important for a family business, while unrelated team members may be chosen based on skills and knowledge. They hypothesize that couples have worked out joint decision-making processes (e.g., about household finances), whereas non-couple family members are more likely to be in conflict with one another due to long-standing family roles, and their analysis of 295 teams from the PSED shows that couple-owned firms have a higher probability of ever having sales than other family firms.

immigrant diversity coefficients are larger (and negative). The addition of industry controls increase the magnitude of the negative age 7 gender diversity coefficient.

Tests of differences between the all-female coefficients in Table 3.1 and the couple and gender diversity coefficients in Table 3.4 suggest that married couple ownership is associated with a greater propensity to be high growth than all-female ownership or by women and men not married to each other (though the differences between couples with unequal operation and gender diversity are significant only at the 10 percent level at age 0). All-female ownership is associated with a higher propensity for high growth than ownership by women and men who aren't married to each other.

A final issue concerning the founding process is whether the business is a genuinely new idea or a franchise of an existing firm. A franchise relationship provides the entrepreneur with a ready-made business model, which could be conducive to faster growth early in the lifecycle. It may also require more start-up capital. Jarmin and Krizan (2010) report higher average employment growth for franchise businesses. We find that franchises are many times more likely to experience high growth than independent businesses, though this effect is somewhat smaller at age 7. Franchises use more start-up capital and operate in industries with a larger share of high-growth firms, as suggested by the sharply reduced coefficients in the presence of financial and industry controls.

3.3 Robustness

Research on the determinants of firm growth is marked by inconsistencies of results across studies. McKelvie and Wiklund (2010) point to several measurement issues potentially contributing to the inconsistencies, such as when in the lifecycle to begin tracking the firm, organic vs. acquisitive growth, and how the growth is measured. Growth factors could also vary with macroeconomic conditions, which may be particularly relevant for the sample in this study, given that the Great Recession began soon after the firms in our sample started up.

To assess the importance of these and other concerns, we have conducted robustness exercises along several dimensions. As Shane (2008) notes, firms may need time to complete their initial hiring process, in which case employment in the first quarter of life may not be the right time to measure start-up size. When replacing age 0 employment with age 1 employment (four quarters after birth) in the employment transition matrix, we find very similar patterns, as shown in Appendix Tables A1-A3. Table A1 shows that firms starting large generally either stay large or exit, firms starting large have a much higher propensity to be large at age 7 than firms starting smaller, cohort employment is highly concentrated in large firms at birth, and the concentration is even higher at age 7. The employment transition matrix between age 0 and age 1 in Table A2 shows that firms that are large at age 0 make up the bulk of firms that are large at age 1 (60.5 percent), and the share of firms that are below 20 employees at age 0 that grow to 20 or more by age 1 is minor, for instance only 12 percent of firms with 10-19 employees at age 0 do so.

The patterns of association of founder characteristics with employment size are also very similar at age 0 and age 1. As shown in the regression results in Table A3, when replacing the top 5 percent of employment at age 0 dummy with that at age 1, differences in results are nearly all statistically insignificant. The only exceptions are coefficients that are still negative, but larger in magnitude, on all immigrant and on started the business more than two years before.

Another potential concern is the choice of start-up year, which in the case of 2007 immediately precedes the Great Recession. To investigate the extent to which the results are sensitive to this choice, we run regressions using the only other data containing all the same variables, from the 2012 SBO. Table A4 shows results for the four 2012 quarterly start-up cohorts, constructed according to the same procedures as described above for 2007. Most patterns are quite similar for the 2012 and 2007 cohorts at age 0 (of course we cannot yet examine the 2012 cohort at age 7). Comparing the 2012 relative to the 2007 cohort, there are a few statistically significant differences: the coefficients on all immigrant and primarily operated by the wife are negative and significant in 2012; and the coefficients on two unrelated owners, at least three unrelated owners, franchise, and start-up capital 1m and more are still positive but smaller in magnitude.

Multicollinearity is another issue for interpreting the results. The unconditional results in Table 2 have broadly similar patterns to the conditional results in Table 3, but there are some differences as noted above. A more important issue is the possibility that some variables are jointly determined with firm size and growth, and for this reason we present the base specification that excludes financing and industry and show that indeed other variables have different estimated coefficients depending on whether financing and industry are included. Among the variables in the base specification, we identify founding team variables and time between initial start-up and hiring the first employee as those most likely to be problematic. We estimate regressions excluding these variables from the base specification and find qualitatively similar results.[34]

Our definition of business entry is based on the first quarter with paid employees, but we have also taken into account information on activities prior to hiring, in two different ways. First, the SBO asks each owner to report when the business was "established" (with no further definition). Second, we use information from the Census Bureau on businesses reporting revenue but no employees. From each of these, we construct variables capturing activity prior to hiring. Including these variables in the regressions had only negligible impact on the other results.

We also worried that results might be sensitive to the choice of the threshold for defining high growth. The estimates presented above use a top 5 percent threshold, but we have also run regressions using top 2 percent and top 10 percent thresholds, with results available upon request. The general patterns are very similar, but the magnitudes of several of the effects are monotonically increasing as the threshold rises from the top 10 to 5 to 2 percent, including the negative all female (at firm age 0), veteran, gender diversity, and ethnic diversity (at firm age 0) effects and the positive owner age (at firm age 0), bachelor's and graduate degree, prior business, two and three or more unrelated owners, and start-up finance effects.

A related issue is that our analysis includes exiting firms in the calculation of the top 5 percent at age 7, but because of the high exit rate, the firms in the top 2 percent of this distribution at age 7 are roughly the same as the top 5 percent of survivors.

Other robustness checks include the following: We have estimated logistic regressions in place of linear probability models. Since there is some question whether firms with very high employment when they first appear are really start-ups, we have run regressions dropping all firms

[34] Since we omit founding team variables like couples and different types of diversity in these specifications, we control for owner-level gender, ethnicity, race, and immigrant variables rather than firm-level variables indicating whether all of the owners are in the particular category.

with 100 or more employees in their first quarter.[35] We have estimated regressions without adjusting for boundary changes. We have also estimated regressions with owner share weights, but not LBD weights, to examine the sensitivity of the results to LBD weighting. The results from all of these variants are qualitatively similar to those in Table 3, and they are available upon request.

4. Conclusion

Motivated by previous research findings on the predominance of job creation among entrants and young firms, this paper has analyzed high-growth entrepreneurship using a unique data set. The data are based on a large, random, and nationally representative source that permits us to follow an entry cohort over time, dealing with the challenges of defining when the firm starts and tracking it over time. The data we have constructed are also unusual in containing information immediately after the firm starts up, including a rich set of variables on founder characteristics.

Our empirical results confirm the finding in previous research of large skewness in the employment size distribution, whereby a small fraction of firms account for most employment in the U.S. economy. We add to this result by documenting the importance of high-growth entrepreneurship both at entry and at age 7 and the high persistence of size over the first 7 years of the firm's life. We also show that not only are large entrants most likely to be large firms at age 7, but also that among growing firms, the average job creation is largest among large entrants. Large entrants also destroy the most jobs on average, so the "up or out dynamic" of Decker et al. (2014) is strongest for this group. By contrast, small entrants are most likely to remain small, conditional on survival, and are unlikely to grow, consistent with Hurst and Pugsley (2011). Exit rates are high across the board, only declining slightly with start-up size. We also find that most of the size variation is within rather than between narrow (6-digit) industries and little is explained by differences in the amount of start-up capital.

These results motivate our detailed examination of founder characteristics that may predict high growth, here defined as the top ventile of the size distribution at age 0 and age 7. Controlling in a base specification for other aspects of demographic, human capital, and founding team characteristics, but not for start-up finance and industry, we find a much lower probability (34 percent) for women to operate a high-growth business. This large gender gap falls to about 20 percent when we control for the amount of start-up finance, which could result from discrimination in financial access or from non-pecuniary motives for founding the business, neither of which we can observe directly. But the gap rises to about 40 percent when we add detailed industry effects, which suggests that women-owned businesses tend to be disproportionately in high-growth sectors. The estimated effects are similar at age 0 and age 7.

By contrast with the large and statistically significant gender gap, we generally find only modest differences with respect to race, ethnicity, and nationality. Unconditional odds ratios do imply a lower probability of operating a high-growth firm for Hispanic, African-American, and Asian owners, but coefficients for these categories are generally insignificant when we control for other variables in the base specification. The major exception is African-American owners at firm age 0, who have a 28 percent lower probability of entering in the largest 5 percent of businesses, but this difference disappears by firm age 7. When the amount of start-up capital is controlled the

[35] Firms with 100 or more employees are 0.5 percent of the firms in the sample in Tables 2 and 3.

racial gap also falls and becomes statistically insignificant at age 0, and it moves in a positive direction at age 7, a result which is consistent with financial discrimination at start-up. But it rises in magnitude when detailed industry controls are added, implying that African Americans select into industries with higher shares of large entrants. We find no significant differences for immigrants in any specification, a finding at odds with some popular beliefs and prior research on immigrant entrepreneurship (e.g., Kerr and Kerr 2017).

Concerning the age of the founder(s), the high-growth probability is clearly positively sloped at firm age 0, but then flattens by the time firms reach age 7. The positive profile at age 0 might be explained by lower skills or greater financial constraints faced by younger entrepreneurs, and consistent with the latter we find that controlling for start-up finance yields a flatter founder age profile at that firm age. The flat profile in the base specification at age 7 might be explained by leveling out of the financial constraints as the firm ages, but when we control for start-up finance, the slope of the founder age profile becomes negative, implying that the effect of any tougher financial constraints for younger entrepreneurs tends to persist. A negative slope is inconsistent with general labor market experience playing an important role in entrepreneurial human capital.

With respect to formal education, a striking finding at firm age 0 is that bachelor's degrees are associated with a much higher (31 percent) probability of large size than either high school or graduate degrees. The difference vis-à-vis high school largely persists to age 7, but it becomes negligible with respect to graduate education. Controlling for start-up finance actually yields a significantly lower probability of high growth for those with graduate versus only high-school education, although again this disappears 7 years later. Controlling for detailed industry raises the graduate education coefficient, implying that this group tends to choose sectors with relatively small firms, perhaps because many of them work in professions such as law, medicine, or accounting. In any case, our results do not support an important role for graduate education in producing high-growth entrepreneurs. Perhaps less surprising, we find that military experience is negatively associated and prior business ownership is positively associated with high growth, in both cases strong results that are robust over time and across specifications.

Finally, concerning founding teams we find that businesses with more founders are more likely to be large at start-up and subsequently. The differences are larger than any others in the data: in the baseline specification, firms with at least three founders are 230 percent more likely and those with two founders other than a couple are about 100 percent more likely to be in the top 5 percent than single owners. Diversity in age, race/ethnicity, and immigration status have little association with the high growth probability, while gender diversity is negatively associated.

Some of these results contradict patterns that were previously reported by other researchers in related research. The differences may be explained by sample sizes and representativeness, focus on averages versus high growth, definitions of high growth, or analyzing a cross-section versus following an entry cohort. In other cases, the results in this paper confirm previous research, putting them on a more secure footing. A series of robustness checks confirm that our findings are not sensitive to small changes in specification or in the definition of high growth.

A general pattern worthy of note is that not only is firm size highly persistent from entry to age 7, but also so are the relationships of high growth with characteristics. There is a general tendency for an attenuation of coefficients from age 0 to age 7, suggesting increasing difficulty in accounting for growth heterogeneity. But except in a few cases we have noted, the qualitative patterns are similar, and most of the differences are statistically insignificantly different from zero at conventional levels. Thus, the patterns observed at age 0 already embody most of what one can learn 7 years later.

On the other hand, results are frequently sensitive to controlling for the amount of start-up finance and for the industry in which the firm operates. We have argued that these variables are particularly suspect for the possibility of correlation with important unobservables such as motivations and skills of the entrepreneur and the quality of the business idea, so we have excluded them from our base specification, but in several cases adding them to a richer specification is helpful in illuminating the patterns with respect to other variables. Future research could fruitfully focus on these relationships.

References

Acemoglu, Daron, Ufuk Akcigit, Nicholas Bloom, and William R. Kerr, 2013, "Innovation, Reallocation, and Growth," NBER Working Paper No. 18993.

Acs, Zoltan J., William Parsons and Spencer Tracy, 2008, "High-Impact Firms: Gazelles Revisited," Washington D.C.: Small Business Administration.

Baptista, Rui, Murat Karaoz, and Mendonca, 2014, "The Impact of Human Capital on Early Success of Necessity vs. Opportunity-Based Entrepreneurs," *Small Business Economics,* 42(4), 831-847.

Bates, Timothy, 1990, "Entrepreneur Human Capital Inputs and Small Business Longevity," *Review of Economics and Statistics*, 72(4), 551-559.

Birch, David L., 1979, *The Job Generation Process*, unpublished report prepared by the MIT Program on Neighborhood and Regional Change for the Economic Development Administration, U.S. Department of Commerce, Washington, DC.

Birch, David L., 1981, "Who Creates Jobs?" *The Public Interest*, 65, 3-14.

Birch, David L., 1987, *Job Creation in America: How Our Smallest Companies Put the Most People to Work*, Free Press, New York.

Blanchard, Lloyd, Bo Zhao, and John Yinger, 2008, "Do Lenders Discriminate against Minority and Woman Entrepreneurs?" *Journal of Urban Economics*, 63(2), 467–97.

Blanchflower, David G., Phillip B. Levine, and David J. Zimmerman, 2003, "Discrimination in the Small-Business Credit Market," *The Review of Economics and Statistics* 85(4), 930–43.

Brannon, David L., Johan Wiklund, and J. Michael Haynie, 2013, "The Varying Effects of Family Relationships in Entrepreneurial Teams," *Entrepreneurship Theory and Practice*, 37(1), 107-132.

Brown, Charles, J. Hamilton, and James Medoff, 1990, *Employers Large and Small*, Harvard University Press: Cambridge, MA.

Brown, J. David, and John S. Earle, 2017, "Finance and Growth at the Firm-Level: Evidence from SBA Loans," *Journal of Finance,* 72, 1039–1080.

Brown, J. David, John S. Earle, and Dana Lup, 2005, "What Makes Small Firms Grow? Finance, Human Capital, Technical Assistance, and the Business Environment in Romania," *Economic Development and Cultural Change*, 54(1), 33-70.

Cabral, Luís M. B., and José Mata, 2003, "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, 93(4), 1075–90.

Coad, Alex, Julian S. Frankish, Paul Nightingale, and Richard G. Roberts, 2014, "Business Experience and Start-up Size: Buying More Lottery Tickets Next Time Around?" *Small Business Economics*, 43, 529-547.

Coleman, Susan, and Alicia Robb, 2014, "Access to Capital by High-Growth Women-Owned Businesses," report SBAHQ-13-Q-0A63 for the National Women's Business Council.

Daunfeldt, Sven-Olov, and Daniel Halvarsson, 2015, "Are High-Growth Firms One-Hit Wonders? Evidence from Sweden," *Small Business Economics,* 44, 361-383.

Daunfeldt, Sven-Olov, Dan Johansson, and Daniel Halvarsson, 2015, "Using the Eurostat-OECD Definition of High-Growth Firms: A Cautionary Note," *Journal of Entrepreneurship and Public Policy*, 4(1), 50-56.

Davis, Steven J., John Haltiwanger, and Scott Schuh, 1996, *Job Creation and Destruction*, MIT Press.

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.

________________, 2016, "Where Has All the Skewness Gone? The Decline in High Growth (Young) Firms in the U.S.," *European Economic Review*, 86, 4-23.

Dunne, Paul, and Alan Hughes, 1994, "Age, Size, Growth and Survival: UK Companies in the 1980s," *Journal of Industrial Economics*, 42, 115–140.

Eurostat-OECD, 2008, *Eurostat-OECD Manual on Business Demography Statistics*, OECD Publishing: Paris, France.

Evans, David S., 1987, "The Relationship Between Firm Growth, Size and Age: Estimates for 100 Manufacturing Industries," *Journal of Industrial Economics*, 35, 567–581.

Fairlie, Robert W., and Alicia M. Robb, 2007, "Why Are Black-Owned Businesses Less Successful than White-Owned Businesses? The Role of Families, Inheritances, and Business Human Capital," *Journal of Labor Economics*, 25(2), 289–323.

Fairlie, Robert W., and Alicia M. Robb, 2009, "Gender Differences in Business Performance: Evidence from the Characteristics of Business Owners Survey," *Small Business Economics*, 33(4), 375-395.

Fairlie, Robert, Alicia Robb, and David T. Robinson, 2016, *Black and White: Access to Capital among Minority-Owned Startups*, SIEPR Working Paper No. 17-03.

Foster, Lucia, Cheryl Grim, and John Haltiwanger, 2016, "Reallocation in the Great Recession: Cleansing or Not?" *Journal of Labor Economics*, 34(1), S293-S331.

Frank, Murray Z. 1988. “An Intertemporal Model of Industrial Exit,” *The Quarterly Journal of Economics*, 103(2), 333–344.

Garnsey, Elizabeth, Erik Stam, and Paul Heffernan, 2006, “New Firm Growth: Exploring Processes and Paths,” *Industry and Innovation,* 13(1), 1-20.

Geroski, Paul, 1995, “What Do We Know About Entry?” *International Journal of Industrial Organization,* 13(4), 421-440.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda, 2013, “Who Creates Jobs? Small versus Large versus Young,” *Review of Economics and Statistics*, 95(2), 347–361.

Headd, Brian, and Radwan Saade, 2008, “Do Business Definition Decisions Distort Small Business Research Results?” SBA Office of Advocacy Working Paper.

Holzl, Werner, 2014, “Persistence, Survival and Growth: A Closer Look at 20 Years of Fast-Growing Firms in Austria,” *Industrial and Corporate Change*, 12(1), 199-231.

Hoogendoorn Sander, and Mirjam van Praag, 2012, “Ethnic Diversity and Team Performance: A Field Experiment,” Tinbergen Institute Discussion Paper TI 2012-068/3.

Hoogendoorn, Sander, Hessel Oosterbeek, and Mirjam van Praag, 2013, “The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment,” *Management Science*, 59(7).

Hopenhayn, Hugo A., 1992, “Entry, Exit, and Firm Dynamics in Long Run Equilibrium,” *Econometrica,* 60 (5): 1127–1150.

Hurst, Erik, and Benjamin Wild Pugsley, 2011, “What Do Small Businesses Do?” *Brookings Papers on Economic Activity*, Fall 2011, 73-118.

Hurst, Erik G., and Benjamin W. Pugsley, 2017, “Wealth, Tastes, and Entrepreneurial Choice,” in *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar, eds.

Jarmin, Ron, and C.J. Krizan, 2010, “Past Experience and Future Success: New Evidence on Owner Characteristics and Firm Performance,” CES Discussion Paper 10-24.

Jarmin, Ron, C.J. Krizan, and Adela Luque, 2014, “Owner Characteristics and Firm Performance During the Great Recession,” CES Discussion Paper 14-36.

Jovanovic, Boyan, 1982, “Selection and the Evolution of Industry,” *Econometrica*, 50(3), 649–670.

Kalleberg, Arne L., and Kevin T. Leicht, 1991, “Gender and Organizational Performance: Determinants of Small Business Survival and Success,” *The Academy of Management Journal*, 34(1), 136–161.

Kerr, Sari Pekkala, and William R. Kerr, 2017, “Immigrant Entrepreneurship,” in *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar, eds., University of Chicago Press.

Lazear, Edward P., 2004, “Balanced Skills and Entrepreneurship,” *American Economic Review*, 94, 208-211.

Lazear Edward P., 2005, “Entrepreneurship,” *Journal of Labor Economics*, 23, 649-680.

McKelvie, Alexander, and Johan Wiklund, 2010, “Advancing Firm Growth Research: A Focus on Growth Mode Instead of Growth Rate,” *Entrepreneurship Theory and Practice*, 34(2), 261–288.

Melitz, Marc J., 2003, “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71(6), 1695–1725.

Neumark, David, Brandon Wall, and Junfu Zhang, 2011, “Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series,” *Review of Economics and Statistics*, 93(1), 16-29.

Nightingale, Paul, and Alex Coad, 2014, “Muppets and Gazelles: Political and Methodological Biases in Entrepreneurship Research,” *Industrial and Corporate Change*, 23(1), 113-143.

OECD, 2010, *High-Growth Enterprises: What Governments Can Do to Make a Difference*, OECD Studies on SMEs and Entrepreneurship, OECD Publishing.

Robb, Alicia M., and Robert W. Fairlie, 2009, “Determinants of Business Success: An Examination of Asian-Owned Businesses in the US,” *Journal of Population Economics*, 22, 827-858.

Ruef, Marin, Howard E. Aldrich, and Nancy M. Carter, 2003, “The Structure of Founding Teams: Homophily, Strong Ties, and Isolation Among U.S. Entrepreneurs,” *American Sociological Review*, 68(2), 195-222.

Shane, Scott, 2008, *The Illusions of Entrepreneurship: The Costly Myths That Entrepreneurs, Investors, and Policy Makers Live By,* New Haven: Yale University Press.

Shaw, Kathryn L., and Anders Sorensen, 2017, “The Productivity Advantage of Serial Entrepreneurs,” NBER Working Paper No. 23320.

Stangler, Dane, 2010, “High-Growth Firms and the Future of the American Economy,” Kauffman Foundation Research Series: Firm Formation and Economic Growth.

Storey, D. J., 1994, "New Firm Growth and Bank Financing," *Small Business Economics,* 6(2), 139–150.

U.S. Census Bureau, 2016, Business Dynamics Statistics: Firm Characteristics Data Tables – Firm Age, https://www.census.gov/ces/dataproducts/bds/data_firm.html, accessed on September 6, 2017.

Table 1.1 Employment Category Transition Matrices from Age 0 to Age 7: Row Percent

		Age 7								
		0	1	2-4	5-9	10-19	20+	Column Total	Age 0 Emp Share	Age 7 Emp Share
Age 0	1	67.3	16.5	10.8	3.4	1.3	0.8	44.7	7.1	15.9
	2-4	61.8	6.6	18.2	8.4	3.2	1.8	33.4	13.9	23.1
	5-9	59.1	2.4	9.5	16.0	8.9	4.1	11.9	12.2	15.0
	10-19	57.3	1.8	4.0	9.2	17.1	10.7	6.1	12.7	13.0
	20+	56.5	0.6	1.6	2.2	7.6	31.4	4.1	54.1	33.0
	Row Total	63.4	10.0	12.3	6.9	4.1	3.3	100.0		100.0
Age 7 Emp Share		0.0	2.8	9.4	12.6	15.2	60.1		100.0	

Employment in the start quarter and the same quarter at age 7 are from the Business Register (BR), and firms are tracked over time using the Longitudinal Business Database (LBD). The sample is all firms starting in one of the quarters of 2007, and the sample size is about 603,000. Each cell represents the percentage of firms in the age 0 size category in the particular row that transition to the age 7 size category in the column. The Age 0 and Age 7 shares are the age 0 size category's percent of employment at age 0 and age 7, respectively.

Table 1.2 Employment Category Transition Matrices from Age 0 to Age 7: Column Percent

		Age 7						
		0	1	2-4	5-9	10-19	20+	Total
Age 0	1	47.4	73.8	39.1	22.2	14.7	10.2	44.7
	2-4	32.5	22.0	49.3	40.8	26.5	17.9	33.4
	5-9	11.1	2.9	9.2	27.6	25.8	14.4	11.9
	10-19	5.5	1.1	2.0	8.1	25.4	19.3	6.1
	20+	3.6	0.2	0.5	1.3	7.6	38.2	4.1

Employment in the start quarter and the same quarter at age 7 are from the Business Register (BR), and firms are tracked over time using the Longitudinal Business Database (LBD). The sample is all firms starting in one of the quarters of 2007, and the sample size is about 603,000. Each cell represents the percentage of firms in the age 7 size category in the particular column that have transitioned from the age 0 size category in the row.

Table 1.3 Average Jobs Gained/Lost Per Firm in Category

		$Emp_7=0$	$Emp_7<Emp_0$	$Emp_7=Emp_0$	$Emp_7>Emp_0$
Age 0	1	-1.0	N.A.	0.0	5.8
	2-4	-2.6	-2.4	0.0	8.3
	5-9	-6.5	-5.7	0.0	12.0
	10-19	-13.2	-11.2	0.0	17.6
	20+	-100.4	-78.5	0.0	52.3

Employment in the start quarter and the same quarter at age 7 are from the Business Register (BR), and firms are tracked over time using the Longitudinal Business Database (LBD). The sample is all firms starting one of the quarters of 2007, and the sample size is about 603,000. Emp_0 is employment at age 0, and Emp_7 is employment at age 7.

Table 2.1 Descriptive Statistics for Independent Variables:
Percent by Gender, Age, Ethnicity, Race, and Immigrant

	All	Age 0		Age 7		Age 0	Age 7
Founder characteristic		Top 5%	Lower 95%	Top 5%	Lower 95%	Top 5%/All	
All female	16.0	8.2	16.5	8.5	16.4	0.51	0.53
All male	51.4	60.2	50.9	60.6	50.9	1.17	1.18
Age <35	19.6	11.9	20.0	19.6	19.6	0.61	1.00
Age 35-44	31.3	29.2	31.4	31.8	31.3	0.93	1.02
Age 45-54	29.3	33.5	29.1	29.0	29.3	1.14	0.99
Age 55-64	15.4	19.0	15.3	15.0	15.5	1.23	0.97
Age 65 or over	4.0	5.9	3.9	4.0	4.0	1.48	1.00
All Hispanic	5.0	3.6	5.1	4.0	5.0	0.72	0.80
All non-Hispanic	93.0	94.1	92.9	93.2	92.9	1.01	1.00
All White	86.4	87.8	86.4	86.9	86.4	1.02	1.01
All African American	2.8	1.5	2.9	2.2	2.8	0.54	0.79
All Asian	4.8	3.7	4.9	4.3	4.8	0.77	0.90
All other minority race	3.3	4.4	3.3	3.6	3.3	1.33	1.09
All immigrant	15.0	14.2	15.1	14.3	15.1	0.95	0.95
All U.S.-Born	80.7	80.1	80.7	79.2	80.8	0.99	0.98

The number of observation is about 3,000 for the top 5% and 52,800 for the bottom 95% at age 0, 3,100 for the top 5% and 52,700 for the bottom 95% at age 7.

Table 2.2 Descriptive Statistics for Independent Variables: Percent by Type of Human Capital

	All	Age 0		Age 7		Age 0	Age 7
Founder characteristic		Top 5%	Lower 95%	Top 5%	Lower 95%	Top 5%/All	
Less than high school	3.5	2.8	3.5	2.0	3.5	0.80	0.57
High school	17.9	15.1	18.1	14.4	18.1	0.84	0.80
Vocational school	5.7	4.1	5.8	4.0	5.8	0.72	0.70
Some college	16.9	15.6	16.9	15.9	16.9	0.92	0.94
Associate degree	6.1	4.3	6.2	5.0	6.2	0.70	0.82
Bachelor's degree	29.2	38.8	28.7	36.2	28.9	1.33	1.24
Graduate degree	19.9	18.1	19.9	21.4	19.8	0.91	1.08
Veteran	9.0	8.1	9.0	7.1	9.1	0.90	0.79
Non-veteran	91.0	91.9	91.0	92.9	90.9	1.01	1.02
Prior business	50.8	67.6	49.9	62.8	50.1	1.33	1.24
No prior business	49.2	32.4	50.1	37.2	49.9	0.66	0.76

The number of observation is about 3,000 for the top 5% and 52,800 for the bottom 95% at age 0, 3,100 for the top 5% and 52,700 for the bottom 95% at age 7.

Table 2.3 Descriptive Statistics for Independent Variables:
Percent by Size and Characteristics of Founding Team

	All	Age 0		Age 7		Age 0	Age 7
Founder characteristic		Top 5%	Lower 95%	Top 5%	Lower 95%	Top 5%/All	
One owner	50.6	29.3	51.8	28.1	51.9	0.58	0.56
Equally operated by couple	12.1	11.0	12.2	11.6	12.2	0.91	0.96
Primarily husband	10.8	7.4	11.0	7.4	11.0	0.69	0.69
Primarily wife	4.3	3.1	4.4	3.2	4.4	0.72	0.74
Family other than couple	8.3	16.4	7.9	16.2	7.9	1.98	1.95
Two unrelated owners	10.5	15.9	10.2	16.3	10.2	1.51	1.55
At least 3 owners	5.9	18.5	5.2	19.1	5.2	3.14	3.24
Multi-generation	2.9	6.1	2.8	6.2	2.8	2.10	2.14
No multi-generation	97.1	93.9	97.2	93.8	97.2	0.97	0.97
Gender diversity	5.4	10.1	5.1	8.7	5.2	1.87	1.61
No gender diversity	94.6	89.9	94.9	91.3	94.8	0.95	0.97
Ethnic diversity	4.4	4.5	4.3	5.5	4.3	1.02	1.25
No ethnic diversity	95.6	95.5	95.7	94.5	95.7	1.00	0.99
Immigrant diversity	4.3	5.7	4.2	6.5	4.2	1.33	1.51
No immigrant diversity	95.7	94.3	95.8	93.5	95.8	0.99	0.98
Franchise	4.2	18.0	3.5	12.8	3.8	4.29	3.05
Non-franchise	95.8	82.0	96.5	87.2	96.2	0.86	0.91

The number of observation is about 3,000 for the top 5% and 52,800 for the bottom 95% at age 0, 3,100 for the top 5% and 52,700 for the bottom 95% at age 7.

Table 2.4 Descriptive Statistics for Independent Variables: Percent by Amount of Start-up Finance

	All	Age 0		Age 7		Age 0	Age 7
Founder characteristic		Top 5%	Lower 95%	Top 5%	Lower 95%	Top 5%/All	
No capital needed	6.9	4.9	7.0	4.1	7.1	0.71	0.59
Capital under 5k	21.1	7.0	21.8	6.7	21.8	0.33	0.32
Capital 5k to 10k	9.7	3.0	10.1	3.0	10.1	0.31	0.31
Capital 10k to 25k	12.8	4.6	13.3	6.4	13.2	0.36	0.50
Capital 25k to 50k	9.8	4.6	10.0	7.2	9.9	0.47	0.73
Capital 50k to 100k	11.0	8.4	11.2	10.1	11.1	0.76	0.92
Capital 100k to 250k	11.3	15.0	11.1	15.5	11.1	1.33	1.37
Capital 250k to 1m	7.8	25.1	6.9	22.9	7.0	3.22	2.94
Capital 1m and more	2.5	15.3	1.8	12.4	2.0	6.12	4.96
Don’t know amount	7.0	12.1	6.7	11.6	6.8	1.73	1.66

The number of observation is about 3,000 for the top 5% and 52,800 for the bottom 95% at age 0, 3,100 for the top 5% and 52,700 for the bottom 95% at age 7.

Table 3.1. LPM Estimates for Top 5% of Employment at Age 0 and Age 7: Gender and Age

	Base		Plus Start-up Fin.		Plus Industry	
Characteristic at Age 0	Top 5% Emp_0	Top 5% Emp_7	Top 5% Emp_0	Top 5% Emp_7	Top 5% Emp_0	Top 5% Emp_7
All female	-1.68	-1.66	-1.09^	-1.17^	-2.04^	-2.53^
	(0.31)	(0.31)	(0.30)	(0.31)	(0.34)	(0.35)
Age 35-44	1.35	-0.16+	0.99^	-0.48+^	1.28^	-0.40+
	(0.30)	(0.36)	(0.29)	(0.35)	(0.29)	(0.35)
Age 45-54	2.10	-0.56+	1.43^	-1.12+^	1.67	-1.13+
	(0.33)	(0.36)	(0.32)	(0.36)	(0.32)	(0.36)
Age 55-64	2.37	-0.92+	1.57^	-1.53+^	1.74	-1.55+
	(0.41)	(0.43)	(0.40)	(0.42)	(0.40)	(0.42)
Age 65 or over	3.43	-0.96+	1.83^	-2.20+^	1.46	-2.45+
	(0.77)	(0.68)	(0.74)	(0.68)	(0.72)	(0.70)
Capital 5k to 10k			-0.28	-0.26	-0.46	-0.36
			(0.28)	(0.27)	(0.29)	(0.28)
Capital 10k to 25k			-0.31	0.44+	-0.68^	0.37+
			(0.26)	(0.29)	(0.28)	(0.30)
Capital 25k to 50k			0.09	1.44+	-0.78^	1.08+^
			(0.35)	(0.41)	(0.37)	(0.41)
Capital 50k to 100k			0.83	1.88+	-0.61^	1.10+^
			(0.40)	(0.45)	(0.42)	(0.46)
Capital 100k to 250k			2.68	3.37	1.04^	2.49+^
			(0.54)	(0.54)	(0.53)	(0.54)
Capital 250k to 1m			11.10	10.29	8.45^	8.78^
			(0.90)	(0.86)	(0.85)	(0.85)
Capital 1m and more			24.58	19.30+	21.50^	17.22+^
			(1.83)	(1.73)	(1.77)	(1.73)
Don't know amount			5.65	5.63	4.14^	4.86^
			(0.62)	(0.65)	(0.61)	(0.63)
No capital needed			1.64	1.30	1.24^	0.94^
			(0.39)	(0.42)	(0.39)	(0.41)
R^2	0.058	0.041	0.099	0.069	0.194	0.139

The number of observations is about 55,800. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables 3.2, 3.3 and 3.4. The omitted categories are under 35 for owner age and less than $5,000 for start-up capital amount. The regressions also include start quarter dummies, and those in the last two columns also include six-digit NAICS industry dummies. + signifies that the coefficient at age 7 is statistically significantly different at the 5% level from the coefficient for the otherwise similar specification for age 0 (column 2 vs. 1, 4 vs. 3, and 6 vs. 5). ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification for this age (column 3 vs. 1, 4 vs. 2, 5 vs. 3, and 6 vs. 4).

Table 3.2. LPM Estimates for Top 5% of Employment at Age 0 and Age 7: Ethnicity, Race, and Immigrant

	Base		Plus Start-up Fin.		Plus Industry	
Characteristic at Age 0	Top 5% Emp_0	Top 5% Emp_7	Top 5% Emp_0	Top 5% Emp_7	Top 5% Emp_0	Top 5% Emp_7
All Hispanic	-0.47	-0.23	0.19^	0.35^	-0.91^	-0.40^
	(0.59)	(0.65)	(0.57)	(0.64)	(0.59)	(0.63)
All African American	-1.40	0.04	-0.80^	0.56^	-1.99^	-1.35^
	(0.61)	(0.70)	(0.60)	(0.69)	(0.59)	(0.67)
All Asian	-0.62	0.00	-0.99^	-0.47^	-3.40^	-1.78^
	(0.74)	(0.78)	(0.73)	(0.77)	(0.78)	(0.82)
All other minority race	0.89	-0.16	0.50	-0.47	-0.02	-1.25^
	(0.95)	(0.85)	(0.94)	(0.86)	(0.93)	(0.85)
All immigrant	0.24	0.15	-0.04^	-0.14^	-0.86^	-0.67^
	(0.49)	(0.48)	(0.48)	(0.47)	(0.48)	(0.48)

The number of observations is about 55,800. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables 3.1, 3.3 and 3.4. The omitted category for race is White. The regressions also include start quarter dummies, and those in the last two columns also include six-digit NAICS industry dummies. + signifies that the coefficient at age 7 is statistically significantly different at the 5% level from the coefficient for the otherwise similar specification for age 0 (column 2 vs. 1, 4 vs. 3, and 6 vs. 5). ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification for this age (column 3 vs. 1, 4 vs. 2, 5 vs. 3, and 6 vs. 4).

Table 3.3. LPM Estimates for Top 5% of Employment at Age 0 and Age 7: Human Capital

	Base		Plus Start-up Fin.		Plus Industry	
Characteristic at Age 0	Top 5% Emp_0	Top 5% Emp_7	Top 5% Emp_0	Top 5% Emp_7	Top 5% Emp_0	Top 5% Emp_7
Less than high school	0.03	-0.91	0.25	$-0.69^{\wedge}$	-0.26	-0.90
	(0.65)	(0.62)	(0.65)	(0.62)	(0.69)	(0.62)
Vocational school	-0.21	-0.14	-0.07	0.02	$0.63^{\wedge}$	0.21
	(0.52)	(0.51)	(0.51)	(0.50)	(0.51)	(0.49)
Some college	0.38	0.79	$0.21^{\wedge}$	0.66	$0.66^{\wedge}$	0.80
	(0.38)	(0.39)	(0.37)	(0.39)	(0.38)	(0.38)
Associate degree	-0.51	0.31	-0.47	0.34	-0.13	0.20
	(0.46)	(0.52)	(0.45)	(0.50)	(0.46)	(0.50)
Bachelor's degree	1.54	1.45	$0.90^{\wedge}$	$0.95^{\wedge}$	$1.78^{\wedge}$	1.15
	(0.36)	(0.36)	(0.36)	(0.36)	(0.38)	(0.38)
Graduate degree	-0.07	1.29^{+}	$-0.96^{\wedge}$	$0.57^{+\wedge}$	$1.18^{\wedge}$	$1.41^{\wedge}$
	(0.37)	(0.39)	(0.37)	(0.39)	(0.47)	(0.48)
Veteran	-1.49	-1.30	$-1.09^{\wedge}$	$-0.99^{\wedge}$	-1.08	-1.16
	(0.41)	(0.39)	(0.39)	(0.38)	(0.38)	(0.37)
Prior business	2.05	1.60	$1.34^{\wedge}$	$0.98^{\wedge}$	$1.08^{\wedge}$	1.01
	(0.24)	(0.25)	(0.24)	(0.24)	(0.24)	(0.24)

The number of observations is about 55,800. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables 3.1, 3.2 and 3.4. The omitted category for owner education is high school diploma or GED. The regressions also include start quarter dummies, and those in the last two columns also include six-digit NAICS industry dummies. + signifies that the coefficient at age 7 is statistically significantly different at the 5% level from the coefficient for the otherwise similar specification for age 0 (column 2 vs. 1, 4 vs. 3, and 6 vs. 5). ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification for this age (column 3 vs. 1, 4 vs. 2, 5 vs. 3, and 6 vs. 4).

Table 3.4. LPM Estimates for Top 5% of Employment at Age 0 and Age 7: Founding Team

	Base		Plus Start-up Fin.		Plus Industry	
Characteristic at Age 0	Top 5% Emp$_0$	Top 5% Emp$_7$	Top 5% Emp$_0$	Top 5% Emp$_7$	Top 5% Emp$_0$	Top 5% Emp$_7$
Equally operated by couple	0.49	1.38	-0.10$^{\wedge}$	0.75$^{\wedge}$	-0.83$^{\wedge}$	0.07$^{\wedge}$
	(0.45)	(0.47)	(0.45)	(0.46)	(0.45)	(0.46)
Primarily husband	-0.14	0.29	-0.26	0.12$^{\wedge}$	-0.02	0.20
	(0.40)	(0.41)	(0.40)	(0.41)	(0.39)	(0.40)
Primarily wife	-0.19	0.26	-0.27	0.04$^{\wedge}$	-1.22$^{\wedge}$	-1.39$^{\wedge}$
	(0.62)	(0.60)	(0.60)	(0.61)	(0.62)	(0.63)
Family other than couple	6.25	7.27	4.21$^{\wedge}$	5.42$^{\wedge}$	3.82	5.13
	(0.74)	(0.80)	(0.71)	(0.77)	(0.71)	(0.75)
Two unrelated owners	4.32	5.02	3.26$^{\wedge}$	3.98$^{\wedge}$	2.64$^{\wedge}$	3.59$^{\wedge}$
	(0.57)	(0.58)	(0.55)	(0.57)	(0.53)	(0.56)
At least 3 owners	11.73	13.16	7.51$^{\wedge}$	9.57$^{\wedge}$	6.46$^{\wedge}$	8.57$^{\wedge}$
	(0.98)	(0.99)	(0.97)	(1.01)	(0.95)	(1.00)
Multi-generation	-0.00	0.79	-0.51	0.32$^{\wedge}$	-0.67	0.11
	(1.26)	(1.28)	(1.23)	(1.24)	(1.22)	(1.24)
Gender diversity	-1.57	-3.61^{+}	-0.94$^{\wedge}$	-3.14$^{+\wedge}$	-2.23$^{\wedge}$	-4.26$^{\wedge}$
	(0.94)	(0.89)	(0.91)	(0.88)	(0.90)	(0.87)
Ethnic diversity	-1.86	-1.15	-1.35$^{\wedge}$	-0.76$^{\wedge}$	-1.46	-0.97
	(0.76)	(0.81)	(0.74)	(0.80)	(0.74)	(0.79)
Immigrant diversity	-0.05	0.43	-0.50$^{\wedge}$	0.06$^{\wedge}$	-0.91	-0.36
	(0.87)	(0.91)	(0.84)	(0.90)	(0.84)	(0.89)
Franchise	16.06	9.27^{+}	12.61$^{\wedge}$	6.20$^{+\wedge}$	9.98$^{\wedge}$	4.66$^{+\wedge}$
	(1.45)	(1.27)	(1.41)	(1.25)	(1.35)	(1.24)

The number of observations is about 55,800. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables 3.1, 3.2 and 3.3. The omitted owner type category is one owner. The regressions also include start quarter dummies, and those in the last two columns also include six-digit NAICS industry dummies. + signifies that the coefficient at age 7 is statistically significantly different at the 5% level from the coefficient for the otherwise similar specification for age 0 (column 2 vs. 1, 4 vs. 3, and 6 vs. 5). ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification for this age (column 3 vs. 1, 4 vs. 2, 5 vs. 3, and 6 vs. 4).

APPENDIX

Table A1.1 Employment Category Transition Matrices from Age 1 to Age 7 LBD 2007 Start-up Cohort: Row percent

		Age 7							
		0	1	2-4	5-9	10-19	20+	Column Total	Age 1 Share
Age 1	0	87.9	5.0	4.4	1.6	0.7	0.4	32.7	0.0
	1	58.9	25.2	12.2	2.7	0.8	0.3	23.0	5.0
	2-4	50.0	9.0	26.7	10.5	2.9	1.0	23.8	13.8
	5-9	46.2	2.8	12.8	22.9	11.5	3.7	10.8	15.2
	10-19	45.4	1.4	4.2	11.3	24.6	13.2	5.7	16.4
	20+	42.6	0.7	1.7	2.5	8.6	43.9	4.0	49.6
	Row Total	63.4	10.0	12.3	6.9	4.1	3.3	100	
Age 7 Emp Share		0.0	2.8	9.4	12.6	15.2	60.1		100

Data source is Business Register (BR) from the U.S. Census Bureau. The sample is all firms starting in 2007, and the sample size is about 603,000, rounded to the nearest 100 for disclosure reasons. Each cell represents transition probability across different size categories between age 1 and age 7. Last column and row represents percent of employment share at age 1 and age 7 respectively.

Table A1.2 Employment Category Transition Matrices from Age 1 to Age 7 LBD 2007 Start-up Cohort: Column percent

		Age 7						
		0	1	2-4	5-9	10-19	20+	Total
Age 1	0	45.3	16.4	11.7	7.8	5.4	4.2	32.7
	1	21.3	58.0	22.8	8.9	4.2	2.1	23.0
	2-4	18.7	21.5	51.6	36.3	16.8	6.9	23.8
	5-9	7.9	3.0	11.3	36.2	30.6	12.1	10.8
	10-19	4.1	0.8	1.9	9.4	34.6	22.7	5.7
	20+	2.7	0.3	0.6	1.4	8.3	52.1	4.0

Data source is Business Register (BR) from the U.S. Census Bureau. The sample is all firms starting in 2007, and the sample size is about 603,000, rounded to the nearest 100 for disclosure reasons. Each cell represents transition probability across different size categories between age 1 and age 7.

Table A2.1 Employment Category Transition Matrices from Age 0 to Age 1: Row percent

		Age 1							
		<=0	1	2-4	5-9	10-19	20+	Column Total	Age 0 Share
Age 0	1	40.0	41.9	14.5	2.4	0.8	0.4	44.7	7.1
	2-4	29.6	11.1	44.7	11.5	2.2	1.0	33.4	13.9
	5-9	23.5	3.1	17.0	40.7	13.0	2.7	11.9	12.2
	10-19	21.4	3.2	4.6	16.4	42.4	12.0	6.1	12.7
	20+	22.3	0.9	2.2	2.8	12.9	59.0	4.1	54.1
	Row Total	32.7	23.0	23.8	10.8	5.7	4.0	100	
Age 1 Emp Share		0.0	5.0	13.8	15.2	16.4	49.6		100

Data source is Business Register (BR) from the U.S. Census Bureau. The sample is all firms starting in 2007, and the sample size is about 603,000, rounded to the nearest 100 for disclosure reasons. Each cell represents transition probability across different size categories between age 0 and age 1. Last column and row represents percent of employment share at age 0 and age 1 respectively.

Table A2.2 Employment Category Transition Matrices from Age 0 to Age 1: Column percent

		Age 1						
		<=0	1	2-4	5-9	10-19	20+	Total
Age 0	1	54.6	81.4	27.3	10.0	6.4	4.9	44.7
	2-4	30.2	16.0	62.7	35.3	12.9	8.2	33.4
	5-9	8.5	1.6	8.5	44.5	26.9	8.1	11.9
	10-19	4.0	0.9	1.2	9.2	44.7	18.3	6.1
	20+	2.8	0.2	0.4	1.0	9.1	60.5	4.1

Data source is Business Register (BR) from the U.S. Census Bureau. The sample is all firms starting in 2007, and the sample size is about 603,000, rounded to the nearest 100 for disclosure reasons. Each cell represents transition probability across different size categories between age 0 and age 1.

Table A3.1. LPM Estimates for Top 5% of Employment at Age 1: Gender and Age

	Base	Plus Start-up Fin.	Plus Industry
Characteristic at Age 0	Top 5% Emp_1	Top 5% Emp_1	Top 5% Emp_1
All female	-1.71	-1.07^	-2.40^
	(0.31)	(0.30)	(0.34)
Age 35-44	0.90	0.52^	0.69
	(0.33)	(0.32)	(0.32)
Age 45-54	0.97	0.24+^	0.40+
	(0.34)	(0.33)	(0.33)
Age 55-64	1.42	0.55+^	0.63+
	(0.43)	(0.41)	(0.41)
Age 65 or over	1.82	0.06+^	-0.19+
	(0.78)	(0.77)	(0.76)
Capital 5k to 10k		-0.32	-0.45
		(0.25)	(0.26)
Capital 10k to 25k		0.03	-0.20
		(0.26)	(0.28)
Capital 25k to 50k		0.56	-0.11^
		(0.35)	(0.36)
Capital 50k to 100k		0.94	-0.13^
		(0.40)	(0.42)
Capital 100k to 250k		2.62	1.45^
		(0.53)	(0.53)
Capital 250k to 1m		11.78	9.85^
		(0.90)	(0.85)
Capital 1m and more		28.04	25.20^
		(1.91)	(1.86)
Don't know amount		5.93	4.65^
		(0.65)	(0.62)
No capital needed		1.85	1.46^
		(0.43)	(0.42)
R^2	0.059	0.109	0.201

The number of observations is about 55,800. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables A3.2, A3.3 and A3.4. The omitted categories are under 35 for owner age and less than $5,000 for start-up capital amount. The regressions also include start quarter dummies, and the one in the last column also includes six-digit NAICS industry dummies. + signifies that the coefficient at age 1 is statistically significantly different at the 5% level from the coefficient for the otherwise similar specification for age 0. ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification.

Table A3.2. LPM Estimates for Top 5% of Employment at Age 1:
Ethnicity, Race, and Citizenship

	Base	Plus Start-up Fin.	Plus Industry
Characteristic at Age 0	Top 5% Emp_1	Top 5% Emp_1	Top 5% Emp_1
All Hispanic	-0.58	0.12^	-0.87^
	(0.61)	(0.59)	(0.61)
All African American	-0.27	0.36^	-1.59^
	(0.72)	(0.72)	(0.71)
All Asian	-0.09	-0.45^	-2.43^
	(0.73)	(0.71)	(0.77)
All other minority race	0.14	-0.30^	-1.20^
	(0.85)	(0.86)	(0.86)
All immigrant	-0.54	-0.84^	-1.54^
	(0.47)	(0.46)	(0.47)

The number of observations is about 55,800. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables A3.1, A3.3 and A3.4. White is the omitted category for race. The regressions also include start quarter dummies, and the one in the last column also includes six-digit NAICS industry dummies. + signifies that the coefficient at age 1 is statistically significantly different at the 5% level from the coefficient for the otherwise similar specification for age 0. ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification.

Table A3.3. LPM Estimates for Top 5% of Employment at Age 1: Human Capital

	Base	Plus Start-up Fin.	Plus Industry
Characteristic at Age 0	Top 5% Emp_1	Top 5% Emp_1	Top 5% Emp_1
Less than high school	-0.02	0.22	-0.18
	(0.65)	(0.65)	(0.69)
Vocational school	-0.40	-0.27	0.28^
	(0.51)	(0.49)	(0.48)
Some college	0.66	0.46^	0.77^
	(0.40)	(0.39)	(0.38)
Associate degree	0.13	0.18	0.25
	(0.51)	(0.49)	(0.49)
Bachelor's degree	1.63	0.93^	1.54^
	(0.37)	(0.36)	(0.38)
Graduate degree	0.61	-0.35^	1.37^
	(0.38)	(0.37)	(0.47)
Veteran	-1.13	-0.69^	-0.65
	(0.44)	(0.42)	(0.40)
Prior business	2.22	1.45^	1.20^
	(0.25)	(0.23)	(0.23)

The number of observations is about 55,800. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables A3.1, A3.2 and A3.4. High school diploma or GED is the omitted category for owner education. The regressions also include start quarter dummies, and the one in the last column also includes six-digit NAICS industry dummies. + signifies that the coefficient at age 1 is statistically significantly different at the 5% level from the coefficient for the otherwise similar specification for age 0. ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification.

Table A3.4. LPM Estimates for Top 5% of Employment at Age 1: Founding Team

	Base	Plus Start-up Fin.	Plus Industry
Characteristic at Age 0	Top 5% Emp_1	Top 5% Emp_1	Top 5% Emp_1
Equally operated by couple	0.03	-0.58^	-1.35^
	(0.45)	(0.44)	(0.44)
Primarily husband	-0.12	-0.25	-0.11
	(0.41)	(0.40)	(0.39)
Primarily wife	0.04	-0.02	-1.38^
	(0.63)	(0.62)	(0.64)
Family other than couple	5.94	3.72^	3.20^
	(0.77)	(0.73)	(0.72)
Two unrelated owners	4.59	3.44^	2.84^
	(0.59)	(0.56)	(0.54)
At least 3 owners	12.60	7.97^	6.55^
	(0.99)	(0.99)	(0.97)
Multi-generation	2.35	1.82	1.65
	(1.43)	(1.37)	(1.34)
Gender diversity	-1.77	-1.05^	-2.36^
	(0.96)	(0.93)	(0.91)
Ethnic diversity	-1.30	-0.75^	-1.01
	(0.82)	(0.80)	(0.78)
Immigrant diversity	-0.52	-1.02^	-1.47
	(0.89)	(0.86)	(0.83)
Franchise	14.91	11.23^	8.86^
	(1.44)	(1.38)	(1.32)

The number of observations is about 55,800. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables A3.1, A3.2 and A3.3. The omitted owner type category is one owner. The regressions also include start quarter dummies, and the one in the last column also includes six-digit NAICS industry dummies. + signifies that the coefficient at age 1 is statistically significantly different at the 5% level from the coefficient for the otherwise similar specification for age 0. ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification.

Table A4.1. SBO 2012 LPM Estimates for Top 5% of Employment at Age 0: Gender and Age

	Base	Plus Start-up Fin.	Plus Industry
Characteristic at Age 0	Top 5% Emp_0	Top 5% Emp_0	Top 5% Emp_0
All female	-2.08	-1.33^	-2.14^
	(0.44)	(0.43)	(0.49)
Age 35-44	0.69	0.37^	0.51
	(0.51)	(0.50)	(0.49)
Age 45-54	2.02	1.23^	1.47
	(0.55)	(0.54)	(0.54)
Age 55-64	1.12	0.29^	0.58
	(0.63)	(0.62)	(0.62)
Age 65 or over	3.55	2.26^	2.11
	(1.08)	(1.06)	(1.03)
Capital 5k to 10k		0.68	0.23^
		(0.46)	(0.49)
Capital 10k to 25k		-0.34	-1.25^
		(0.44)	(0.46)
Capital 25k to 50k		0.39	-1.00^
		(0.56)	(0.59)
Capital 50k to 100k		2.05	0.27^
		(0.73)	(0.73)
Capital 100k to 250k		5.58	2.56^
		(1.01)	(0.97)
Capital 250k to 1m		11.89	8.57^
		(1.63)	(1.51)
Capital 1m and more		15.06+	13.30+^
		(2.87)	(2.89)
Don't know amount		8.07	7.35+^
		(0.86)	(0.83)
No capital needed		1.35	1.48
		(0.64)	(0.63)
R^2	0.039	0.068	0.211

The number of observations is about 21,000. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables A4.2, A4.3 and A4.4. The omitted categories are under 35 for owner age and less than $5,000 for start-up capital amount. The regressions also include start quarter dummies, and the one in the last column also includes four-digit NAICS industry dummies. + signifies that the coefficient with the SBO 2012 cohort is statistically significantly different at the 5% level from the coefficient for an otherwise similar specification for the 2007 cohort. ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification.

Table A4.2. SBO 2012 LPM Estimates for Top 5% of Employment at Age 0: Ethnicity, Race, and Citizenship

	Base	Plus Start-up Fin.	Plus Industry
Characteristic at Age 0	Top 5% Emp_0	Top 5% Emp_0	Top 5% Emp_0
All Hispanic	-0.51	-0.11$^{\wedge}$	-0.19
	(0.70)	(0.70)	(0.69)
All African American	-1.08	-0.29$^{\wedge}$	-1.24$^{\wedge}$
	(0.76)	(0.76)	(0.77)
All Asian	0.67	0.16$^{\wedge}$	-1.53$^{\wedge}$
	(0.89)	(0.87)	(0.89)
All other minority race	-0.02	0.06	0.23
	(0.83)	(0.83)	(0.83)
All immigrant	-2.37^{+}	-2.41^{+}	-2.36^{+}
	(0.55)	(0.55)	(0.56)

The number of observations is about 21,000. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables A4.1, A4.3 and A4.4.White is the omitted category for race. The regressions also include start quarter dummies, and those in the last two columns also include four-digit NAICS industry dummies. $^{+}$ signifies that the coefficient with the SBO 2012 cohort is statistically significantly different at the 5% level from the coefficient for an otherwise similar specification for the 2007 cohort. $^{\wedge}$ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification.

Table A4.3. SBO 2012 LPM Estimates for Top 5% of Employment at Age 0: Human Capital

	Base	Plus Start-up Fin.	Plus Industry
Characteristic at Age 0	Top 5% Emp_0	Top 5% Emp_0	Top 5% Emp_0
Less than high school	-1.18	-0.98	-1.99$^\wedge$
	(0.99)	(1.01)	(1.02)
Vocational school	-1.54	-1.27	-0.43$^\wedge$
	(0.70)	(0.71)	(0.73)
Some college	0.50	0.78$^\wedge$	0.94
	(0.65)	(0.64)	(0.64)
Associate degree	-0.28	0.07$^\wedge$	0.14
	(0.84)	(0.82)	(0.82)
Bachelor's degree	1.73	1.40$^\wedge$	2.37$^\wedge$
	(0.61)	(0.60)	(0.64)
Graduate degree	-0.17	-0.62$^\wedge$	1.14$^\wedge$
	(0.58)	(0.58)	(0.70)
Veteran	-1.00	-0.41$^\wedge$	-0.20
	(0.82)	(0.80)	(0.79)
Prior business	2.88	2.21$^\wedge$	2.11$^+$
	(0.40)	(0.39)	(0.38)

The number of observations is about 21,000. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables A4.1, A4.2 and A4.4. High school diploma or GED is the omitted category for owner education. The regressions also include start quarter dummies, and the one in the last column also includes four-digit NAICS industry dummies. $^+$ signifies that the coefficient with the SBO 2012 cohort is statistically significantly different at the 5% level from the coefficient for an otherwise similar specification for the 2007 cohort. $^\wedge$ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification.

Table A4.4. SBO 2012 LPM Estimates for Top 5% of Employment at Age 0: Founding Team

	Base	Plus Start-up Fin.	Plus Industry
Characteristic at Age 0	Top 5% Emp_0	Top 5% Emp_0	Top 5% Emp_0
Equally operated by couple	0.81	0.48	-0.26^
	(0.98)	(0.96)	(0.95)
Primarily husband	-0.97	-1.27	-1.52
	(0.86)	(0.84)	(0.82)
Primarily wife	-2.74+	-2.91+	-3.01
	(0.71)	(0.71)	(0.71)
Family other than couple	4.09	2.88^	2.46
	(1.15)	(1.12)	(1.08)
Two unrelated owners	2.20	1.16^	1.05
	(1.00)	(0.99)	(0.97)
At least 3 owners	5.31+	3.07+^	2.84+
	(1.76)	(1.77)	(1.73)
Multi-generation	-2.30	-2.97^	-3.29
	(1.75)	(1.71)	(1.67)
Gender diversity	3.00+	3.13+	2.61
	(1.65)	(1.63)	(1.58)
Ethnic diversity	0.36	0.69	0.51
	(1.33)	(1.30)	(1.27)
Immigrant diversity	0.73	0.40	0.24
	(1.50)	(1.48)	(1.46)
Franchise	9.93+	7.45+^	4.56^
	(1.73)	(1.67)	(1.61)

The number of observations is about 21,000. Standard errors clustered by firm are in parentheses. Coefficients and standard errors are multiplied by 100 for ease of reading. Each regression also includes the variables in Tables A4.1, A4.2 and A4.3.The omitted owner type category is one owner. The regressions also include start quarter dummies, and the one in the last column also includes four-digit NAICS industry dummies. + signifies that the coefficient with the SBO 2012 cohort is statistically significantly different at the 5% level from the coefficient for an otherwise similar specification for the 2007 cohort. ^ signifies that the coefficient is statistically significantly different at the 5% level from the coefficient for the previous specification.