The Composition of Firm Workforces from 2006–2022: Findings from the Business Dynamics Statistics of Human Capital Experimental Product

by

Martha Stinson
U.S. Census Bureau

Sean Wang
U.S. Census Bureau

CES 25-20 **April 2025**

The research program of the Center for Economic Studies (CES) produces a wide range of economic analyses to improve the statistical programs of the U.S. Census Bureau. Many of these analyses take the form of CES research papers. The papers have not undergone the review accorded Census Bureau publications and no endorsement should be inferred. Any opinions and conclusions expressed herein are those of the author(s) and do not represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Republication in whole or part must be cleared with the authors.

Papers in the series are posted on www.census.gov/library/working-papers/series/ces-wp.html. For information about the series, contact Sean Wang, Editor, Discussion Papers, U.S. Census Bureau, Center for Economic Studies, 4600 Silver Hill Road, Washington, DC 20233, CES.Working.Papers@census.gov.

Abstract

We introduce the Business Dynamics Statistics of Human Capital (BDS-HC) tables, a new Census Bureau experimental product that provides public-use statistics on the workforce composition of firms and its relationship to business dynamics. We use administrative W-2 filings to combine population-level worker demographic data with longitudinal business data to estimate the demographic and educational composition of nearly all non-farm employer businesses in the United States between 2006 and 2022. We use this newly constructed data to document the evolution of employment, entry, and exit of employers based on their workforce compositions. We also provide new statistics on the interaction between firm and worker characteristics, including the composition of workers at startup firms. We find substantial changes between 2006 and 2022 in the distribution of employers along several dimensions, primarily driven by changing workforce compositions within continuing firms rather than the reallocation of employment between firms. We also highlight systematic differences in the business dynamics of firms by their workforce compositions, suggesting that different groups of workers face different economic environments due to their employers.

* Corresponding author: sean.y.wang@census.gov. We thank Emek Basker, Andrew Foote, Lucia Foster, Nathan Goldschlag, Cheryl Grim, Shawn Klimek, Kevin McKinney, Bitsy Perlman, Joe Staudt, Kirk White, the Business Economics Research Group, and the LEHD Research Group for comments and feedback. We are also grateful to Kim Blair, Marina Krylova, Michael Lavern, and Erin Wrona for programming and disclosure review support. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7508149: CBDRB-FY25-0043, CBDRB-FY25-CES020-002).

1 Introduction

Firm performance is dispersed, and labor markets are segmented. Significant differences in firm productivity, job growth, and survival remain even when comparing firms operating in the same industries, locations, and points in the business cycle (Syverson, 2011; Kim et al., 2024; Foster et al., 2008). Meanwhile, the composition of workers varies across firms due to both economic and non-economic factors (Topel and Ward, 1992; Altonji and Blank, 1999; Bayer et al., 2008; Hellerstein and Neumark, 2008; Ferguson and Koning, 2018; Card et al., 2018; Dostie et al., 2023; Wang, 2021; Karahan et al., 2024; Kline, 2024). Understanding the firm-level relationship between business dynamics and workforce composition is therefore important for understanding both the evolution of the workplace over time and for assessing the contribution of employers to differences in workers' labor market outcomes. Yet, there are few public statistics that provide detailed data on the sorting patterns of workers across individual firms.

In this paper, we introduce the Business Dynamics Statistics of Human Capital (BDS-HC) tables, a new Census Bureau experimental product that begins to fill this gap. These are the first public-use statistics that use W-2 filings to combine longitudinal business data with population-level worker demographic data to classify firms based on the characteristics of their workers. This enables us to describe the relationship between business dynamics and workforce composition for nearly all non-farm employer businesses in the United States. We designed these tables to flexibly document the distribution and evolution of employment growth from 2006 to 2022 for groups of firms defined by their share of workers that belong to each category of the six workforce characteristics that we consider: age, sex, race, ethnicity, nativity, and education. By integrating firm-level data with worker-level data, the BDS-HC provides insight into how worker composition varies by business characteristics like the establishment's sector, state, firm size, and firm age, and it shows the exposure of different types of workers to business outcomes like job growth, start-ups, and exits.

We highlight several findings from these new tables. First, the labor market is systematically segmented by the demographic composition of workers. Across every workforce characteristic that we consider, we find that employers are more homogeneous than would be expected if workers and employers matched randomly. Second, workplaces are substantively different in 2022 than in 2006 in several dimensions. The shift in employment towards firms with older, more racially and ethnically heterogeneous, more multinational, and more college-educated workforces accelerated during the Great Recession and the subsequent recovery; this shift was driven by workforce composition changes within continuing firms rather than the reallocation of employment between firms. Third, labor market segmentation creates consistently different economic environments for different types of workers. Between 2006–2022, the average net job creation rate (controlling for differences in firm size) was as high as 2.7% per year for firms with predominantly college graduates, and as low as -1.9% per year for firms with a high share of workers over age 55. Firms with a lower share of White workers (or, equivalently, a higher share of non-White workers) were especially volatile. Workers were exposed to more job creation from employers due to establishment births, but they were also exposed to more job destruction due to establishment closures. Finally, we highlight the workforce composition at small start-up firms. Relative to small older firms, employment at small start-up firms is skewed towards firms with workers that are younger, more heterogeneous by race, ethnicity, and nativity, but more homogeneous by sex.

A key innovation in our product is the use of administrative W-2 filings for worker-firm linkages.

Linking workers to firms requires a source of administrative, job-level data that records both the identity of the employee and the employer in the employment relationship. By using W-2 filings, we ensure that the universe of firms in the worker-firm data is based upon federal tax data, which also corresponds to the universe of firms underlying all the core and experimental Business Dynamics Statistics (BDS) tables. Furthermore, these W-2 records allow us to define the universe of workers at a firm in a given year as the set of all individuals who received positive W-2 wages. We can link these individuals to population-level demographic data at the Census Bureau that are compiled from administrative SSN applications, Decennial Censuses, and American Community Surveys. As a result, we can assign demographic characteristics to firms that, together, constitute 98% of total employment tracked by the BDS over our sample period.

The structure of the BDS-HC tables matches the structure of the core BDS tables. For each demographic category, we assign an establishment to one of six groups based on the share of the parent firm's overall workforce that belongs to the demographic category: 0–10%, 10–25%, 25–50%, 50–75%, 75–90%, or 90–100%. We compute the total firm, establishment, and employment stock and flow measures following the definitions and tabulation methods from Chow et al. (2021), stratifying by the groups in each workforce composition category. As a result, we can directly compare differences in employer outcomes by firm workforce composition to differences by other firm characteristics included in the BDS, and we can also create tabulations based on their interactions. In the first release, we produce public-use statistics separately for each combination of workforce composition group by firm size category. We then aggregate employment totals across firm size categories to calculate the economy-wide employment statistics reported in this working paper. We also interact the tables with firm age, state, and industry sector to provide more granularity on the characteristics, location, and activities of firms with different types of workforces.

These new tables build on decades of work to develop statistics that track the dynamics of businesses and the dynamics of jobs. The core Business Dynamics Statistics (BDS) tables contain annual statistics on job creation and destruction, entry and exit from the perspective of the employer, without distinguishing between the types of workers at the employer. The Quarterly Workforce Indicators (QWI), created by the Longitudinal Employer Household Dynamics (LEHD) program, contain both employment totals and counts of hires, separations, and overall turnover from the perspective of the worker. Because the unit of observation is an employee-employer match (job), the QWI can add both characteristics of the workers and characteristics of the firm and show how job-level dynamics differ by their associated worker or firm characteristics. However, the QWI tables do not provide any information about the overall workforce composition or dynamics of firms. The BDS-HC product bridges this gap, using matched employee-employer data to provide new details about the workers that constitute the firm's total employment. Thus, from the QWI we learn about outcomes of workers, stratified by characteristics of the firm, and from BDS-HC we learn about outcomes of firms, stratified by characteristics of their workers.

It is important to note that these tables are not designed to provide a causal analysis of either the effects of firm characteristics on worker composition or the effects of workforce characteristics on business outcomes. Instead, they are intended to systematically describe the patterns of sorting between workers and firms in the labor market and the resulting differences in economic environments faced by different groups of workers. We hope these tables and their underlying data will be a useful basis for future work to understand the economic and non-economic mechanisms behind these

patterns.

The rest of the paper proceeds as follows. Section 2 describes the input data for the tables and our linking methodology. Section 3 specifies the definition of each workforce category used for tabulation. Section 4 reports findings on the distribution and evolution of firm employment by workforce composition. Section 5 decomposes the exposure of different worker demographic groups to business dynamics, including job creation, destruction, and business exit. Section 6 explores the demographic composition and employment dynamics at small start-ups firms. Finally, Section 7 concludes with some directions for future work.

2 Data

This BDS experimental product brings together three different types of data. First, we use longitudinal data on employment at establishments and firms. These data allow us to calculate changes in employment and business entry and exit over time, and they are the source of all business outcomes in our tables. Next, we use demographic data on people to describe the characteristics of workers. Finally, we use data on worker-firm matches (jobs) to link people to their employers at the firm level. We combine these three sources of data to classify firms by the share of each demographic group in their workforce and calculate the business dynamics of firms in each classification group. In this section, we describe each type of data and its source, explain the worker-firm linking process, and present some statistics about the coverage of our data linkages.

2.1 Business data

Our tabulations of business dynamics are based on data from the Longitudinal Business Database (LBD), an establishment-level time series that spans from 1976–2022. The LBD is created by linking annual snapshots of the Census Bureau's Business Register (BR), a nearly complete list of non-government, non-farm business establishments operating in the United States, where an establishment is defined as a business entity engaging in a specific type of activity at a specific address. The BR is created from a combination of survey and tax data and serves as both a sampling frame for business surveys and a data source for Census Bureau data products. The LBD creates longitudinal identifiers to track individual business establishments over time and firm identifiers to link establishments that are owned by a common entity. Using the establishment identifiers, we can measure establishment age and employment changes from year to year. Crucial to our linking strategy, the BR and LBD contain Employer Identification Numbers (EINs), which are used to file tax returns. The LBD assigns an EIN to each establishment. Firms with multiple establishments may have multiple EINs covering different establishments within the firm. The LBD is updated each year to implement linking and processing improvements and to add another year of data. See Chow et al. (2021) for details on the creation of the LBD.

The LBD is the source data for all BDS tabulations. We currently use the 2022 vintage, the latest available. The core BDS consists of tables that sum year-to-year job creation and destruction (i.e., employment increases or decreases) by individual establishments within groups defined by characteristics of either the establishment, the firm, or both. In addition, BDS tables report establishment, firm, and employment totals, splitting establishments into ones that enter, exit, or continue relative to the prior year. Job creation and destruction are likewise split by whether the

establishment is creating or destroying jobs by entering or exiting, or whether the establishment is simply growing or shrinking. Finally, we define net job creation for a group of establishments as the difference between employment growth and employment decline. We calculate a growth rate by dividing net jobs added by average employment for the group of firms between years t and t+1, calling this the net job creation rate (NJCR).[1]

To group establishments by firm characteristics, we assign all establishments in the firm to a category based on some shared trait. The two main firm characteristics are firm size and age. Firm size is the sum of employment at all establishments owned by the firm. Firm age begins at zero in the year that the first establishment of the firm is born and then accrues by one for each year thereafter. The innovation of the BDS-HC tables is to create new firm characteristics based on the demographics of all the workers employed by the firm and to stratify the standard BDS tables by these characteristics.

2.2 Demographic data

We obtain information on the characteristics of workers from a demographics file assembled by the LEHD program. The Individual Characteristics File (ICF) combines data from a variety of sources, including the Census Numident file, the 2000 Decennial Census long and short forms, the 2010 Decennial Census short form, and the ACS (2001-present).[2] People are identified by a PIK, a unique identifier that tracks an individual across employers and across data sources (see Wagner and Lane (2014) for details on PIK assignment). Some individuals have missing data for one or more characteristics, either because they were not found in the relevant source file or because the characteristic was missing in the source file. Around 5% of individuals have missing sex, age, and place of birth (sourced predominantly from the Numident), about 20% are missing race and ethnicity (sourced from either the short form Census or the ACS), and around 85% are missing education (sourced from the 2000 Census long form or the ACS). When multiple sources of data exist for an individual, the ICF uses the most recent record for each demographic characteristic, so the only time-varying worker characteristic is worker age. For more details on the LEHD ICF, see Graham et al. (2022).

Education has an exceptionally high missing rate because education information is only available from the 2000 Decennial Census long form and the ACS surveys, and because we only use responses when the respondent is at least 25 years old at the time of the survey to ensure we are usually observing completed education. As a result, the share of workers with a valid education response depends heavily on the workers' ages. Figure 1 plots the share of individuals in the demographics file who have a valid survey response on educational attainment. Individuals who were born in 1998 or later were below the age threshold for every year of the ACS and the 2000 Decennial Census and would not have a valid educational response in our data. Individuals born between 1976 and 1997 would be covered by later waves of the ACS, but they would be too young to have a valid educational response in the earlier waves of the ACS or the 2000 Decennial. The share of individuals who have a valid education response declines at a consistent rate across birth years, which is consistent with

[1] See `https://www.census.gov/programs-surveys/bds.html` for formal definitions of BDS statistics and a full list of publicly available tables.

[2] Although some versions of the ICF only include workers who held a UI-eligible job in one of the state-years covered by the LEHD program, we use a demographics file that includes all individuals who were ever covered by one of the component demographics files, regardless of their work status.

the sampling design of the ACS.[3]

To ensure that our educational measures are not confounded by changes in data coverage due to sampling issues, we only use education responses collected from workers who were born before 1976. In principle, the workers in this group have consistent coverage in educational attainment data over our time series. However, in practice, Figure 1 shows that some differences in data coverage by birth year remain for workers born before 1976 due to potentially non-random differences in survey response, sampling coverage, or entry or exit from the sampling frame (e.g., due to migration or mortality).

2.3 Linking employers to workers

To combine data on firms and people, we use administrative IRS W-2 records. These tax forms are filed annually by employers to report federal income tax and social security tax withholding for individual workers. These records contain both PIKs and EINs, allowing us to link our worker characteristic file to our LBD business data. These linked employer-employee data define the universe of workers who were employed at each firm in the LBD. The key advantage of using W-2 data in this setting is that they correspond to the primary source of employer data in the LBD. The LBD uses annual and quarterly federal payroll tax filings (IRS Forms 941, 943, 944), in conjunction with Census surveys, to track establishment- and firm-level employment over time. These filings are required for all employers who pay wages subject to federal income tax withholding or social security and Medicare taxes. Employers are similarly required to file individual W-2 forms for all employees who have either income tax, social security, or Medicare tax withholding. Moreover, employers are instructed to use the same EIN to identify themselves on the employee-level W-2 filings and the employer-level employment tax returns (Internal Revenue Service, 2025). As a result, the universe of firms in the W-2 data closely corresponds to the universe of firms in the LBD data.

Table 1 summarizes the average share of employment, payroll, and establishments that can be matched at different levels. The W-2 data identifies employers at the EIN-level and do not separately track worker employment or earnings at individual establishments. On the other hand, the LBD tracks businesses at the establishment-level and also records the EIN and parent firm associated with each establishment each year. However, there is some ambiguity in the organization of establishments under different EINs in multi-unit firms within the LBD. As a result, for some establishments in the LBD, their EINs do not match any W-2 filings because another EIN at the firm was used to file W-2s for those workers. Thus, we create a firm-level W-2 match by pooling all workers associated with any EINs belonging to the firm.[4] As shown in row 1, accounting for these firm-level linkages allows us to match 96.8% of LBD establishments, accounting for 98.3% of employment and 98.6% of payroll. Alternatively, if we had restricted ourselves to direct EIN matches, we would only have matched 95.0% of establishments, 94.8% of employment, and 94.6% of payroll.

To assess the potential share of employers that may be unmatched due to out-of-date EINs, we

[3]The ACS sampling rate is lower in the first few years of its introduction, so the difference in educational data availability within the 1976-1980 birth cohorts is correspondingly smaller.

[4]In some cases, an EIN may be associated with multiple parent firms in the LBD. These cases can arise for a number of reasons. For example, multiple business units may operate independently under the same EIN, and they were assigned different firm-level identifiers by the Business Register for firm-level surveys. These cases may also arise in error if an out-of-date or incorrect EIN was assigned to an establishment. To avoid potentially assigning a small number of EIN-PIK matches to a large number of firms, we assign each EIN to its modal firm in each given year based on the total employment, payroll, and count of establishments associated with the EIN in the LBD.

also tested a strategy to link the LBD to the W-2 data using any EINs that were assigned to an establishment in the 5 years before or after the target year. These match rates are shown in the third row of Table 1. While this method did find more matches than the EIN-only links, the increase in match rates is small relative to the increase from using same-year, firm-level links. In addition, this strategy introduces potential ambiguity in the timing of worker matches to the firm and differentially lower match rates for the beginning and end years of our time series (since we would not have the data to look both forward and backward for those years).

Based on our findings in Table 1, we link employers to workers at the firm-level by aggregating W-2 filings to the firm-level based on EIN-to-firm linkages from the LBD and assign each establishment in the LBD the workforce composition of the parent firm. This is equivalent to creating an employment-weighted average of the shares at individual establishments. We assess the quality of these matches by comparing the firm totals in each dataset. The total firm payroll from the establishment-level LBD data and the worker-level W-2 are closely related, which further supports the correspondence between employers in the two datasets and our linking approach. Table 2 reports regression estimates of the relationship between the total log payroll and employment between the two datasets for the approximately 85 million firm-year observations in our matched data. The two measures of payroll are extremely similar, which is reassuring since the concept of payroll is the same in both sources (total dollar amount paid to workers in a year). The estimated coefficient of the bivariate regression in Column 1 is close to one, and the adjusted R-squared is also close to one, indicating that the variance of any remaining differences due to rounding, measurement error on either side, or error in associated EINs is generally small. We add firm fixed effects and year fixed effects to the regression in Column 2 to assess the within-firm relationship between the two sets of measures over time. The estimated coefficient falls slightly, as would be expected given that the inclusion of fixed effects would amplify any attenuation due to classical measurement error in the explanatory variable. However, both the estimated coefficient and the adjusted R-squared remain large.

The regressions in Columns 3 and 4 mirror the regressions in Columns 1 and 2, but use employment rather than payroll as the focal variable in both datasets. The concept of employment in the two datasets differs. Employment in the LBD is the number of workers working at the firm during the pay period covering March 12th, whereas employment in the W-2 data is the total number of workers who received pay from the firm at any point in the year. Correspondingly, the relationship between the two employment measures is lower than the two payroll measures, but strongly positive nevertheless.

Finally, we assess whether we are differentially matching employers by year or by size. Figure 2 reports the share of each dataset matched under our firm-level links separately for each year. Between 2005 and 2021, the overall share of total employment or payroll matched in either the LBD or the W-2 is roughly constant. Meanwhile, the share of LBD establishments matched increases over time, and the share of W2 EINs matched drops between 2007 and 2008 and remains flat after 2009. The share of matched employers in the LBD drops in 2022, due to lags in the receipt of W-2 records in the last year of the data. Our main tables only use data from the last year to assign demographics for new, start-up firms. Hence, data for those firms should be interpreted with this difference in representativeness in mind.

Figure 3 plots the average log payroll and log employment for firms in the matched data separately

by each of 40 equal-sized bins in the firm payroll and employment distributions. For payroll, whose definition is consistent in the two datasets, we find that the average LBD total for each bin is nearly identical to the average W-2 total. This suggests that the quality of our firm matches is similar for large and small firms. For employment, the average LBD total in each bin is consistently below the W-2 total, which is consistent with the more restrictive point-in-time definition of employment in the LBD. The relative difference between W-2 and LBD employment is higher for larger firms. Given that the two measures of payroll are similar for those firms, the differences in employment may be driven by higher levels of turnover or seasonal employment at larger firms.

3 Defining workforce composition

To classify firms by their workforce composition, we apply a single methodology to each of the six demographic characteristics that we consider: age, sex, race, ethnicity, nativity, and education. In order to tabulate different demographic characteristics in a comparable manner, we use a classification strategy that can be applied uniformly to each category within each demographic trait that we consider. As a result, our approach can be used for binary worker characteristics (sex, ethnicity, and nativity), categorical characteristics (race and education), as well as continuous characteristics (age).

The key basis for our classification approach is that we treat each category within a demographic characteristic as a distinct individual characteristic, and we separately consider the share of the firm's workforce in each year that belongs to that demographic category. Wherever possible, we use a definition for each demographic category that is consistent with the demographic categories tabulated in the Quarterly Workforce Indicators, so our firm-based statistics can be directly combined with their job-based statistics. For the workforce age composition tables, we split worker age into 7 detailed categories: 14–18, 19–24, 25–34, 35–44, 45–54, 55–64, and over 65, as well as into 3 coarse categories: 14–24, 25–54, and over 55. For the workforce racial composition tables, we separately calculate the share of the firm's workforce that belongs to each of the six racial categories available in our data: White, Black, American Indian or Alaskan Native, Asian, Native Hawaiian or Pacific Islander, and two or more races. Since the sex, ethnicity, and nativity characteristics we use are binary, we only tabulate the workforce share for one of the two categories, and the share for the other category is defined implicitly as its complement.[5] Finally, we define the workforce educational composition by the highest degree attained: less than high school, high school, some college, and bachelor's degree or higher.

We define the firm's workforce share for each demographic category as the share of matched workers with a valid W-2 earnings record and defined demographic characteristics who belong to that category. We calculate this share separately for each year, so the firm's workforce share can change over time either due to worker turnover or due to changes in the time-varying characteristics of incumbent workers.[6] When calculating shares, we weight each worker at the firm equally. If some groups of workers are more likely to be employed on a part-time or partial-year basis, these workers would be overrepresented in the W-2 data relative to a point-in-time employment definition.[7]

[5]We currently tabulate using the race and ethnicity definitions of the 2010 and 2020 Decennial Censuses to be consistent with the QWI.

[6]The ICF uses the most recent data available for each worker demographic characteristic, so the only time-varying characteristic in our case is worker age.

[7]Since we only observe total annual earnings in the W-2 data, we are not able to construct alternatives that weight

Although the LEHD ICF includes imputed demographic variables, we only use the non-imputed measures to calculate each firm-level share. This is equivalent to assuming that workers who are missing demographic measures have the same average composition as other workers at the firm. The appropriateness of this assumption may depend on the demographic variable in question and the type of analysis being conducted, so we also separately tabulate all tables based on the firm's workforce share that is missing each demographic trait. This allows us to directly assess the missing patterns and make alternative assumptions about their assignment process and their implications on overall employment trends.

Figure 4 plots the total share of employment in our sample, broken out by the share of the firm's workforce that is missing each demographic characteristic. The patterns are consistent with the general missing data patterns reported for the LEHD ICF. Information on age, sex, and place of birth is collected in the Numident for every worker with a Social Security Number (SSN). Correspondingly, Figures 4a, 4b, and 4e show that the overwhelming majority of employment is at firms that have observable age, sex, and nativity data for over 90% of their workforce. Data on race and ethnicity are only collected in the Decennial Censuses and the ACS, and would not cover recent immigrants or non-respondents. As a result, Figures 4c and 4d show that a greater share of employment is at employers who are missing race or ethnicity data for more than 10% of their workforce, but there are very few employers who are missing this data for 25% or more of their workforce. On the other hand, given the measurement issues described in Section 2.2, it is unsurprising firms tend to be missing education data for the majority of their workforces, either because those workers were born after 1976 (and therefore excluded from the educational sample) or because they were not surveyed by the 2000 Census long-form or any subsequent ACS. Nevertheless, Figure 4f shows that 11.1% of total tabulated employment over this period is at firms where we observe educational attainment for 25-50% of the firm's workforce (i.e., 50-75 % of education is missing), and another 54.2% of total tabulated employment is at firms where we observe educational attainment for 10-25% of the firm's workforce (i.e., 75-90% of education is missing). In total, at least 66.1% of tabulated employment belongs to firms where at least 10% of the firm's workforce is born before 1976 and has a valid educational response from either the 2000 Census or the ACS.

After defining a firm's workforce share for each demographic category, we then partition these shares into six discrete groups, and we use these groups to tabulate business outcomes for our tables. The six intervals are symmetric and are based on common distributional thresholds: 0–10%, 10–25%, 25–50%, 50–75%, 75–90%, and 90–100%. Consequently, it is straightforward to combine tabulated output to produce statistics based on coarser bins (e.g., by combining 50–75%, 75–90%, and 90–100% into a single 50–100% group). However, note that, like in other BDS products, it is not possible to combine statistics across multiple, non-mutually exclusive firm characteristics into an aggregate category.[8] So, we also tabulate tables based on an explicitly defined set of coarse age categories: 14–24, 25–54, and over 55.

All workforce composition measures are defined at the firm-level and assigned to each of their establishments. We separately tabulate business dynamics based on an establishment's initial workforce composition and its average workforce composition. The two different timing definitions are

individuals by total hours or duration of employment during the year. The impact of these alternative methods is a topic for future research.

[8]For example, if a firm has 0–10% of workers in the 55-64 age category and 0-10% of workers in the over 65 age group, its total share of workers over age 55 can either be 0–10% or 10–25%.

created to match the timing of the firm size measures from the core BDS. The establishment's initial firm workforce composition is defined as the establishment's firm workforce shares from the prior year.[9] Meanwhile, the establishment's average workforce composition is defined as the average of the establishment's firm workforce share in the prior year and in the current year (weighted by total firm employment in each year).[10] For definitions of each workforce composition variable, see the codebook in Appendix Section C. For expositional convenience, whenever we refer to an establishment's workforce characteristics or firm size in this paper, we refer to that establishment's initial workforce characteristics or initial firm size. We explicitly compare the difference from using initial vs average measures of firm characteristics when we discuss the worker exposure to job creation, job destruction, and business dynamics in Section 5.

In the first release, we compile and publish five sets of tables, which are listed in Table A.1. We produce a base set of two-way tables stratified by the initial workforce composition for each demographic characteristic category interacted with the initial firm size. We also produce a corresponding set of tables stratified by the average workforce composition for each demographic characteristic category interacted with average firm size. Finally, we produce three additional sets of three-way tables by interacting the initial workforce composition by initial firm size tables with either firm age, the establishment's industry sector, or the establishment's state.[11]

4 Distribution of employment by workforce composition

We introduce our tables by describing the cross-sectional distribution of employment across firm demographic categories in the most recent year of our data, 2022. As expected, the most common firm workforce composition group for employers tends to correspond to the average prevalence of the demographic category among the working population. However, for every demographic characteristic that we consider, we find more dispersion in the composition of firm workforces than what would be expected if workers and firms matched randomly, even after accounting for the dispersion from small employers.

We then expand our discussion of total employment from 2022 to the entire 2006–2022 period covered by our tables. Over our sample period, we find that employment is increasingly concentrated at firms whose workforces are older, more racially and ethnically heterogeneous, more foreign-born, and more college-educated. The reallocation in employment during the Great Recession was a major driving force in these changes, but many demographic changes continued in the subsequent recovery up to 2020. Using the longitudinal job growth measures available in the BDS-HC tables, we estimate that if the workforce composition of each firm had remained fixed over time, the counterfactual distribution of employment would have remained stable or moved in the opposite direction. This suggests that the trends that we observe are primarily driven by changes in the composition of workers within continuing firms, rather than the reallocation of employment between firms.

[9] If the establishment's firm workforce share for the prior year is missing, then we use the establishment's firm workforce share from the current year.

[10] If the establishment's firm workforce share for the prior or current year is missing, then we only use the establishment's firm workforce share for the non-missing year

[11] All industry codes are on a NAICS 2017 basis.

4.1 Cross-sectional differences in 2022

The dark bars in Figure 5 report the total employment in 2022 at firms whose workforce compositions fall under each segment of the distribution. To compute the data for these figures, we use the tabulated 2022 employment for each group in the workforce composition by firm size tables, and we aggregate employment in each workforce composition group across all firm size categories.[12] In this paper, we report one focal demographic category for each demographic characteristic, but analogous figures can be constructed for any other tabulated demographic categories since all tables follow a consistent structure.

The overall employment by firm workforce composition generally follows an approximately single-peaked distribution. For each of the six demographic characteristics that we consider—age, sex, race, ethnicity, nativity, and education—the majority of total employment is at firms whose workforce composition is at or close to the modal group. The Business Dynamics Statistics reports that in 2022, the total employment of all tabulated firms is 133.7 million. Since our tables are built using the same universe of firms and sources of business data, corresponding measures can be compared directly. Figure 5a shows that in 2022, 65.0 million workers work at firms that have 10–25% workers over age 55, consisting of 48.6% of total tabulated employment. Employment at firms with less than 10% or 25–50% of workers over age 55 is smaller but still significant, with 31.9 million (23.9% of total employment) and 30.3 million workers (22.7% of total employment), respectively. Finally, firms with a majority of workers over age 55 (i.e., greater than 50%) are uncommon and represent only 5.1 million workers (3.8% of total employment). The remaining 1.3 million in employment are at firms that do not have available data to calculate firm workforce characteristics. Age data at the Census is relatively complete due to the availability of the Numident, so the main reason for the lack of data for the age tables is that we cannot find matching W-2 records corresponding to the firm.

What distribution of employment should we expect to see? One possible benchmark is the share of workers who belong to the same demographic category. Since our demographic definitions are based on the definitions used in the QWI, we can, in most cases, calculate the population share of this type of worker using the publicly available QWI tables. In 2021, the reference year for the initial workforce composition measures, 23.5% of nationwide employment is held by workers over age 55. If the workforce composition of every firm reflected this nationwide average, then all employment would correspondingly be at firms where 10–25% of workers are over age 55. Our data show that while this segment is indeed the modal group, only 48.6% of employment falls within this group.

We next show that the degree of observed dispersion in the data cannot be explained by random matching at small firms, which are more likely to have an extreme share of workers in any given category. To do this, we calculate a predicted distribution of employment based on the national share of workers in each demographic category and the distribution of employment by firm size. We would expect small firms to exhibit more dispersion in firm workforce characteristics, even if matching in the labor market were completely random, since the variance of the mean over Bernoulli draws decreases with the sample size. To assess the degree of dispersion that can be explained by small firms, we use the employment distribution from the initial firm size tables in the core BDS, and we calculate the total employment that would be expected to fall under each firm workforce composition group given the existing firm-size distribution in the U.S. To establish a conservative upper bound on the degree of dispersion, we assume that all firms in each detailed firm size category

[12] For aggregation details, see Appendix Section B.1.

are the smallest size possible within that category. Within a firm size category, we predict the probability of a firm falling in each segment of the workforce composition distribution using the binomial distribution.[13] The light bars in Figure 5 report the total expected employment under random matching. We estimate that at most 12.6 million workers would be employed at firms with less than 10% of their workforce over age 55. This is significantly smaller than the 31.9 million workers we actually observe in the same group. Even after accounting for the dispersion due to small firms, firms have substantially fewer older workers than we would expect with random assignment. Similarly, we estimate that under random matching, 20.6 million workers would be employed at firms with 25-50% of their workforce over age 55, whereas we observe 30.3 million workers employed at firms in this group.

We apply this same exercise to the dispersion in the firm workforce female share in Figure 5b, the firm workforce share White in Figure 5c, the firm workforce share Hispanic in Figure 5d, the firm workforce share foreign-born in Figure 5e, and the firm workforce share with a college degree in Figure 5f. We find similar patterns for each characteristic. The modal group in the firm workforce distribution is close to the general prevalence of the demographic category in the working population. This means that the most common type of employer is one whose workforce composition is close to the composition of the general working population. However, in each case, the distribution of employment is substantially more dispersed than what would be expected under random matching, even after accounting for the firm size distribution. The difference between expected employment and actual observed employment for the modal group ranges from 31.5 million for the firm workforce share over age 55 to as high as 68.6 million for the firm workforce share with a college degree. In other words, at least 68.6 million workers are employed in firms where the firm's workforce educational distribution is not representative of the nationwide working population. We clearly reject the model that firms draw workers from a single, common distribution, as measured by any of the worker characteristics we examine. It is important to note that our tables are not designed to directly isolate any specific economic or non-economic explanations for this dispersion, but instead are designed to systematically understand the magnitude and correlations of this dispersion with key factors like geography, sector, and firm age and size.

4.2 Changes in workplace composition, 2006–2022

Figure 6 expands the employment measures from 2022, as reported in the dark bars in Figure 5, to the entire 2006–2022 period covered by our matched demographic data. While firms' sex distributions remained relatively stable over this period—changes to total employment of firms belonging to one segment of the workforce sex composition moves roughly in parallel with changes to firms in the other segments, we find substantial reallocation in employment across different segments of the workforce composition distribution in the other five demographic categories.

Figure 6a shows that the typical workplace has become much older between 2006 and 2022. Total employment at firms with less than 10% of their workers over age 55 started at a high of 45.3 million in 2006 and steadily dropped in subsequent years. Most of this decline occurred over 2006–2011, a period when employment losses around the Great Recession coincided with the aging of many Baby Boomers (individuals born between 1946 and 1964) into the "over age 55" group. Meanwhile, total employment at firms with at least 25% of their workers over age 55 started at a low of 12.8 million

[13] For details on our methodology, see Appendix Section B.2.

in 2006 and grew steadily over time until 2021, when it reached a high of 35.5 million. By 2022, a greater share of workers were employed at firms where at least 25% of workers were over age 55 than at firms where less than 10% of workers were over age 55.

Figures 6c, 6d, and 6e show that the typical firm employs more racially and ethnically heterogeneous workers and more workers born in other countries. Employment at predominantly White firms—firms where over 90% of the workforce is White—was the second largest employer group in 2006, but between 2006 and 2022, the total employment in this group steadily declined from 37.5 million to 29.9 million in 2022. On the other hand, employment at firms that were at least 25% non-White (or, equivalently, firms that are less than 75% White) increased from 34.7 million in 2006 to 57.3 million in 2022, despite experiencing large declines in employment between 2008 and 2010. Today, workers are more likely to work at a firm that is less than 75% White than to work at a firm that is more than 90% White. The trends in employment by firms' workforce share Hispanic or share foreign-born are similar. Employment at predominantly non-Hispanic firms—firms where less than 10% of the workforce is Hispanic—was as high as 69.1 million in 2006, and declined to 57.5 million by 2022. The changes by nativity are qualitatively similar but dampened. Employment at predominantly native-born firms—firms where less than 10% of the workforce is foreign-born—was as high as 59.1 million in 2006, and declined to 53.9 million by 2022. In both cases, workers in 2022 were more likely to work at a firm that is at least 10% Hispanic or at least 10% foreign-born than a firm where over 90% of workers are non-Hispanic or native-born, whereas the opposite was true in 2006.

Finally, Figure 6f highlights the reallocation of employment towards firms where a greater share of workers have college degrees. Note that by restricting our measure of education to workers who were born before 1976, we are abstracting away from any workforce composition changes that would arise from changes in the college attainment rate for younger workers. Even when we focus on workers who were over 30 years old in 2006 (and over 46 years old by 2022), we still find that the employment at firms with less than 10% of college graduates (among workers born before 1976) declined from 28.1 million in 2006 to 24.2 million in 2022. Meanwhile, employment at firms where more than half of the workers have college degrees (again among workers born before 1976) increased from 21.5 million to 29.6 million in 2022. Moreover, employment declines during the Great Recession seem to have been concentrated in employers where less than half of their workforce had a college degree. Total employment in these firm categories dropped sharply after 2008 and remained depressed for years. On the other hand, employment losses at firms where more than half of their workforce had a college degree were negligible in 2008 and were quickly reversed by subsequent gains after 2009.

Since we calculate the composition of firm workforces separately by year, the patterns we observe in Figure 6 may be driven either by within-firm changes to the composition of their workforces, or the reallocation of employment between firms (e.g., from different rates of job creation or job destruction). For example, total employment at predominantly White firms may have fallen either because more non-White workers joined those firms so their firm workforce share White dropped below 90%, or because establishments at those firms were more prone to job destruction rather than job creation. One major advantage of integrating our workforce composition measures into the Business Dynamics Statistics infrastructure is that the establishments are linked over time, so we can directly assess the relative contribution of each of these mechanisms.

Figure 7 reports the counterfactual path of employment for each workforce composition measure

under the scenario that firm workforce compositions are fixed over time. To construct this counterfactual, we first take the total employment in each firm workforce composition group as calculated in 2006. For each year after 2006, we then apply the year-over-year net job creation rate to the previous year's counterfactual employment total.[14] This approach captures changes in employment due to job creation or job destruction from continuing establishments as well as from entry and exit, but excludes any changes in employment due to continuing employers leaving a given firm workforce characteristic group (i.e., due to changes in the composition of their workforce). In each demographic category, the counterfactual path of employment is different from or the opposite of what we observe in the data. If the distribution of worker characteristics at each firm remained constant over time, employment would have shifted towards firms with fewer workers over age 55 and remained largely stable across the firm workforce composition distribution in the other categories. Although firms with predominantly White, non-Hispanic, U.S.-born, or non-college-educated workforces would still have experienced sharp employment declines around the Great Recession, their employment totals would have recovered and increased further in the subsequent decade. Instead, we observe in Figure 6 that employment at these firms continued to decline, suggesting that the long-run changes over this period are primarily driven by systematic changes to workforces within firms.

5 Worker exposure to business dynamics

Although the analysis in Section 4.2 shows that the changes in the demographic composition of employers over 2006-2022 have been primarily driven by within-firm changes in their workforce composition, there may nevertheless be significant differences in the business dynamics of firms with different workforce compositions. In this section, we use longitudinal outcomes from the LBD to directly assess the degree to which different types of workers are exposed to different degrees of job creation, job destruction, entry, and exit by their employers. These differences should be interpreted with caution since the degree of dispersion in a firm's workforce demographics is mechanically related to its size. However, we still find large differences in the economic environments faced by different groups of workers even after we account for firm size. Some of these differences—particularly by the firm workforce share female—can be explained by differences in the employer's sector, but the other differences remain or become larger.

Small firms tend to exhibit more variation in firm workforce characteristics. Some of this greater dispersion may be due to fundamental differences in the organization and hiring practices of these employers. However, if there is at least some idiosyncratic randomness in the matching between workers and firms, then we would expect the dispersion in workforce composition at smaller firms to be mechanically greater due to their smaller number of employees.[15] Past research has emphasized the importance of firm size as a source of differences in firm dynamics, and simply aggregating the job creation or destruction rate of firms across workforce composition groups is likely to reflect differences related to size rather than workforce composition (Haltiwanger et al., 2013).

To ensure that we isolate potential differences in job creation or destruction due to firm workforce characteristics rather than simply size, in Figure 8 we report average employment dynamics rates

[14] For details, see Appendix Section B.3.

[15] As an extreme example, consider a firm with only a single employee. The workforce composition for any demographic characteristic at this firm can only alternate between 0% and 100%, depending on the characteristics of the employee.

by firm workforce composition while holding the firm size distribution constant. For each workforce characteristic group, we take the average of its firm-size-specific rates while weighting by the average, economy-wide firm-size distribution of denom.[16] This ensures that any differences between two firm workforce characteristic groups are driven by differences in the business dynamics of firms within the same firm size category, rather than differences in the distribution of employment across firm size categories. Finally, to reduce the contribution of small categories where the job dynamics rates are noisily estimated, we only include firm size categories with at least 20 employees (where the workforce composition and business dynamics rates are both likely to be more stable).

We highlight several facts from our comparisons. First, the dispersion in firm workforce compositions is related to real differences in the economic environments faced by different groups of workers. Even after we control for differences in the firm size distribution across groups, we still find that the average net job creation rate for firms in different workforce composition groups ranges from -1.9% per year, for employers with 25–50% of workers over age 55, to 2.7% per year, for employers with over 90% of workers holding a college degree. We find that the net job creation rates tend to be higher for employers with workforces that are younger, female, non-White, Hispanic, foreign-born, and college-educated. At the same time, we find that net job creation rates are also lower for employers that are over 90% female, 75–90% non-White, over 90% Hispanic, or 50–75% foreign-born. Although the total employment for firms that belong to these outlier categories tends to be relatively small, any general conclusions about the relationship between net job creation rates and firm workforce characteristics should be made with caution.

The entry and exit of establishments can explain some differences in net job creation rates between different employers, but often, the degree of churn from business entry and exit varies independently from differences in net job creation rates. For example, 57% of the difference in net job creation between firms with 25–50% versus less than 10% of workers over age 55 can be explained by lower rates of job creation from births at firms with more workers over age 55. On the other hand, employers with workforces that are less White or more Hispanic experience both more job creation from births, job destruction from exits, and overall higher net job creation rates, suggesting that both growth at existing firms and the contribution of entry and exit of employers are important for understanding the contribution of firms to economic outcomes for these workers.

Research on business dynamics has also highlighted the importance of accounting for mean reversion when comparing differences in net job creation by firm size (Davis et al., 1996; Kim et al., 2024). Although we do not explicitly consider differences by firm size, we still use size-contingent measures of job growth to calculate all average differences in this section. To assess the degree to which our conclusions depend on different measures of firm size, we repeat our analysis using the average workforce composition by average firm size tables (rather than initial workforce composition by initial firm size) in Figure A.1. Consistent with past work, we find that using initial firm size can understate the overall rates of job creation for medium and large firms relative to using average firm size, but the qualitative patterns by workforce composition and the overall degree of dispersion in job creation rates remain similar.

Finally, we assess the degree to which different dynamics for firms with different workforce compositions can be explained by geographical or sectoral differences. It is important to emphasize

[16] Denom is the average employment of each establishment between the current and prior year and is used as the denominator for all employment dynamics rates in the BDS and BDS-HC tables. For details on the construction of these weighted means, see Appendix Section B.4.

that our results describe the degree of segmentation in labor markets and the consequences of that segmentation for different types of workers; they should not be interpreted as the causal effect of workforce composition on net job creation or establishment entry and exit. Furthermore, the distribution of workers across different geographical areas or sectors is itself an equilibrium outcome that is determined by both economic and non-economic factors. So, this decomposition exercise primarily serves to understand the degree to which these firm-level differences may follow existing geographic or sectoral boundaries, and whether they may be captured in existing worker-level data.

Figure 9 reports average business dynamics rates by firm workforce composition while holding both state and firm size constant across all workforce composition groups. Specifically, for each group, we take the average of its state-and-firm-size-specific rates while weighting by the average, economy-wide state-by-firm-size distribution of denom. This then ensures that any differences between segments in the firm workforce characteristic distribution are driven only by differences in the net job creation rates of firms within the same state and firm size category. Figure 10 reports the average net job creation rate by firm workforce composition while holding both the firm size and the sectoral composition constant across all demographic characteristics. The construction of these figures is analogous to the construction of Figure 9, but uses the sector-and-firm-size specific growth rates and denom weights.

We find that most differences in the business dynamics of employers remain substantial after controlling for either state or sector, so the differences that we observe between employers would not be captured by more aggregate state- or sector-level worker data.[17] When comparing Figure 8 to Figure 9, employment dynamics rates for most groups with a substantial number of workers remain similar, and differences largely arise from a greater dispersion in noisily estimated employment dynamic rates in groups with relatively few firms and workers. This is to be expected given that our decomposition uses employment growth rates from smaller, state-specific cells that may exhibit more idiosyncratic fluctuations in growth rates.

When comparing Figure 8 to Figure 10, the difference in net job creation rates by the employers' workforce age, nativity, and educational composition increases slightly, so the sectoral sorting of older, foreign-born, and college-educated workers slightly offsets the within-sector differences in the net job creation rates of employers. Meanwhile, the differences in net job creation rates by the workforce racial and ethnic composition shrank slightly. The only characteristic where controlling for sector makes a substantial difference is the firm workforce sex composition: the positive relationship between firm workforce share female and net job creation rates reverses after we control for the sector of the employer. This suggests that the higher rates of net job creation over our sample period for firms with a higher share of female workers may have been driven by the relatively strong growth of sectors like Educational Services (61) and Heath Care and Social Assistance (62), the two sectors with the highest share of majority-female employers.

[17] This also consistent with the findings from our preliminary analysis using a subsample of single-unit firms, as reported in Beem et al. (2022). When we regress the average net job creation rate or the establishment exit rate on (lagged) measures of the firm's workforce composition, the differences often become larger after we include detailed fixed effects for the establishment's 6-digit NAICS industry and state/MSA.

6 The workforce composition and dynamics of new firms

Having considered general trends in growth and workforce composition among all firms, in this section, we detail the workforce composition of start-up firms. Start-up firms in the BDS are defined as firms that are less than one year old in March of any given year, having hired their first employees after March in the prior year. These firms account for a large share of job creation and are often seen as key to business dynamism (Haltiwanger et al., 2013; Decker et al., 2014; Haltiwanger et al., 2016; Decker et al., 2016). To place our findings in context, we document similarities and differences of start-ups' workforces relative to older firms, defined as those who have employed workers for at least 11 years but were born after 1976. Because over 98% of start-ups are in the 1-19 employees size category, we restrict our analysis to both start-ups and older firms in this size range. After discussing the workforce composition of these firms at their genesis, we then turn to the growth patterns of these small, young firms in the first one to five years after their formation.

6.1 Workforce composition of start-ups compared to older firms

We begin by examining the age composition of the workforce at small start-up firms compared to small older firms. In Figure 11a, we see that 67% of employment at start-ups is at firms with less than 25% of workers in the 19–24 age range. In contrast, 82% of employment at older firms is at firms with less than 25% of these young workers. Correspondingly, employment at start-up firms is more likely to skew towards young workers. 10% of employment is at firms with at least 50% of age 19–24 workers. Only 3% of older firm employment is found in firms that have similarly young workers. While 6% of start-up employment is at firms with an unknown share of young workers, this percentage of missing data is not enough to offset the differences between start-up and older firms. Even if all this employment fell in the less than 25% young workers category, there would still be a gap of eight percentage points between start-ups and firms at least 11 years old.

When we look at the distribution of employment across firms divided by the share of workers age 25–34 and age 35–44 in Figures 11b and 11c, we see very similar patterns. Relative to older firms, employment at start-ups is more likely to be at firms with higher shares of workers in these early-career age ranges. However, beginning with the age group 45–54, the trend shifts and employment at start-up firms is equally likely compared to older firms to be at employers with less than 25% of their workers in this age range. For workers age 55–64, 80% of start-up firm employment is at firms with less than 25% of such workers, as opposed to only 69% of employment at older firms, as shown in Figure 11e. Even the missing rate of six percentage points is not enough to potentially erase the full gap between the start-up and older firms. In contrast, employment at older firms is more highly represented in the 25-50% share category relative to start-up firms, with 21% of older firm employment falling there relative to only 8% of start-up firm employment.

In summary, employment at older firms skews towards older workers, between the ages of 45–64, while employment at start-ups is more likely to have high shares of workers below age 45. This finding is consistent with Ouimet and Zarutskie (2014), which finds a high propensity for new firms to hire relatively young workers.

While the distribution of workers by age seems different between start-ups and older firms, the distribution by sex at first appears more similar. As shown in Figure 12a, the share of start-up employment found at firms that are less than 25% or more than 75% female is slightly higher than

the share of older firm employment in these categories, but the differences are small, on the order of one to two percentage points. However, when we consider the 6% of employment at start-up firms whose sex distribution is unknown, the potential for differences grows. If the start-ups with missing demographic information have the same distribution of workers by sex as the start-up firms with non-missing data, then we can normalize the employment shares for the known categories and predict that 31% of start-ups would have less than 25% female workers.[18] This share is 3 percentage points higher than at older firms. For the largest female share category, employment at start-up firms is 4 percentage points higher than at older firms. We conclude that employment at start-ups is more likely to be at firms that are at the ends of the female share distribution. A higher share of start-up employment is at either heavily male or heavily female firms, compared to old firm employment which is just as likely to be at firms with 50-75% female workers as at firms that have more than 75% female workers.

We next consider the share of foreign-born workers and the share of Hispanic workers at start-up firms compared to older firms in Figures 12b and 12c. A higher share of start-up employment is found in firms with 75% or greater share of foreign-born workers relative to older firms (13% versus 7%). Likewise, the share of start-up employment at firms with less than 25% foreign-born workers is only 65% versus 75% in older firms. Even if all the employment in the missing foreign-born share category was in fact in the less than 25% category, a gap would remain between start-ups and older firms. Start-up employment is over-represented in firms with high shares of foreign-born workers. Similar patterns hold for Hispanic workers where a higher share of start-up employment is at firms with 75% or more Hispanic workers. Finally, with regards to race, in Figure 12d we observe 79% of small older firm employment at firms that are 75% or more White whereas 62% of start-up employment is found at similar firms. Even with 10% of employment being at unassigned firms, this is not enough to close the gap between start-ups and older firms. Conversely, as shown in Figures 12e and 12f, start-up employment is more likely to be at firms with 25% or more Black workers and 75% or more Asian workers.

In summary, on an employment-weighted basis, small start-up firms have higher shares of young workers and non-White or Hispanic workers relative to small older firms, although the differences based on age are much larger than the differences by race. Employment is distributed unevenly across all firms in terms of sex composition, skewing towards firms with either relatively small or large shares of female workers. However, this is more true for start-ups than for older firms. Finally, employment at start-ups is disproportionately found at firms with a high share of foreign-born and Hispanic workers and lower shares of White workers.

6.2 Growth rates of young firms

We next consider what happens to start-ups in terms of job growth during the early years of their existence. We do this by examining net job creation rates (NJCR) for young firms age 1-5 who are initially small, meaning they employed between 1 and 19 people in March of the previous year. We

[18]The first category (less than 25% female) accounts for 29.3% of all start-up employment or 31.2% of the employment at start-up firms with known shares of female workers. If employment at start-up firms with missing data is distributed across the four share categories in the same proportions as employment at firms with known data, we would expect the actual (but unknown) share of employment to also be 31.2% for the smallest share category. For this to be true, worker demographic data must be missing at random, conditional on firm size and age. This means firms that cannot be linked to their workers are not more or less likely to have high or low shares of female workers, and we can essentially ignore these observations when we calculate the share of employment in each group.

then compare the NJCR of young firms that have a high share of a certain category of worker to those that employ workers spread more broadly across categories.

As shown in Figure 13, we find that higher net job creation rates are often associated with establishments belonging to firms that have 50% or fewer workers in a particular age category. In contrast, those firms with a 75% or greater share of workers in a single age category have lower net job creation rates. For the 19-24, 25-34, and 35-44 age categories, firms with between 25-50% of workers in these age ranges have the highest growth rates. For older age categories (45-54 and 55-64), firms with less than 25% of workers in these age categories have the highest growth rates. Firms with greater than 75% of any age category of workers consistently have negative net job creation rates. On average, growing firms have a workforce that spans the age distribution.

In Figure 14 we show results for other demographic groups. Firms with 25-50% female workforces generally saw strong positive net job creation rates, especially in 2022 coming out of the pandemic. The one exception was the Great Recession, when net job creation rates fell dramatically for all firms, regardless of their share of female workers. Firms that are less than 25% female did the worst in the Great Recession. As was the case in Section 5, this result reflects differences in industry composition for female versus male workers. In 2009, the industry sectors with the lowest net job creation rates were Construction (23), Administrative Support and Waste Management (56), Manufacturing (31-33), and Agriculture Support-Forestry-Fishing-Hunting (11). Employment in these sectors in 2009 was concentrated in firms with mostly male workers, especially when we focus on small firms. Firms that are less than 25% female (or more than 75% male) constituted an average of 55.6% of overall sectoral employment at small firms in these industries. These differences in sorting across sectors by sex translated into differences in the exposure of male and female workers to cyclical economic activity.

Young firms with 25-50% foreign-born workers also had high net job creation rates, whereas those with over 75% of foreign-born workers usually had negative net job creation rates between 2006 and 2022. The same pattern holds for Hispanic, Black, and Asian workers. Firms with at least a quarter but less than half of their workers in a race or ethnicity category saw high growth in employment, whereas those that were predominantly minority had lower growth rates. This also applies to the share of White workers. Firms with 25-75% White workers had higher NJCRs than firms with more than 75% White workers.

We again emphasize that the BDS-Human Capital tables are not meant to imply causal relationships. The demographic characteristics of a firm's workforce do not necessarily cause a firm to either grow or struggle. Rather, these tables are intended to provide a picture of the business dynamics different types of workers have been exposed to between 2006 and 2022. It is also important to note that the BDS-HC firm age tables do not take account of industry sector and hence cannot account for differences in establishment outcomes that stem from differences in types of business activity. Despite these caveats, we learn important facts about start-ups and young firms from these statistics. Start-ups tend to employ younger, non-White, Hispanic, or foreign-born workers, but they also tend to be more concentrated by sex. Establishments belonging to young firms that tended to employ high shares of similar workers created fewer jobs on net. More research is needed to determine if other types of firm heterogeneity are also associated with these different growth rates.

7 Conclusion

The Business Dynamics Statistics of Human Capital experimental product adds novel statistics describing the sorting between workers and firms and its consequences. We show that firms vary substantially more in workforce composition than would be predicted by random matching of workers and firms. We also show that the distribution of employment by firm workforce composition has changed over time, driven primarily by changes in the composition of workers at continuing firms. These differences in composition are also correlated with differences in business outcomes. Business dynamics, including establishment-level employment growth and decline and entry and exit, vary substantially by firms' workforce composition. This implies that different types of workers are exposed to different economic environments, even within the same broad geographic area (state) or the same type of business activity (industry sector). Finally, we show that small, startup firms skew toward young workforces and that in the early years of existence, establishments with workers spread across the age distribution have higher rates of net job creation.

The BDS-HC tables are the result of ongoing work at the U.S. Census Bureau to measure the interactions between workers and firms in the labor market. For future versions of the product, we plan to expand the BDS-HC tables to provide additional statistics based on the interaction of firm characteristics or worker characteristics, like firm age by sector or worker age by sex. We also plan to expand the time period covered by our tables to include both updated data after 2022 and new sources of data on worker-firm linkages before 2005, giving us a more comprehensive view of the evolution of firm workforces over time. Finally, we hope to add other firm and worker characteristics, including data on educational attainment for younger workers and the demographic characteristics of business owners. We are also working to use the findings from our work to continue improving other BDS products and their underlying microdata. For example, we plan to incorporate workforce composition data into our BDS product that focuses on single-unit firms (BDS-SU-REV) to provide additional detail on firm payroll, revenue, and quarterly dynamics by workforce composition. In addition, we are exploring methods to use our worker-firm data to provide new avenues for validating and refining the outcomes and linkages in the LBD. Ultimately, these future developments all build towards a better understanding of the employment relationship and the impact of these connections on both firms and workers.

References

Altonji, Joseph and Rebecca Blank (1999) "Race and gender in the labor market," Handbook of Labor Economics, Elsevier.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa (2008) "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, 116 (6), 1150–1196.

Beem, Richard, Christopher Goetz, Martha Stinson, and Sean Wang (2022) "Business Dynamics Statistics for Single-Unit Firms," Working Papers 22-57, Center for Economic Studies, U.S. Census Bureau.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018) "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36 (S1), S13–S70.

Chow, Melissa, Teresa C. Fort, Christopher Goetz, Nathan Goldschlag, James Lawrence, Elisabeth Ruth Perlman, Martha Stinson, and T. Kirk White (2021) "Redesigning the Longitudinal Business Database," Working Papers 21-08, Center for Economic Studies, U.S. Census Bureau.

Davis, Steve, John Haltiwanger, and Scott Schuh (1996) *Job Creation and Destruction*: MIT Press.

Decker, Ryan, John Haltiwanger, Ron S Jarmin, and Javier Miranda (2014) "The role of entrepreneurship in US 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 US," *European Economic Review*, 86, 4–23.

Dostie, Benoit, Jiang Li, David Card, and Daniel Parent (2023) "Employer policies and the immigrant–native earnings gap," *Journal of Econometrics*, 233 (2), 544–567.

Ferguson, John-Paul and Rembrand Koning (2018) "Firm Turnover and the Return of Racial Establishment Segregation," *American Sociological Review*, 83 (3), 445–474.

Foster, Lucia, John Haltiwanger, and Chad Syverson (2008) "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98 (1), 394–425.

Graham, Matthew, Erika McEntarfer, Kevin McKinney, Stephen Tibbets, and Lee Tucker (2022) "LEHD Snapshot Documentation, Release S2021_R2022Q4," Working Papers 22-51, Center for Economic Studies, U.S. Census Bureau.

Haltiwanger, John, Ron S Jarmin, Robert Kulick, and Javier Miranda (2016) "High growth young firms: contribution to job, output, and productivity growth," in *Measuring entrepreneurial businesses: current knowledge and challenges*, 11–62: University of Chicago Press.

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.

Hellerstein, Judith K. and David Neumark (2008) "Workplace Segregation in the United States: Race, Ethnicity, and Skill," *The Review of Economics and Statistics*, 90 (3), 459–477.

Internal Revenue Service (2025) *Instructions for Form W-2 and Form W-3*, Washington, D.C.

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin (2024) "Demographic Origins of the Start-up Deficit," *American Economic Review*, 114 (7), 1986–2023.

Kim, J. Daniel, Joonkyu Choi, Nathan Goldschlag, and John Haltiwanger (2024) "High-Growth Firms in the United States: Key Trends and New Data Opportunities," Working Papers 24-11, Center for Economic Studies, U.S. Census Bureau.

Kline, Patrick M. (2024) "Firm Wage Effects," NBER Working Papers 33084, National Bureau of Economic Research, Inc.

Ouimet, Paige and Rebecca Zarutskie (2014) "Who works for startups? The relation between firm age, employee age, and growth," *Journal of Financial Economics*, 112 (3), 386–407.

Syverson, Chad (2011) "What Determines Productivity?" *Journal of Economic Literature*, 49 (2), 326–365.

Topel, Robert H. and Michael P. Ward (1992) "Job Mobility and the Careers of Young Men," *The Quarterly Journal of Economics*, 107 (2), 439–479.

Wagner, Deborah and Mary Lane (2014) "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software," CARRA Working Paper 2014-01, Center for Economic Studies.

Wang, Sean (2021) "What Is the Price for Opportunity? The Effects of Employer Learning on Worker Promotions and Turnover," Working Paper.

Tables & Figures

Table 1: LBD and W2 match rates by link type

Match type	Share matched					
	LBD payroll	LBD emp.	LBD estabs	W2 payroll	W2 emp.	W2 EINs
BDS-HC approach:						
FID links	0.986	0.983	0.968	0.988	0.987	0.951
Alternative approaches:						
EIN links only	0.946	0.948	0.950	0.988	0.987	0.951
EIN links (5-year window)	0.961	0.960	0.956	0.995	0.995	0.962

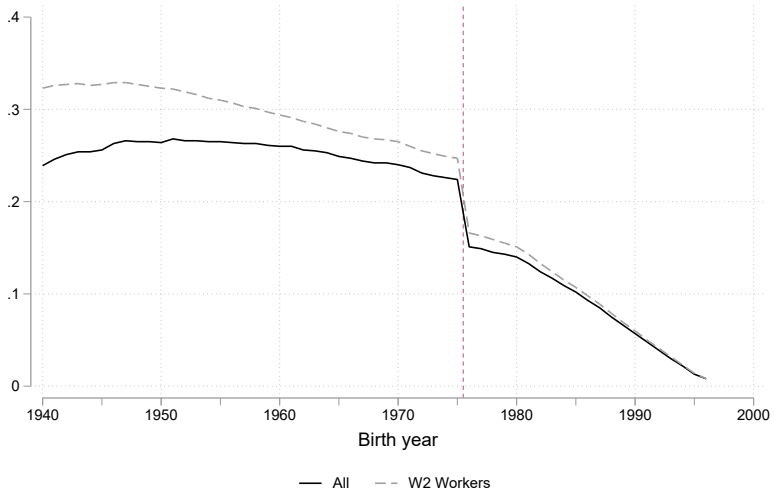
Note: This table reports the average share of the Longitudinal Business Database and W-2 administrative records that can be matched by various linking strategies between 2005 and 2021. The match rates are calculated by year and then averaged across years as an unweighted mean. FID linking is the approach used to create the BDS-HC tables—we combine all EINs associated with a firm in the LBD, and we match an establishment to W-2 records whenever one of the EINs for its parent firm matches the EIN on the W-2 records. EIN linking is a more restrictive approach than our baseline. We only link an establishment to the W-2 records if the EIN associated with the establishment (in the LBD data) matches the EIN on the W-2 records. Finally, EIN linking with a rolling 5-year window expands EIN links to include any EINs associated with the establishment in the 5 years before or after the current year. Data source: LBD and W-2 microdata. DRB approval number: CBDRB-FY25-CES020-002.

Table 2: Matched firm totals, LBD vs W2

	Log LBD Payroll		Log LBD Emp.	
	(1)	(2)	(3)	(4)
Log W2 Payroll	0.983	0.913		
	(0.0000553)	(0.000198)		
Log W2 Emp.			0.860	0.675
			(0.000115)	(0.000219)
Obs	84,870,000	84,870,000	84,870,000	84,870,000
Firms	12,820,000	12,820,000	12,820,000	12,820,000
Adj. R2	0.982	0.988	0.885	0.937
Firm & Year FEs		Yes		Yes

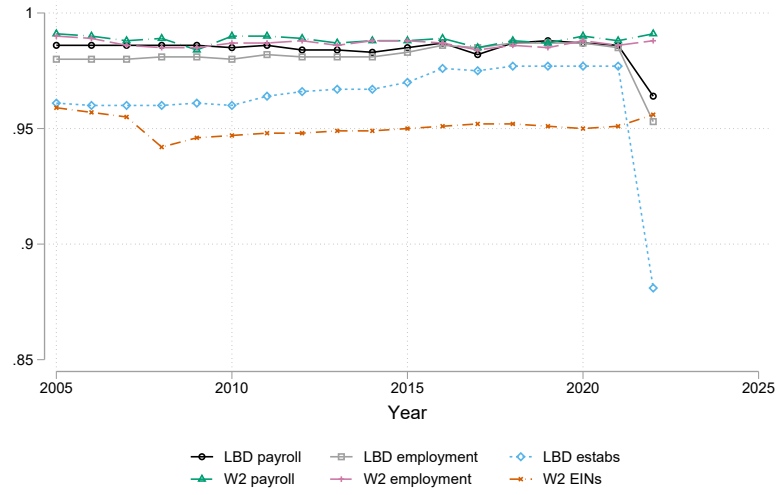
Note: This table reports regression estimates of the relationship between LBD and W2 total employment and payroll for all matched firms in the BDS-HC tables. An observation in the regression is a firm-year. Standard errors in parentheses are clustered by firm. Columns 1 and 3 report bivariate regression estimates, and columns 2 and 4 include both firm fixed effects and year fixed effects. Data source: LBD and W-2 microdata. DRB approval number: CBDRB-FY25-CES020-002.

Figure 1: Education response rates by birth year



Note: This figure plots the share of individuals in the ICF observed demographics file who have an educational response collected after age 25, broken out by the individual's birth year. The black line is based on all individuals that are in the Census Numident, the 2000 or 2010 Censuses, or an American Community Survey. The dashed line restricts the sample to individuals who have received a W-2 earnings record at any point between 2005 and 2022. Data source: LEHD and W2 microdata. DRB approval number: CBDRB-FY25-CES020-002.

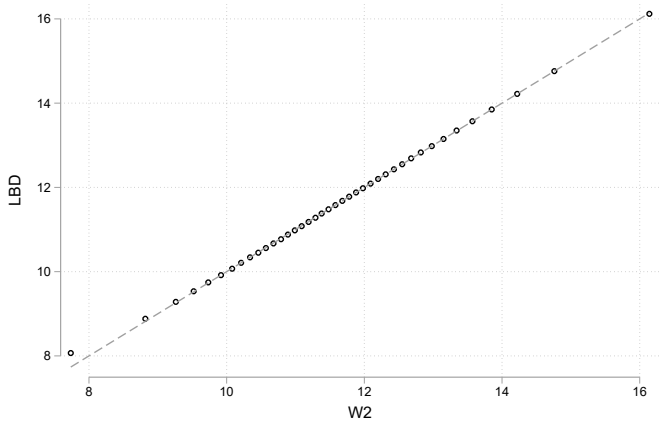
Figure 2: LBD and W2 match rates by year



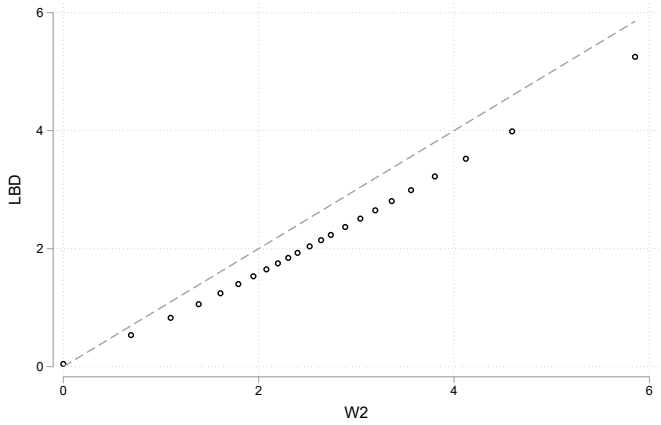
Note: This figure plots the share of the Longitudinal Business Database and W-2 administrative records that can be matched for the BDS-HC tables in each year between 2005 and 2022. The average match rates from 2005 to 2021 correspond to the match rates reported in the FID links row of Table 1. Data source: LBD and W-2 microdata. DRB approval number: CBDRB-FY25-CES020-002.

Figure 3: Matched firm totals, LBD vs W2

(a) Total payroll

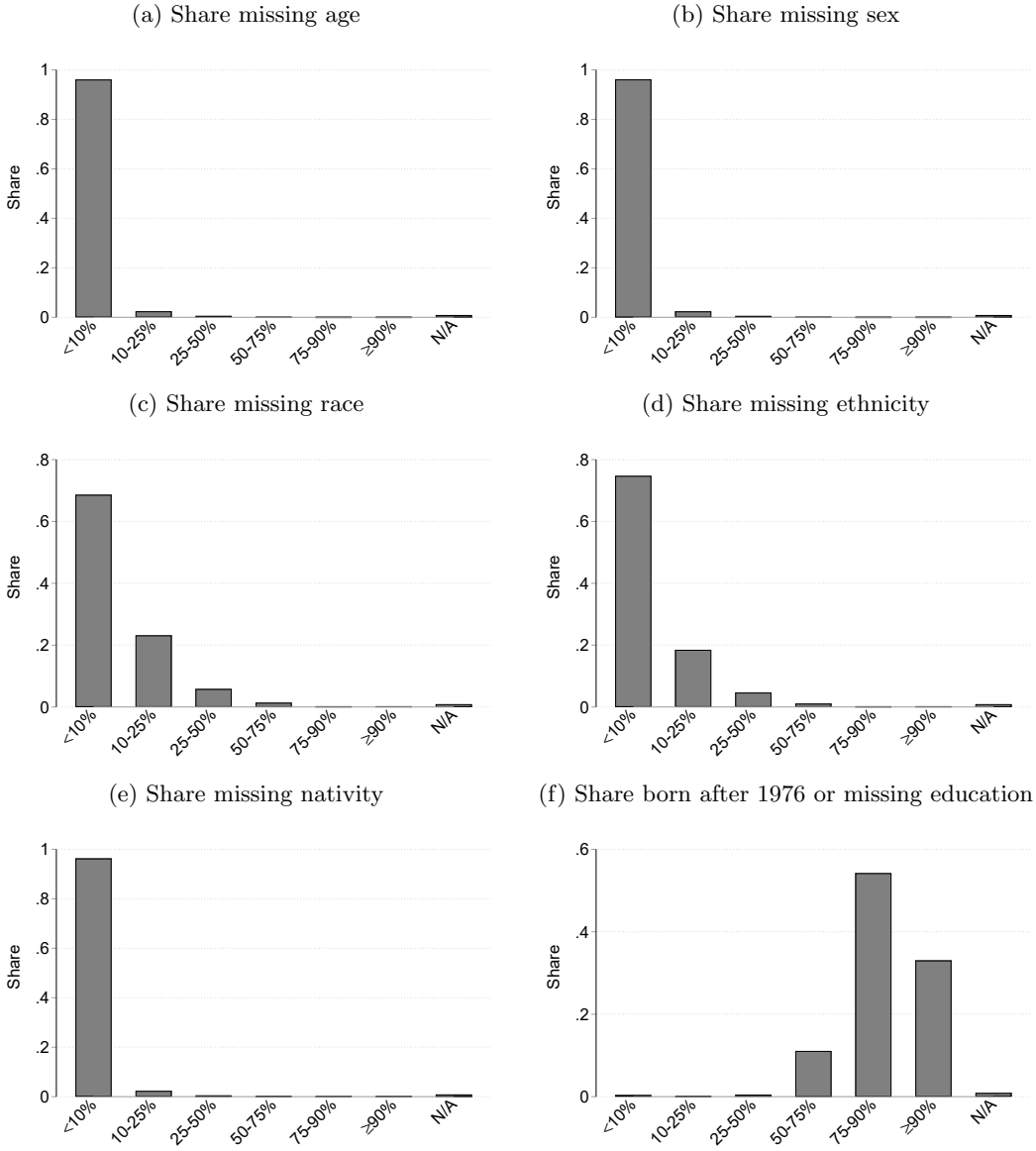


(b) Total employment



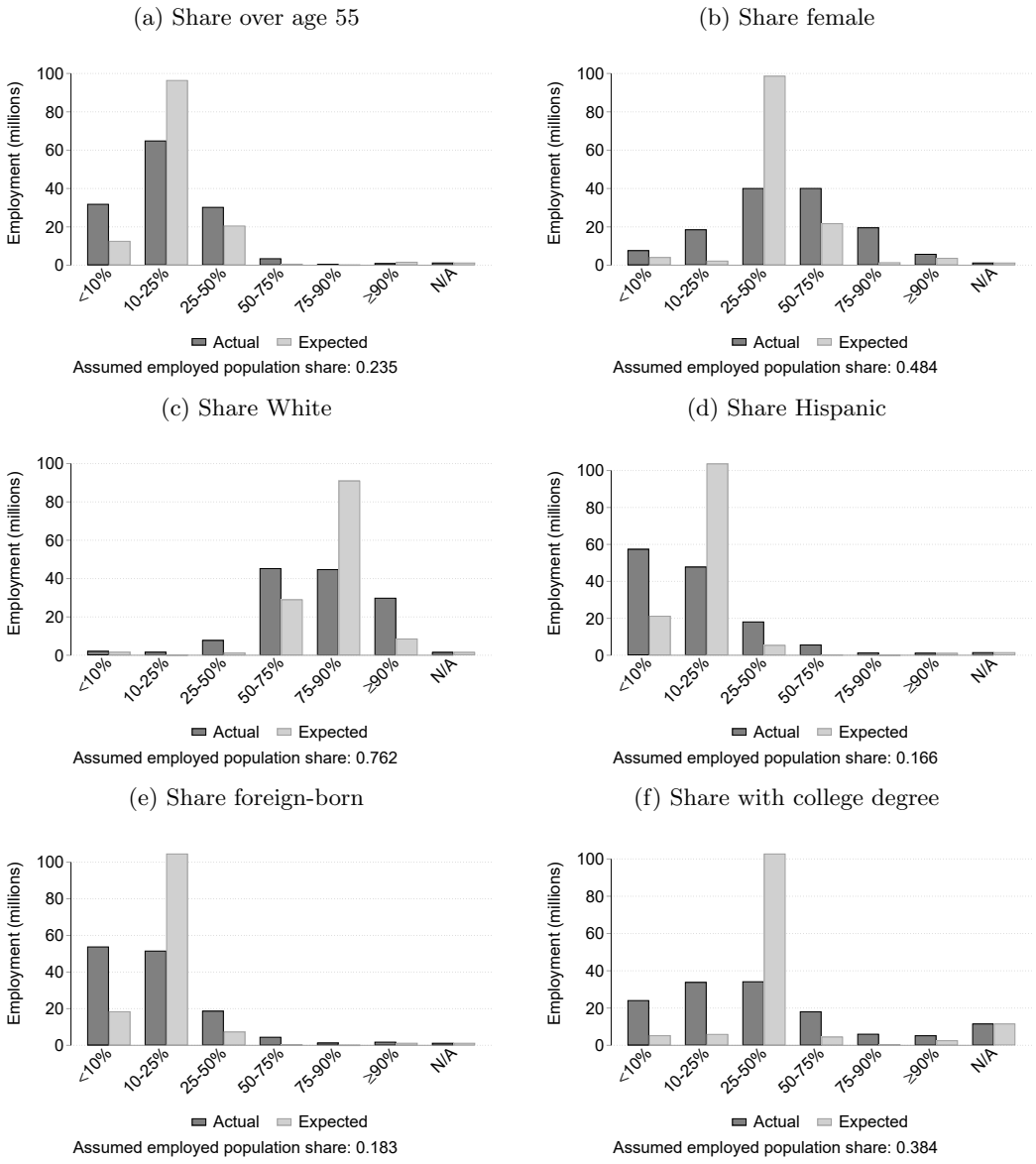
Note: This figure plots the relationship between LBD and W2 total employment and payroll for all matched firms in the BDS-HC tables between 2005 and 2021. We divide all firm-year observations into 40 equally sized groups based on their total W2 payroll (in Panel A) and the total W2 employment (in Panel B). For each group, we calculate the average log payroll and log employment across all firm-years in the group separately by data source. The dashed line is a 45-degree reference line. Data source: LBD and W-2 microdata. DRB approval number: CBDRB-FY25-CES020-002.

Figure 4: Distribution of 2006–2022 employment by firm workforce data coverage



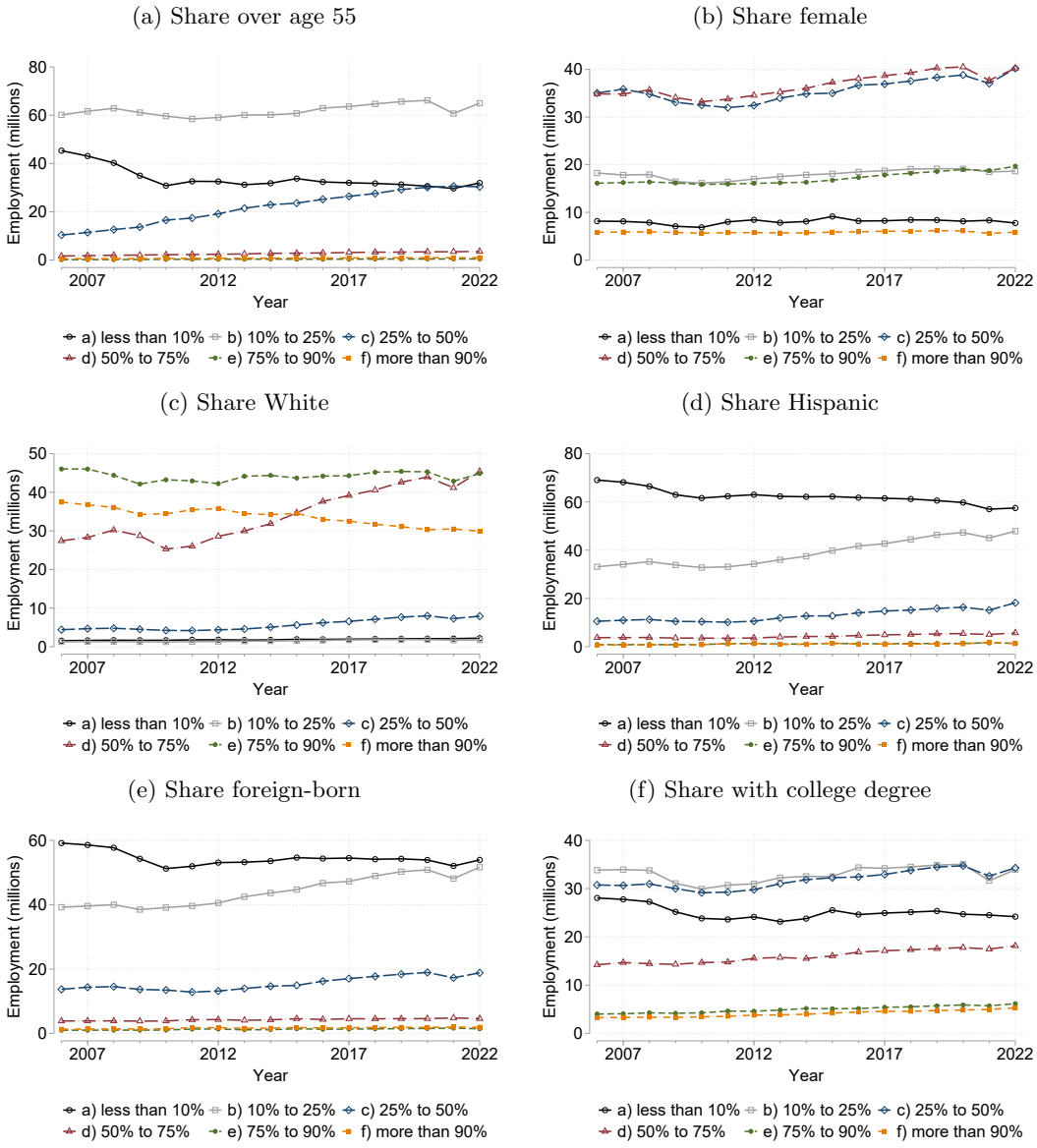
Note: This figure plots the average share of total tabulated employment between 2006 and 2022 based on the share of each firm's workforce that is missing the referenced source of demographic data. Firms that cannot be matched to any workers are separately tabulated in the "N/A" category. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 5: Distribution of total 2022 employment by firm workforce characteristic



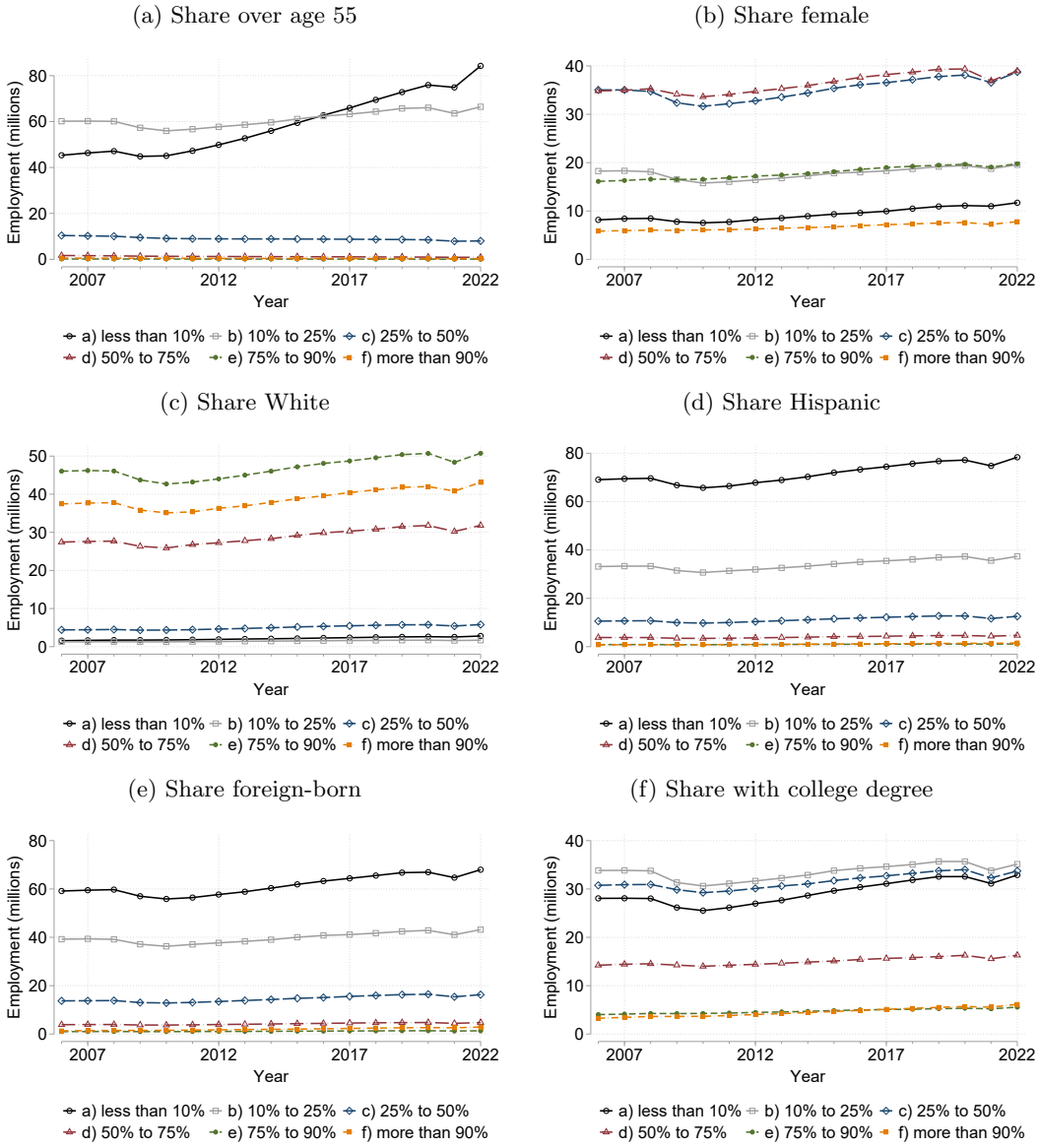
Note: This figure plots the total tabulated employment in 2022 based on the share of each firm's workforce that belongs to the referenced demographic group. Firms that cannot be matched to any workers are separately tabulated in the "N/A" category. The dark bars report the observed employment totals, and the light bars report the expected employment in each category if workers and firms matched randomly and the firm size distribution matches that of the observed U.S. distribution. The assumed population share for the expected employment totals is calculated from the reference year in the QWI (in panels A-D) and the ACS (in panels E and F). Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 6: Annual employment by firm workforce characteristic



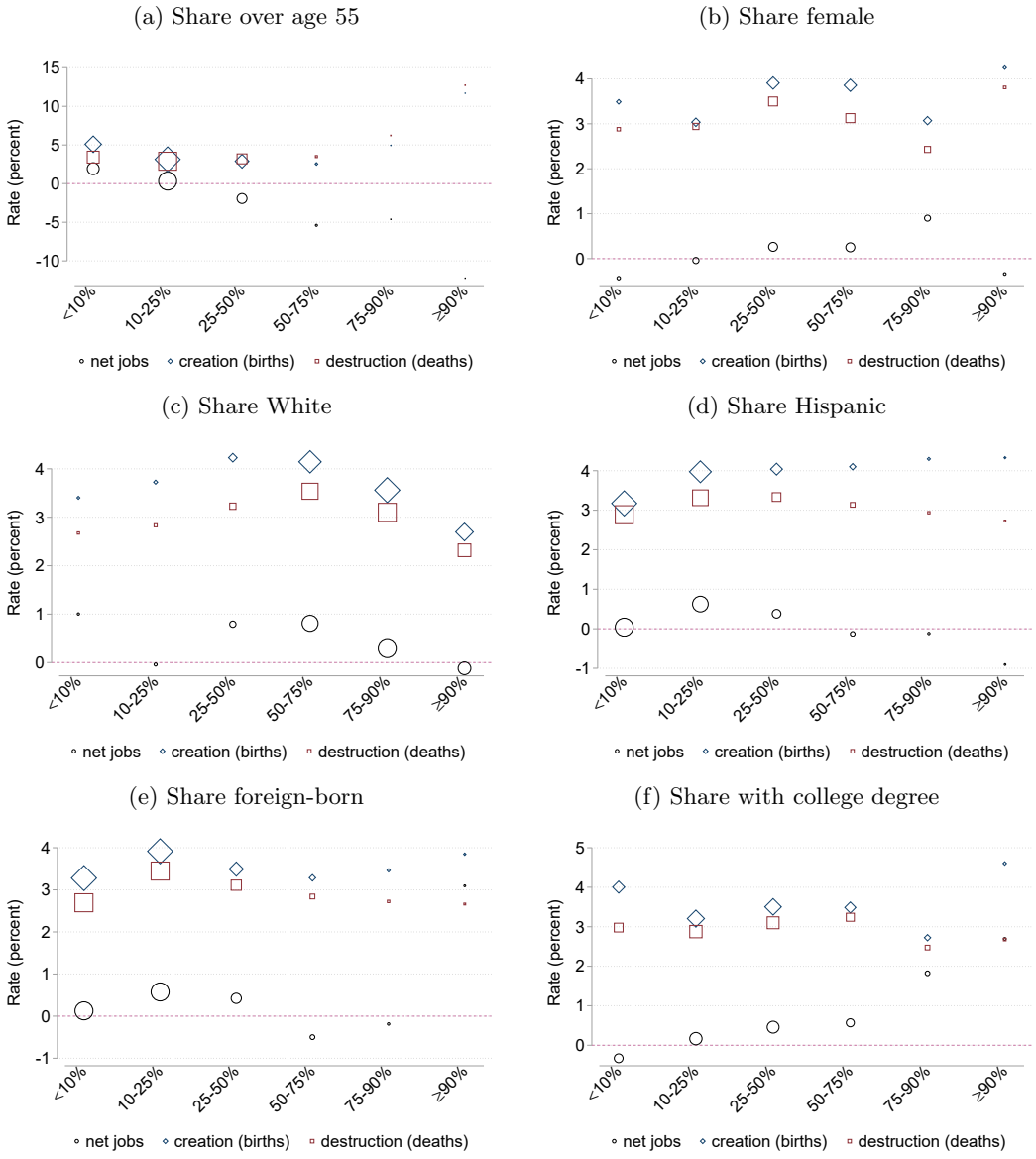
Note: This figure plots the total tabulated employment in each year based on the share of each firm's workforce that belongs to the referenced demographic group. The totals in 2022 also match the observed totals (dark bars) in Figure 5. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 7: Counterfactual annual employment (no changes in firm workforce composition)



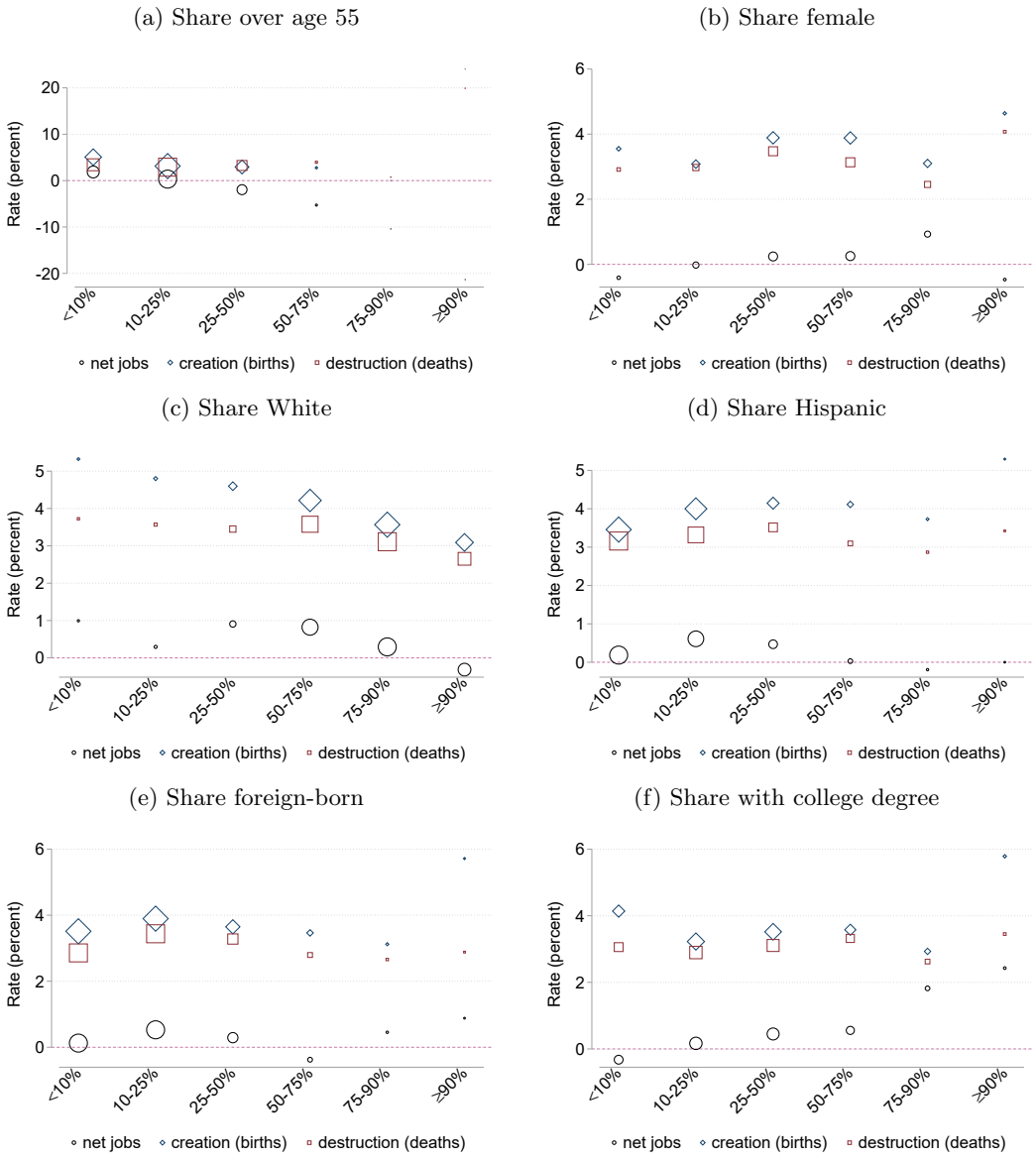
Note: This figure plots the total counterfactual employment in each year after 2006 based on the share of each firm's workforce that belongs to the referenced demographic group, under the assumption that firms' workforce compositions are fixed and do not change over time. Any changes in employment in this figure are only due to job creation, job destruction, or the entry and exit of employers. Any changes due to existing employers changing the composition of their workforces are excluded. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 8: Employment change at medium and large firms by initial firm workforce composition, 2006–2022 (controlling for firm size)



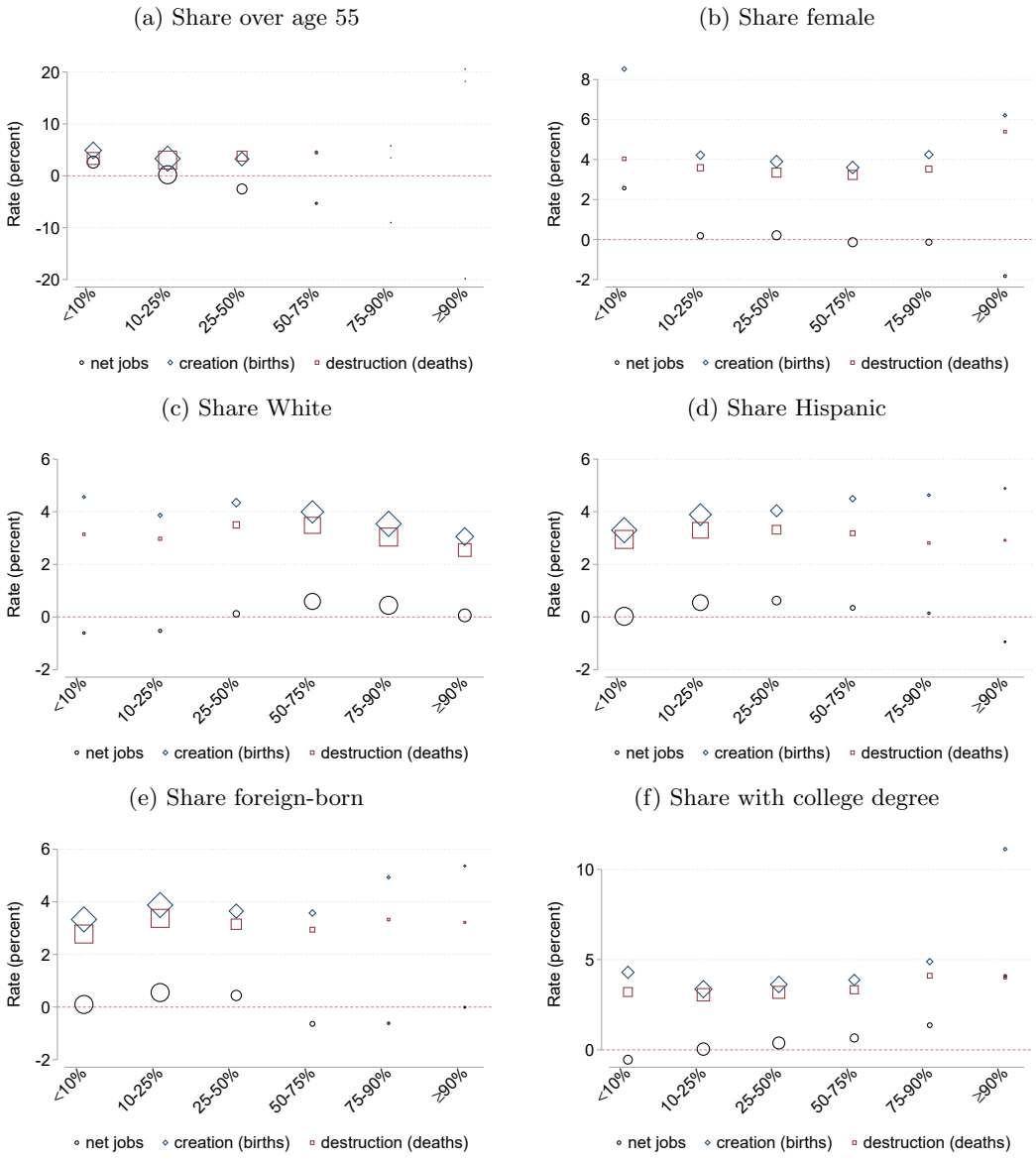
Note: This figure plots the average net job creation rate, job creation rate from births, and job destruction rate from deaths over 2006–2022 based on the share of each firm's workforce that belongs to the referenced demographic group. The size of each point reflects the magnitude of the denominator for each group, which is the average employment of establishments in the group between the current and prior year (denom). Each rate is calculated from firm-size-specific means in the workforce characteristic by firm size tables (excluding firms in the smallest "1-19" firm size category) and reweighted to reflect the firm size distribution of the overall economy. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 9: Employment change at medium and large firms by initial firm workforce composition, 2006–2022 (controlling for state and firm size)



Note: This figure plots the average net job creation rate, job creation rate from births, and job destruction rate from deaths over 2006–2022 based on the share of each firm's workforce that belongs to the referenced demographic group. The size of each point reflects the magnitude of the denominator for each group, which is the average employment of establishments in the group between the current and prior year (denom). Each rate is calculated from firm-size-specific means in the workforce characteristic by state by firm size tables (excluding firms in the smallest "1-19" firm size category) and reweighted to reflect the state and firm size distribution of the overall economy. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 10: Employment change at medium and large firms by initial firm workforce composition, 2006–2022 (controlling for sector and firm size)



Note: This figure plots the average net job creation rate, job creation rate from births, and job destruction rate from deaths over 2006–2022 based on the share of each firm's workforce that belongs to the referenced demographic group. The size of each point reflects the magnitude of the denominator for each group, which is the average employment of establishments in the group between the current and prior year (denom). Each rate is calculated from firm-size-specific means in the workforce characteristic by sector by firm size tables (excluding firms in the smallest "1-19" firm size category) and reweighted to reflect the sector and firm size distribution of the overall economy. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 11: Distribution of total employment at small, start-up firms by firm workforce age

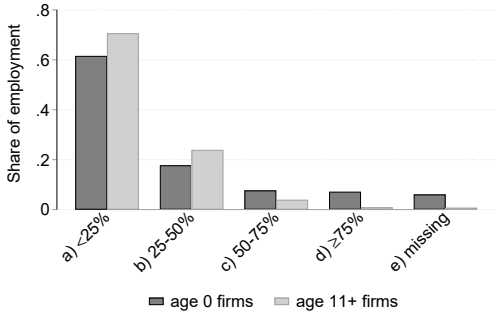
(a) Share 19–24



(b) Share 25–34



(c) Share 35–44



(d) Share 45–54



(e) Share 55–64

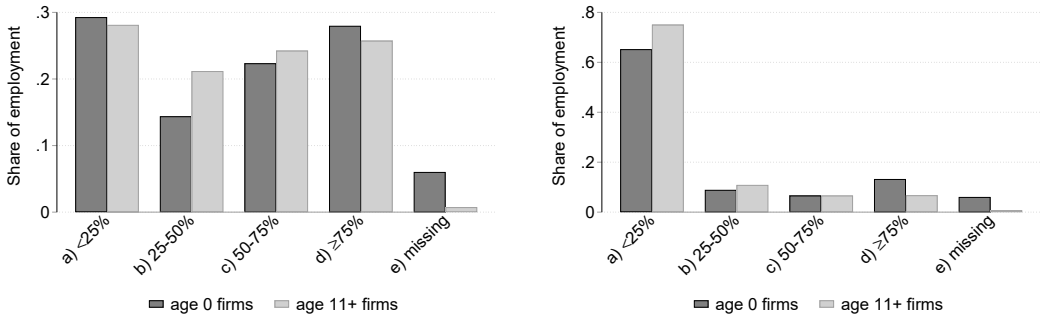


Note: This figure plots the share of employment at age 0 firms (start-ups) and age 11+ firms in 2022 based on the share of each firm's workforce that belongs to the referenced age group. Firms that cannot be matched to any workers are separately tabulated in the missing category. Firms in both age categories are restricted in size to 1-19 employees in the prior year. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 12: Distribution of total employment at small, start-up firms by firm workforce sex, nativity, ethnicity, and race

(a) Share female

(b) Share foreign-born



(c) Share Hispanic

(d) Share White





(e) Share Black

(f) Share Asian

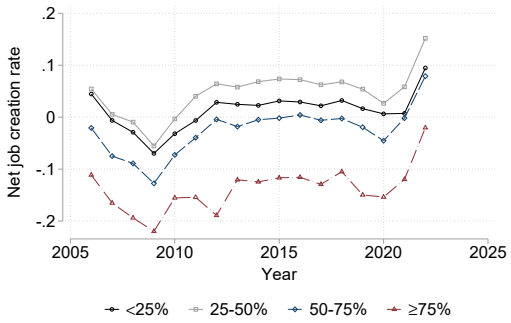




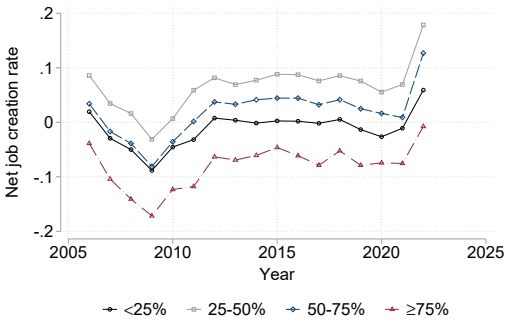
Note: This figure plots the share of employment at age 0 firms (start-ups) and age 11+ firms in 2022 based on the share of each firm's workforce that belongs to the referenced demographic group. Firms that cannot be matched to any workers are separately tabulated in the missing category. Firms in both age categories are restricted in size to 1-19 employees in the prior year. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 13: Net job creation at small, young firms by firm workforce age

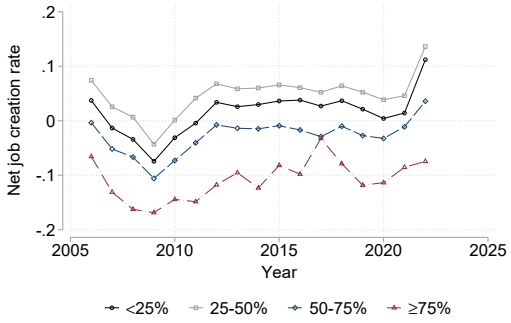
(a) Share 19–24



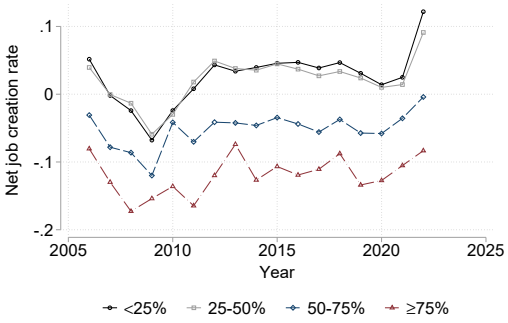
(b) Share 25–34



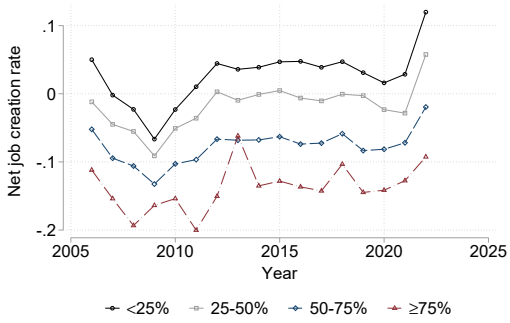
(c) Share 35–44



(d) Share 45–54

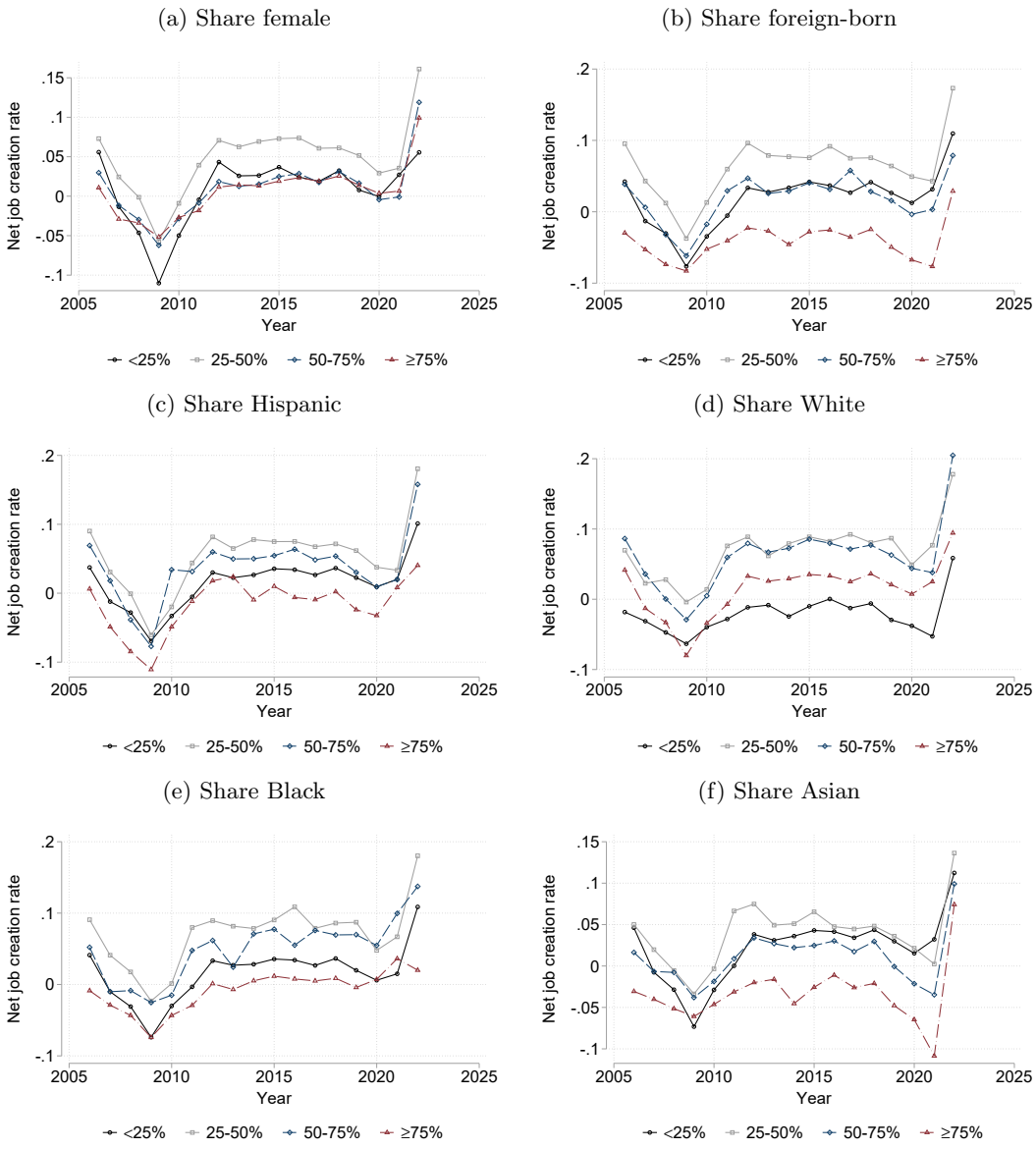


(e) Share 55–64



Note: This figure plots the net job creation rate for small (1-19 employees) and young (age 1-5) firms, stratified by the share of each firm's workforce that belongs to the referenced age group. Net job creation is the difference between the total jobs created by establishments in each category and the total jobs destroyed. Jobs can be created by existing establishments growing or existing firms opening new establishments. Jobs can be destroyed by existing establishments shrinking or existing firms closing establishments. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

Figure 14: Net job creation at small, young firms by firm workforce sex, nativity, ethnicity, and race



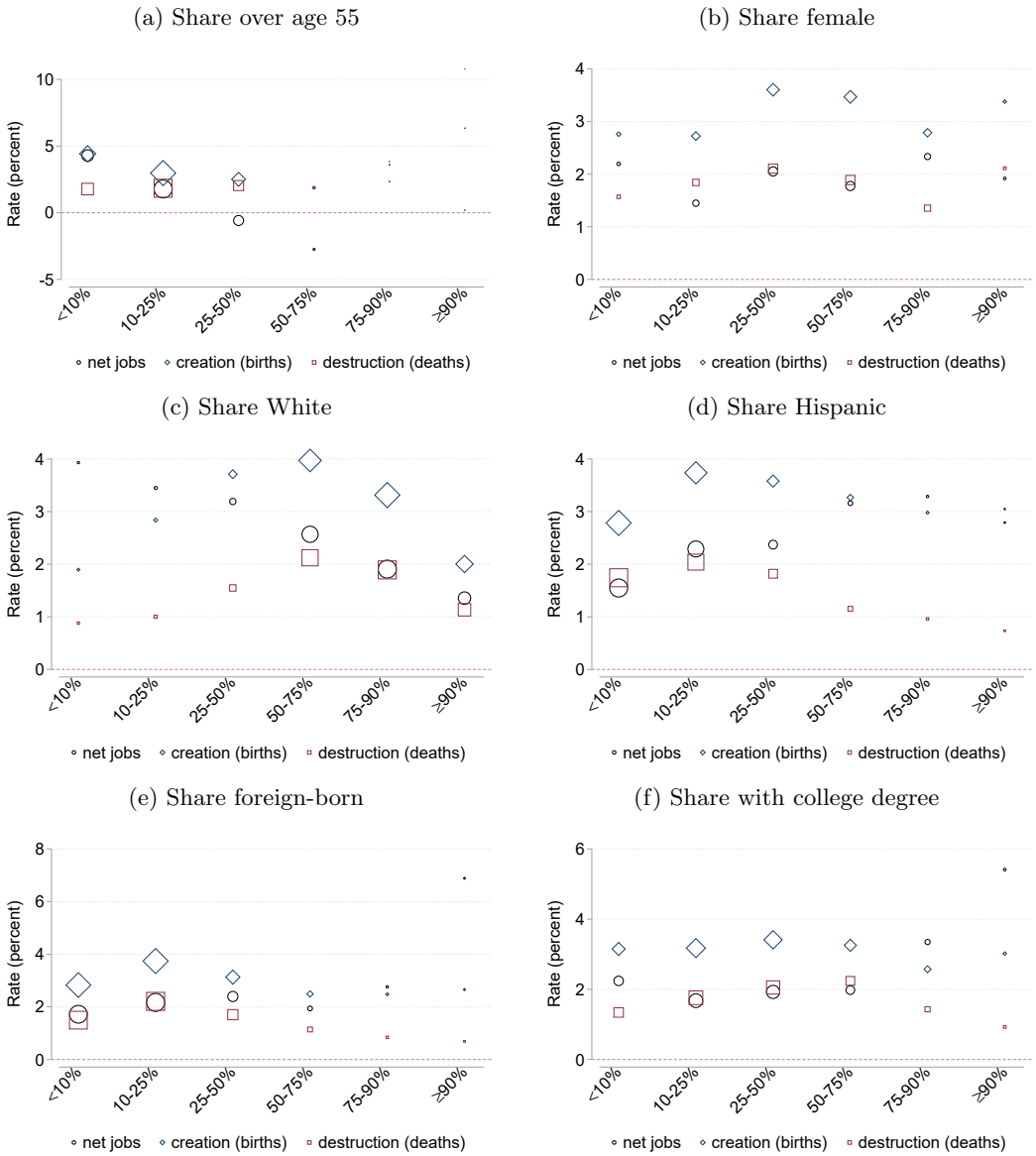
Note: This figure plots the net job creation rate for small (1-19 employees) and young (age 1-5) firms, stratified by the share of each firm's workforce that belongs to the referenced demographic group. Net job creation is the difference between the total jobs created by establishments in each category and the total jobs destroyed. Jobs can be created by existing establishments growing or existing firms opening new establishments. Jobs can be destroyed by existing establishments shrinking or existing firms closing establishments. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

A Additional Tables & Figures

Table A.1: List of BDS-HC tables

Category	Group	Variable name	Table Names				
			by Initial Firm Size Coarse	by Average Firm Size Coarse	by Firm Age Coarse and Initial Firm Size Coarse	by State & Initial Firm Size Coarse	by Sector & Initial Firm Size Coarse
Age	14–18	agein_14_18	bds2022_ifzc_ia_14	bds2022_fzc_aa_14	bds2022_fac_ifzc_ia_14	bds2022_st_ifzc_ia_14	bds2022_sec_ifzc_ia_14
Age	19–24	agein_19_24	bds2022_ifzc_ia_19	bds2022_fzc_aa_19	bds2022_fac_ifzc_ia_19	bds2022_st_ifzc_ia_19	bds2022_sec_ifzc_ia_19
Age	25–34	agein_25_34	bds2022_ifzc_ia_25	bds2022_fzc_aa_25	bds2022_fac_ifzc_ia_25	bds2022_st_ifzc_ia_25	bds2022_sec_ifzc_ia_25
Age	35–44	agein_35_44	bds2022_ifzc_ia_35	bds2022_fzc_aa_35	bds2022_fac_ifzc_ia_35	bds2022_st_ifzc_ia_35	bds2022_sec_ifzc_ia_35
Age	45–54	agein_45_54	bds2022_ifzc_ia_45	bds2022_fzc_aa_45	bds2022_fac_ifzc_ia_45	bds2022_st_ifzc_ia_45	bds2022_sec_ifzc_ia_45
Age	55–64	agein_55_64	bds2022_ifzc_ia_55	bds2022_fzc_aa_55	bds2022_fac_ifzc_ia_55	bds2022_st_ifzc_ia_55	bds2022_sec_ifzc_ia_55
Age	over 65	agein_65_99	bds2022_ifzc_ia_65	bds2022_fzc_aa_65	bds2022_fac_ifzc_ia_65	bds2022_st_ifzc_ia_65	bds2022_sec_ifzc_ia_65
Age	14–24	cagein_14_24	bds2022_ifzc_ica_14	bds2022_fzc_aca_14	bds2022_fac_ifzc_ica_14	bds2022_st_ifzc_ica_14	bds2022_sec_ifzc_ica_14
Age	25–54	cagein_25_54	bds2022_ifzc_ica_25	bds2022_fzc_aca_25	bds2022_fac_ifzc_ica_25	bds2022_st_ifzc_ica_25	bds2022_sec_ifzc_ica_25
Age	over 55	cagein_55_99	bds2022_ifzc_ica_55	bds2022_fzc_aca_55	bds2022_fac_ifzc_ica_55	bds2022_st_ifzc_ica_55	bds2022_sec_ifzc_ica_55
Sex	female	female	bds2022_ifzc_isf	bds2022_fzc_asf	bds2022_fac_ifzc_isf	bds2022_st_ifzc_isf	bds2022_sec_ifzc_isf
Race	White	race_white	bds2022_ifzc_ir_w	bds2022_fzc_ar_w	bds2022_fac_ifzc_ir_w	bds2022_st_ifzc_ir_w	bds2022_sec_ifzc_ir_w
Race	Black	race_black	bds2022_ifzc_ir_b	bds2022_fzc_ar_b	bds2022_fac_ifzc_ir_b	bds2022_st_ifzc_ir_b	bds2022_sec_ifzc_ir_b
Race	AIAN	race_aian	bds2022_ifzc_ir_a	bds2022_fzc_ar_a	bds2022_fac_ifzc_ir_a	bds2022_st_ifzc_ir_a	bds2022_sec_ifzc_ir_a
Race	Asian	race_asian	bds2022_ifzc_ir_s	bds2022_fzc_ar_s	bds2022_fac_ifzc_ir_s	bds2022_st_ifzc_ir_s	bds2022_sec_ifzc_ir_s
Race	NHPI	race_nhpi	bds2022_ifzc_ir_p	bds2022_fzc_ar_p	bds2022_fac_ifzc_ir_p	bds2022_st_ifzc_ir_p	bds2022_sec_ifzc_ir_p
Race	two or more races	race_mixed	bds2022_ifzc_ir_t	bds2022_fzc_ar_t	bds2022_fac_ifzc_ir_t	bds2022_st_ifzc_ir_t	bds2022_sec_ifzc_ir_t
Ethnicity	Hispanic	hispanic	bds2022_ifzc_ih	bds2022_fzc_ah	bds2022_fac_ifzc_ih	bds2022_st_ifzc_ih	bds2022_sec_ifzc_ih
Nativity	foreign-born	foreign	bds2022_ifzc_ifb	bds2022_fzc_afb	bds2022_fac_ifzc_ifb	bds2022_st_ifzc_ifb	bds2022_sec_ifzc_ifb
Education	less than high school	oedu_lths	bds2022_ifzc_ioe_hl	bds2022_fzc_aoe_hl	bds2022_fac_ifzc_ioe_hl	bds2022_st_ifzc_ioe_hl	bds2022_sec_ifzc_ioe_hl
Education	high school	oedu_hsgrad	bds2022_ifzc_ioe_hg	bds2022_fzc_aoe_hg	bds2022_fac_ifzc_ioe_hg	bds2022_st_ifzc_ioe_hg	bds2022_sec_ifzc_ioe_hg
Education	some college	oedu_colsome	bds2022_ifzc_ioe_cs	bds2022_fzc_aoe_cs	bds2022_fac_ifzc_ioe_cs	bds2022_st_ifzc_ioe_cs	bds2022_sec_ifzc_ioe_cs
Education	bachelor's or higher	oedu_colgrad	bds2022_ifzc_ioe_cg	bds2022_fzc_aoe_cg	bds2022_fac_ifzc_ioe_cg	bds2022_st_ifzc_ioe_cg	bds2022_sec_ifzc_ioe_cg
Age	missing data	m_age	bds2022_ifzc_im_a	bds2022_fzc_am_a	bds2022_fac_ifzc_im_a	bds2022_st_ifzc_im_a	bds2022_sec_ifzc_im_a
Sex	missing data	m_sex	bds2022_ifzc_im_sex	bds2022_fzc_am_sex	bds2022_fac_ifzc_im_sex	bds2022_st_ifzc_im_sex	bds2022_sec_ifzc_im_sex
Race	missing data	m_race	bds2022_ifzc_im_r	bds2022_fzc_am_r	bds2022_fac_ifzc_im_r	bds2022_st_ifzc_im_r	bds2022_sec_ifzc_im_r
Ethnicity	missing data	m_hispanic	bds2022_ifzc_im_h	bds2022_fzc_am_h	bds2022_fac_ifzc_im_h	bds2022_st_ifzc_im_h	bds2022_sec_ifzc_im_h
Nativity	missing data	m_foreign	bds2022_ifzc_im_fb	bds2022_fzc_am_fb	bds2022_fac_ifzc_im_fb	bds2022_st_ifzc_im_fb	bds2022_sec_ifzc_im_fb
Education	missing data	m_oedu	bds2022_ifzc_im_oe	bds2022_fzc_am_oe	bds2022_fac_ifzc_im_oe	bds2022_st_ifzc_im_oe	bds2022_sec_ifzc_im_oe

Figure A.1: Employment change at medium and large firms by average firm workforce composition, 2006–2022 (controlling for firm size)



Note: This figure is identical in structure and scope to Figure 8, but it is constructed using measures of average workforce composition and average firm size from the average workforce composition by average firm size tables rather than the initial measure used in Figure 8. For additional details on timing, see Section 3. Data source: BDS-HC tables (2022 vintage). DRB approval number: CBDRB-FY25-0043.

B Additional methodological details

In this section, we provide explicit formulas and additional details about any statistics reported in this paper that are constructed using the publicly available BDS-HC tables. Whenever possible, we use the same outcome variable naming conventions as the BDS-HC tables, and we refer to the specific cells within the tables using parentheses. For example,

$$\text{emp}_{2022}(\text{ifemale} = \text{“10–25\%”}, \text{ifirmsizecoarse} = \text{“20–499”}) \tag{1}$$

refers to the tabulated employment in the year 2022 of all establishments within the initial firm size category of 20–499 employees that have an initial workforce female share of 10–25%.

B.1 Aggregating outcomes across employer categories

In the dark bars in Figure 5, we report the total 2022 employment by each of the six firm workforce characteristics. The methods used to calculate these totals are identical for each workforce characteristic group. For each group g within each workforce characteristic category c, we calculate the total 2022 employment at establishments that belong to this group by summing the employment over all firm size categories in the corresponding initial workforce characteristic by initial firm size table. So, the total employment of a given group is given by

$$\text{emp}_t(c = g) = \sum_s \text{emp}_t(c = g, \text{ifirmsizecoarse} = s) \tag{2}$$

B.2 Estimating expected employment under random matching

In the light bars in Figure 5, we calculate the counterfactual 2022 employment by each of six firm workforce characteristics, under the assumptions that firms and workers match randomly (i.e., independently of any worker or firm characteristics) and that the distribution of employment by firm size categories reflect the observed firm size distribution in the U.S. In other words, we bound the degree of dispersion in firm workforce characteristics that can be explained by the prevalence of small firms.

Calculating the counterfactual dispersion in employment requires two main components, both of which are publicly available from existing U.S. Census Bureau data products. First, for a given workforce characteristic category c, we define the nationwide average share of workers who belong to the demographic category as $\bar{c}$. For age, sex, race, and ethnicity, we calculate these reference population shares using the corresponding nationwide employment totals from the 2021 Quarterly Workforce Indicators (QWI). However, for nativity and educational attainment, directly comparable definitions are not available from the QWI. Instead, we tabulate these shares as person-weighted means using the public-use files from the 2021 American Community Survey (ACS). Second, for the total national employment in each firm size category, we use the detailed initial firm size tables from the BDS. Similar to the BDS-HC workforce shares, the BDS partitions the range of firm sizes into intervals. We use the lower bound for each firm size interval to conservatively bound the overall degree of dispersion that can be attributed to firm size differences.

Under the assumption that firm and workers match randomly, the probability that the number of workers that belong to a given demographic category c at a firm with total employment s follows

a binomial distribution, so $P(i) = \binom{s}{i} \cdot (\bar{c})^i (1-\bar{c})^{s-i}$. It then follows that the total expected employment of firms with a workforce demographic share of n is

$$\text{emp}_t^{\text{E}}(n) = \sum_s \text{emp}_t(\text{ifz} = s) \times \binom{\sigma}{n\sigma} (\bar{c})^{n\sigma} (1-\bar{c})^{\sigma - n\sigma} \tag{3}$$

where $\text{emp}_t(\text{ifz} = s)$ is the total employment of all firms in the initial firm size category s from the BDS, σ is the lower bound of the firm size interval s, and $\bar{c}$ is the average share of the working population with a given characteristic. The expected employment of firms with a workforce demographic share between a given range (a, b) is then given as the sum of $\text{emp}_t^{\text{E}}(n)$ over the intermediate values.

Finally, a small number of employers are missing workforce characteristics because we do not have valid demographic data for any workers at the firm. As a result, they would be excluded from the tabulated employment totals based on workforce composition, but they would be included in the expected employment calculations because they contribute to overall employment in their relevant firm size category. To account for this, we adjust all expected employment totals by the share of employment that is missing each workforce characteristic (i.e., $\text{emp}_t(c = \text{“NA”})$), and we assign the residual expected employment to a corresponding “NA” group. As a result, the tabulated and expected employment in the “NA” group will always be the same for any workforce characteristic.

B.3 Constructing counterfactual employment (holding firm characteristics constant)

In Figure 7, we report the counterfactual total employment in each firm workforce characteristic group while assuming that firms do not change their workforce compositions over time. Specifically, in each year, we calculate the change in employment in a given firm workforce characteristic group due to the job creation, job destruction, entry, or exit of establishments in that group. As a result, any changes in aggregate employment due to establishments switching firm workforce composition groups would be excluded.

We construct this counterfactual measure in two steps. First, we use the corresponding initial workforce characteristic by initial firm size tables to calculate the aggregate annual employment in 2006 by aggregating across all firm size categories (as in Section B.1). We also calculate the average net job creation rate in each workforce characteristic group by taking the ratio of the total net job creation and the total average prior and current year employment (denom) across all firm size categories:

$$\text{NJCR}_t(c = g) = \frac{\sum_s \text{NJC}_t(c = g, \text{ifirmsizecoarse} = s)}{\sum_s \text{denom}_t(c = g, \text{ifirmsizecoarse} = s)} \tag{4}$$

This is consistent with the methodology used to calculate the cell-specific net job creation rates from establishment-level microdata in all BDS-HC tables, as well as the approach used in the core BDS tables and described in Chow et al. (2021).

Finally, we iteratively apply the average net job creation rates from 2007 to year t to the 2006 employment totals to calculate the counterfactual total employment at t:

$$\text{emp}_t^{\text{CF}}(c = g) = \text{emp}_{2006}(c = g) \times \prod_{\tau=2007}^{t} \frac{1 + \text{NJCR}_\tau(c = g)/200}{1 - \text{NJCR}_\tau(c = g)/200} \tag{5}$$

B.4 Comparing average job growth within firm size (or state-by-firm-size or sector-by-firm-size) categories

In Figures 8, 9, and 10, we calculate the average net job creation rate, job creation rate from births, and job destruction rate from deaths for each workforce characteristic group while controlling for group-level differences in firm size (in Figure 8), state and firm size (in Figure 9), both sector and firm size (in Figure 10). The key to this comparison is that we reweight all job growth rates for the interacted tabulation variables within each workforce characteristic group using a common, economy-wide employment distribution for the interacted variables. As a result, any differences in growth rates between any two workforce characteristic groups can only be due to between-group differences in growth rates within the same firm size (or state-by-firm-size or sector-by-firm-size) category, rather than the differential distribution in employment across different categories.

Specifically, for each group g within each workforce characteristic category c, we calculate the average firm-size-invariant net job creation rate as

$$\text{NJCR}_t^{\text{FS}}(c=g) = \sum_s \underbrace{\frac{\text{denom}_t(\text{ifirmsizecoarse}=s)}{\text{denom}_t}}_{\equiv\omega_t(\text{ifirmsizecoarse}=s)} \times \text{NJCR}_t(c=g, \text{ifirmsizecoarse}=s) \tag{6}$$

where $\text{denom}_t(\text{ifirmsizecoarse}=s)$ is the economy-wide average prior and current year employment (denom) of all establishments in firm size category s, as reported in the BDS tables. The calculations for the average job creation rate from births and job destruction rate from deaths are analogous. To see that this approach isolates the between-group differences in growth rates within the same firm size category, notice that the difference in firm-size-invariant net job creation rates between two groups can be written as

$$\begin{aligned}
\Delta\text{NJCR}_t(g,g') =& \text{NJCR}_t^{\text{FS}}(c=g) - \text{NJCR}_t^{\text{FS}}(c=g') \\
=& \sum_s \omega_t(\text{ifirmsizecoarse}=s) \times \text{NJCR}_t(c=g, \text{ifirmsizecoarse}=s) \\
& - \sum_s \omega_t(\text{ifirmsizecoarse}=s) \times \text{NJCR}_t(c=g', \text{ifirmsizecoarse}=s) \\
=& \sum_s \omega_t(\text{ifzc}=s) \times \left[\text{NJCR}_t(c=g, \text{ifzc}=s) - \text{NJCR}_t(c=g', \text{ifzc}=s)\right].
\end{aligned}$$

As a result, any overall differences in the firm-size-invariant net job creation rate between two workforce characteristic groups can only arise if they have different net job creation rates within the same firm size category, weighted by the relative weight of each category over the total, economy-wide employment.

We calculate the additional job dynamics rates in Figure 9 and 10 using a similar approach from the corresponding three-way tables. For example, the average net job creation rate, controlling for state and firm size, is defined as

$$\text{NJCR}_t^{\text{ST-FS}}(c=g) = \sum_s \underbrace{\frac{\text{denom}_t(\text{st}=j, \text{ifzc}=s)}{\text{denom}_t}}_{\equiv\omega_t(\text{st}=j,\text{ifzc}=s)} \times \text{NJCR}_t(c=g, \text{st}=j, \text{ifzc}=s) \tag{7}$$

where $\text{denom}_t(\text{st}=j, \text{ifirmsizecoarse}=s)$ is the economy-wide average prior and current year

employment (denom) of all establishments in state j and firm size category s, as reported in the BDS tables. Meanwhile, the average net job creation rate, controlling for sector and firm size, is defined as

$$\text{NJCR}_t^{\text{SEC-FS}}(c=g) = \sum_s \underbrace{\frac{\text{denom}_t(\text{sector}=j, \text{ifzc}=s)}{\text{denom}_t}}_{\equiv \omega_t(\text{sector}=j, \text{ifzc}=s)} \times \text{NJCR}_t(c=g, \text{sector}=j, \text{ifzc}=s) \quad (8)$$

where $\text{denom}_t(\text{sector}=j, \text{ifirmsizecoarse}=s)$ is the economy-wide average prior and current year employment (denom) of all establishments in sector j and firm size category s, as reported in the BDS tables.

C Codebook

The definitions of all workforce composition variables are as follows. For all other variable definitions, including the definition of other firm characteristics and outcomes, see Chow et al. (2021).

Classification values

We use a consistent set of classification groups for every demographic category in the BDS-HC tables. For each demographic category, we classify all firms into one of seven groups based on the share of their workforce that belongs to that category. The classification groups are defined as the following for all BDS-HC workforce composition variables:

a) **less than 10%**: Firm has less than 10% of workers belonging to the demographic category

b) **10% to 25%**: Firm has 10-25% of workers belonging to the demographic category

c) **25% to 50%**: Firm has 25-50% of workers belonging to the demographic category

d) **50% to 75%**: Firm has 50-75% of workers belonging to the demographic category

e) **75% to 90%**: Firm has 75-90% of workers belonging to the demographic category

f) **more than 90%**: Firm has more than 90% of workers belonging to the demographic category

z) **not available**: Firm does not have matched data on workers in this demographic category

Initial workforce composition variables

Age

- **iagein_14_18**: Initial firm workforce share age 14–18. The share of workers with valid age data at the firm at year t-1 who are age 14–18.
- **iagein_19_24**: Initial firm workforce share age 19–24. The share of workers with valid age data at the firm at year t-1 who are age 19–24.
- **iagein_25_34**: Initial firm workforce share age 25–34. The share of workers with valid age data at the firm at year t-1 who are age 25–34.

- **iagein_35_44**: Initial firm workforce share age 35–44. The share of workers with valid age data at the firm at year t-1 who are age 35–44.
- **iagein_45_54**: Initial firm workforce share age 45–54. The share of workers with valid age data at the firm at year t-1 who are age 45–54.
- **iagein_55_64**: Initial firm workforce share age 55–64. The share of workers with valid age data at the firm at year t-1 who are age 55–64.
- **iagein_65_99**: Initial firm workforce share over age 65. The share of workers with valid age data at the firm at year t-1 who are over age 65.
- **icagein_14_24**: Initial firm workforce share age 14–24. The share of workers with valid age data at the firm at year t-1 who are age 14–24.
- **icagein_25_54**: Initial firm workforce share age 25–54. The share of workers with valid age data at the firm at year t-1 who are age 25–54.
- **icagein_55_99**: Initial firm workforce share over age 55. The share of workers with valid age data at the firm at year t-1 who are over age 55.

Sex

- **ifemale**: Initial firm workforce share female. The share of workers with valid sex data at the firm at year t-1 who are female.

Race

- **irace_white**: Initial firm workforce share White. The share of workers with valid race data at the firm at year t-1 who are White.
- **irace_black**: Initial firm workforce share Black. The share of workers with valid race data at the firm at year t-1 who are Black.
- **irace_aian**: Initial firm workforce share American Indian/Alaska Native. The share of workers with valid race data at the firm at year t-1 who are American Indian/Alaska Native.
- **irace_asian**: Initial firm workforce share Asian. The share of workers with valid race data at the firm at year t-1 who are Asian.
- **irace_nhpi**: Initial firm workforce share Native Hawaiian/Pacific Islander. The share of workers with valid race data at the firm at year t-1 who are Native Hawaiian/Pacific Islander.
- **irace_mixed**: Initial firm workforce share two or more races. The share of workers with valid race data at the firm at year t-1 who are two or more races.

Ethnicity

- **ihispanic**: Initial firm workforce share Hispanic. The share of workers with valid ethnicity data at the firm at year t-1 who are Hispanic.

Nativity

- **iforeign**: Initial firm workforce share foreign-born. The share of workers with valid nativity data at the firm at year t-1 who are foreign-born.

Education

- **ioedu_lths**: Initial firm workforce share with less than high school degree. The share of workers born before 1976 with valid education data at the firm at year t-1 whose highest educational attainment is less than a high school degree.
- **ioedu_hsgrad**: Initial firm workforce share with high school degree. The share of workers born before 1976 with valid education data at the firm at year t-1 whose highest educational attainment is a high school degree.
- **ioedu_colsome**: Initial firm workforce share with some college. The share of workers born before 1976 with valid education data at the firm at year t-1 whose highest educational attainment is some college (less than a bachelor's).
- **ioedu_colgrad**: Initial firm workforce share with bachelor's degree or higher. The share of workers born before 1976 with valid education data at the firm at year t-1 whose highest educational attainment is a bachelor's degree or higher.

Missing data

- **im_age**: Initial firm workforce share missing age data. The share of workers at the firm at year t-1 who are missing age data.
- **im_sex**: Initial firm workforce share missing sex data. The share of workers at the firm at year t-1 who are missing sex data.
- **im_race**: Initial firm workforce share missing race data. The share of workers at the firm at year t-1 who are missing race data.
- **im_hispanic**: Initial firm workforce share missing ethnicity data. The share of workers at the firm at year t-1 who are missing ethnicity data.
- **im_foreign**: Initial firm workforce share missing nativity data. The share of workers at the firm at year t-1 who are missing nativity data.
- **im_oedu**: Initial firm workforce share missing education data. The share of workers at the firm at year t-1 who are born after 1976 or missing education data.

Average workforce composition variables

Age

- **aagein_14_18**: Average firm workforce share age 14–18. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are age 14–18.
- **aagein_19_24**: Average firm workforce share age 19–24. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are age 19–24.

- **aagein_25_34**: Average firm workforce share age 25–34. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are age 25–34.
- **aagein_35_44**: Average firm workforce share age 35–44. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are age 35–44.
- **aagein_45_54**: Average firm workforce share age 45–54. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are age 45–54.
- **aagein_55_64**: Average firm workforce share age 55–64. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are age 55–64.
- **aagein_65_99**: Average firm workforce share over age 65. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are over age 65.
- **acagein_14_24**: Average firm workforce share age 14–24. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are age 14–24.
- **acagein_25_54**: Average firm workforce share age 25–54. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are age 25–54.
- **acagein_55_99**: Average firm workforce share over age 55. The employment-weighted average share of workers with valid age data at the firm at year t-1 and year t who are over age 55.

Sex

- **afemale**: Average firm workforce share female. The employment-weighted average share of workers with valid sex data at the firm at year t-1 and year t who are female.

Race

- **arace_white**: Average firm workforce share White. The employment-weighted average share of workers with valid race data at the firm at year t-1 and year t who are White.
- **arace_black**: Average firm workforce share Black. The employment-weighted average share of workers with valid race data at the firm at year t-1 and year t who are Black.
- **arace_aian**: Average firm workforce share American Indian/Alaska Native. The employment-weighted average share of workers with valid race data at the firm at year t-1 and year t who are American Indian/Alaska Native.
- **arace_asian**: Average firm workforce share Asian. The employment-weighted average share of workers with valid race data at the firm at year t-1 and year t who are Asian.
- **arace_nhpi**: Average firm workforce share Native Hawaiian/Pacific Islander. The employment-weighted average share of workers with valid race data at the firm at year t-1 and year t who are Native Hawaiian/Pacific Islander.
- **arace_mixed**: Average firm workforce share two or more races. The employment-weighted average share of workers with valid race data at the firm at year t-1 and year t who are two or more races.

Ethnicity

- **ahispanic**: Average firm workforce share Hispanic. The employment-weighted average share of workers with valid ethnicity data at the firm at year t-1 and year t who are Hispanic.

Nativity

- **aforeign**: Average firm workforce share foreign-born. The employment-weighted average share of workers with valid nativity data at the firm at year t-1 and year t who are foreign-born.

Education

- **aoedu_lths**: Average firm workforce share with less than high school degree. The employment-weighted average share of workers born before 1976 with valid education data at the firm at year t-1 and year t whose highest educational attainment is less than a high school degree.
- **aoedu_hsgrad**: Average firm workforce share with high school degree. The employment-weighted average share of workers born before 1976 with valid education data at the firm at year t-1 and year t whose highest educational attainment is a high school degree.
- **aoedu_colsome**: Average firm workforce share with some college. The employment-weighted average share of workers born before 1976 with valid education data at the firm at year t-1 and year t whose highest educational attainment is some college (less than a bachelor's).
- **aoedu_colgrad**: Average firm workforce share with bachelor's degree or higher. The employment-weighted average share of workers born before 1976 with valid education data at the firm at year t-1 and year t whose highest educational attainment is a bachelor's degree or higher.

Missing data

- **am_age**: Average firm workforce share missing age data. The employment-weighted average share of workers at the firm at year t-1 and year t who are missing age data.
- **am_sex**: Average firm workforce share missing sex data. The employment-weighted average share of workers at the firm at year t-1 and year t who are missing sex data.
- **am_race**: Average firm workforce share missing race data. The employment-weighted average share of workers at the firm at year t-1 and year t who are missing race data.
- **am_hispanic**: Average firm workforce share missing ethnicity data. The employment-weighted average share of workers at the firm at year t-1 and year t who are missing ethnicity data.
- **am_foreign**: Average firm workforce share missing nativity data. The employment-weighted average share of workers at the firm at year t-1 and year t who are missing nativity data.
- **am_oedu**: Average firm workforce share missing education data. The employment-weighted average share of workers at the firm at year t-1 and year t who are born after 1976 or missing education data.