Who Hires Whom? Entrepreneurial Backgrounds and Labor Market Opportunities*
(PRELIMINARY AND INCOMPLETE, PLEASE DO NOT CIRCULATE OR CITE)

January 2025

Abstract

What are the implications of unequal access to entrepreneurial careers for labor markets? Using data from the U.S. Census and LinkedIn profiles, we document that entrepreneurs are significantly more likely to hire workers from similar social backgrounds (gender, race, age, education, etc.). These effects are quantitatively large across the demographic dimensions available to us. For example, female employee share at female founded startups are 36.4pp higher after controlling for industry-by-metro area-by-cohort, with the corresponding estimates of 51.2pp for Blacks, 37.3pp for Hispanics, and 11.3pp for non-college individuals. Large effects are present in high-growth startups, across industries and occupations, and remain stable across new firm cohorts. In addition, differences across new firms in the composition of their workers strongly persist over the firm's first ten years. We use wage data to untangle whether the relative differences are driven by labor demand or labor supply effects: group-specific wage decompositions suggest that new firms pay higher relative wages to individuals from similar backgrounds to the entrepreneur. We calibrate a labor demand model using our empirical findings to assess the potential wage impacts of reducing access barriers to entrepreneurial careers, finding quantitatively significant increases in relative wage for women (2.77 percent), Blacks (4.36 percent), and Hispanics (2.26 percent).

*Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2229. This research uses data from the Census Bureau's Longitudinal Employer-Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation.

How does the composition of entrepreneurs in an economy impact individuals' labor market opportunities and outcomes? Existing work by Haltiwanger et al. (2013) has documented that young firms play a significant role in net job creation in the economy. In a separate line of work, researchers have documented that specific socioeconomic groups (women, low SES individuals) are dramatically underrepresented in the pool of entrepreneurs and innovators (e.g., Bell et al., 2017; Guzman and Kacperczyk, 2018; Cook et al., 2022; Bento and Hwang, 2023; Bento, 2024). If entrepreneurs are better at managing or creating organizational structures that make use of workers with similar backgrounds to them, then the established facts in the literature would imply that individuals from certain backgrounds would face fewer labor market opportunities. To date, there has been scarce systematic evidence on this issue, given the difficulty in assembling systematic data on the demographics of entrepreneurs and workers.[1]

In this article, we extend this line of research by studying the impacts of "unequal access" to the innovation system on workers. To do so, we compile two complementary datasets on new firms and the demographics of their founders and workers. Our primary data source comes from the Longitudinal Employer-Household Dynamics (LEHD) data collected by the U.S. Census. We also compile a complementary dataset is based on LinkedIn career history data compiled by Revelio Labs, linked to information on founders from Crunchbase, a tracker of venture-backed startups. Using both datasets, we first test whether entrepreneurs are more likely to hire individuals of similar socioeconomic backgrounds, at least in the early years of each firm. We then explore heterogeneity by firm characteristics and worker roles, in order to gain further insights into the potential mechanisms driving the high level results. Next, we use wage data from the LEHD to provide evidence on equilibrium wage differences at the firm level, in order to test whether labor supply or labor demand factors play a larger role in generating the relative quantity of hiring. Finally, we embed our empirical estimates into a model of the labor market in order to assess the potential implications of policies that increase entrepreneurship in the underrepresented groups.

[1]The closest work to ours is Chiplunkar et al. (2024), which documents that female entrepreneurship in India leads to greater labor force participation among women in India. As discussed below, we aim to document this in the U.S. context and use wage data to decompose supply and demand effects. Our work is also related to Gozen et al. (2024), which analyzes differences between male- and female-owned businesses using historical U.S. Census of Manufactures data from the 19th century.

To start, we combine work history data with information about new firms to connect entrepreneurs to the workers at their firms. For the Census sample, we identify new firms using business registry data, recording their industry and location. Next, we identify entrepreneurs using the methodology in Kerr and Kerr (2016). We then construct demographic measures associated with the non-entrepreneur employees at each firm in each year relative to its founding (e.g., 10 percent Hispanic two years after founding). We follow a similar approach using a merge of Crunchbase and LinkedIn work history data.

Next, we establish our core results on the similarity between founders and their employees. We compute simple correlations between the demographics of the founders (e.g., one-third female) and the demographics of the employees at the firm in a given year (e.g., two-thirds female), both raw and adjusted by founding year, location, and industry. We do so for a series of demographic variables: age, gender, race, and college educated. Throughout, we find large "homophily" estimates across all demographic variables. For example, within founding year by metropolitan statistical area (MSA) by industry groups, female-founded firms have a 37pp higher fraction of female workers during the founding year, decreasing to a 36pp difference after five years. We find large estimates in the sample of highest growing firms, in the venture-backed startup sample, within the highest paid and most senior employees, and across founding cohorts.

The core results suggest different relative equilibrium quantities in workers, which we then attempt to decompose into supply versus demand effects. To do so, we run wage decomposition regressions within demographic subgroups (e.g., college educated versus not) to extract firm fixed effects using the approach in Abowd et al. (1999). Using the output, we then compute the difference in firm fixed effects for each demographic (e.g., college-educated estimate minus the estimate for non-college-educated) and run a simple regression versus founder characteristics (e.g., one-half college educated). We find large and positive relative wage effects for workers that share demographics with the founders, ranging from 9.1 percent to 23.3 percent for difference demographic variables. This suggests that the relative demand dominates the relative supply in terms of generating the observed equilibrium quantities. In other words, founders exhibit relatively higher demand for workers from similar backgrounds.

Finally, we build our core findings into a model to assess the impacts of unequal access to firm creation on labor market opportunities and outcomes. Given that labor demand is the stronger force in this setting, we model entrepreneurs as creating a firm production function that makes workers from similar demographic backgrounds more productive and exhibits imperfect substitution between workers from different groups.[2] We can then simulate counterfactuals under different compositions of founders to assess the impact on

We contribute to three strands of literature. First, we extend the literature on unequal access to the innovation system. Existing work in the literature has documented that women, minorities, and individuals from low SES backgrounds are underrepresented in the pool of entrepreneurs and inventors (e.g., Bell et al., 2017; Guzman and Kacperczyk, 2018; Bento and Hwang, 2023; Bento, 2024). Recent work has also documented a significant increase in black and Hispanic entrepreneurship in the past few years.[3] Building on this line of work, several articles have documented that innovators tend to innovate for consumers similar to them (e.g., Koning et al., 2021; Einio et al., 2023). Our proposed study aims to complement the existing work by studying the implications of "unequal access" to the innovation system for workers rather than consumers.[4]

Second, we contribute the literature on new firm creation and hiring. Initial work in the area noted the importance of young firms for net job creation (e.g., Haltiwanger et al., 2013). Newer work in the literature has documented additional issues surrounding pay at startups (e.g., Kim, 2018; Roach and Sauermann, 2023) and the behavior and importance of early workers at startups (Bryan et al., 2022; Choi et al., 2023; Bessen et al., 2023). Our proposed work aims to add additional richness to this line of work by highlighting the workers that gain from new firm creation using systematic data.

Finally, we will also address the literature on labor market opportunities and outcomes. A long line of literature has documented the drivers of the gender wage gap (e.g., Goldin and Rouse, 2000; Tate and Yang, 2015; Card et al., 2016) and the wage differences across racial

[2]This is similar in spirit to the framework in Chiplunkar et al. (2024). We are better able to distinguish between supply and demand effects.

[3]See Brookings Report "Who is driving Black business growth? Insights from the latest data on Black-owned businesses", available at `https://www.brookings.edu/articles/who-is-driving-black-business-growth-insights-from-the-latest-data-on-black-owned-businesses/`.

[4]The closest work in this area is Marinoni and Voorheis (2019), which studies the impact of entrepreneurship on local earnings inequality. Our proposed research focuses more on heterogeneity across startups and the unequal impacts on workers from different backgrounds.

and ethnic groups (e.g., Chetty et al., 2020). This includes specific work on the impact of manager's social backgrounds and experiences on worker outcomes at firms (Tate and Yang, 2015; Ronchi and Smith, 2024). More recent work has studied female entrepreneurship, hiring, and wage setting in other contexts (Chiplunkar et al., 2024; Gozen et al., 2024). Our work aims to assess the role played by entrepreneurs, specifically the types of organizations created by the current pool of entrepreneurs. We conjecture that unequal access to entrepreneurship careers can induce lasting inequalities across workers in terms of job opportunities and wages across socio-demographic groups.

I Data Sources and Methodological Approach

I.A Census LEHD Data

First, we make use of the Longitudinal Employer-Household Dynamics (LEHD) data from the U.S. Census Bureau. The Census data has several advantages. First, it contains administrative wage data, which makes it more complete than LinkedIn data in terms of measuring workers and provides a reliable way to adjudicate between labor supply and labor demand effects. Second, it contains demographic information on age, race/ethnicity, gender, immigration status, and education level, collected from the Decennial Census. Finally, Census data also contains information on industry and location. A disadvantage of the LEHD data is that it is only available for 30 states and the District of Columbia, which presents both measurement and representativeness issues.

Following Kerr and Kerr (2016), we employ the following steps to identify new firms and their founders. First, we identify the top three earners in the company's first year in the LEHD data and record their demographic profiles. Second, because we only have access to data from a select set of states, we filter out companies that had earlier entries in other Census datasets to avoid cases where an established company in an uncovered state opens a new establishment in a covered state. Finally, we compute average demographics of other employees in each year relative to the firm's first year in the Census data.

Table 1 presents summary statistics from a balanced panel of new firms covering the period 1992–2014. We require firms in the balanced panel to have non-founder employees in

Table 1: Summary Statistics – New Firms in LEHD

Panel A: Firm Level

	Mean	S.D.	Pseudo Median
Female Founder Fraction	0.408	0.367	0.333
Black Founder Fraction	0.054	0.170	0.000
Hispanic Founder Fraction	0.105	0.227	0.000
Average Founder Age	38.32	9.142	38.00
College Grad Founder Fraction	0.281	0.296	0.333
Change in Log Employees	0.484	1.049	0.406
Observations		825,000	

Panel B: Firm-by-Year Level

	Mean	S.D.	Pseudo Median
Female Employee Share	0.501	0.332	0.500
Black Employee Share	0.075	0.166	0.000
Hispanic Employee Share	0.131	0.213	0.023
Average Employee Age	32.92	8.722	32.33
College Grad Employee Share	0.226	0.213	0.186
Total Employees	29.06	188.7	10.000
Observations		4,947,000	

Notes: Panel A presents statistics at the new firm level for the balanced sample of firms. These are mainly demographics of the founders but also includes the growth in employees from year 0 to year 5. Panel B presents statistics at the firm-by-year level, focusing on employee demographics in each year. Pseudo medians are averages of data points closest to the median. All statistics are unweighted averages. Observation numbers are rounded based on Census disclosure rules. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

years 0–5 relative to founding and have MSA and industry information. This ensures that we can estimate trends in homophily as firms age without incorporating selection effects from exit and can incorporate detailed fixed effects in our regressions.[5] We use this as the core sample throughout.[6]

The summary stats highlight discrepancies between founder and employee demographics: women, Blacks, Hispanics, younger individuals, and non-college individuals are underrepresented in the founder statistics relative to their fraction in the new-firm employee pool and the general population. In terms of firm size, the average firm grows 48 percent in terms of

[5]As a practical matter, Census disclosures require a minimal number of samples. This approach also implicitly selects for larger and relatively successful firms.

[6]We also provide results using an unbalanced sample of new firms and extend the relative years to 9 years after founding.

employees from year 0 to year 5. The average employees at the firm-year level is 30.

I.B LinkedIn and Crunchbase Data

Second, we use LinkedIn data compiled by Revelio Labs matched to Crunchbase startup data as a complementary data source.[7] Our Revelio Labs data is a April 2023 snapshot of LinkedIn and our Crunchbase data comes from a December 2020.

The key advantages of the Revelio Labs data is that it contains raw and standardized information on the role of the individual at a given firm (nature of job and seniority), which we cannot observe in the LEHD data. This allows us to test whether our observed effects are driven by people in certain roles (e.g., scientists vs. human resources) or seniority levels. Another advantage is that we can identify founders based on LinkedIn job titles, instead of guessing based on the top paid individuals at a firm.

However, the data come with several limitations. First, individuals select into having LinkedIn profiles. Second, the data do not contain wage information. Third, the demographic variables are constructed based on names rather than reported through the decennial Census, and there are fewer readily-available characteristics versus the Census. Finally, the LinkedIn population is heavily skewed towards college-educated individuals.

We again construct two samples. First, we construct a balanced sample of firms that have non-founder employees with LinkedIn profiles in each year up to 5 years after founding. We also construct an unbalanced panel of all Crunchbase firms. Table 2 reports summary statistics corresponding to those in Table 1 for the balanced sample. We again find that employee shares are smaller than average founder demographics, although the differences are much larger for women than for Blacks and Hispanics. Firms in this sample exhibit higher employment growth over the same six year period and have more employees on average. This is unsurprising given that the Crunchbase sample is skewed towards venture-backed and technology firms and we are ensuring survival in our balanced sample.

[7] Revelio Labs provided a crosswalk from their firm ID to Crunchbase firm IDs, which we rely on for our analysis.

Table 2: Summary Statistics – New Firms in the Crunchbase-LinkedIn Sample

Panel A: Firm Level

	Mean	S.D.	Median
Female Founder Fraction	0.20	0.32	0.01
Black Founder Fraction	0.08	0.15	0.02
Hispanic Founder Fraction	0.07	0.17	0.01
Change in Log Employees	1.21	1.31	1.16
Observations		44545	

Panel B: Firm-by-Year Level

	Mean	S.D.	Median
Female Employee Share	0.35	0.26	0.33
Black Employee Share	0.09	0.10	0.06
Hispanic Employee Share	0.09	0.14	0.05
Total Employees	85.14	1910.43	10.00
Observations		267270	

Notes: Panel A presents statistics at the new firm level for the balanced sample of firms. These are mainly demographics of the founders but also includes the growth in employees from year 0 to year 5. Panel B presents statistics at the firm-by-year level, focusing on employee demographics in each year. Source: Crunchbase-LinkedIn merger.

I.C Methodological Approach

To estimate whether founders hire individuals from similar backgrounds, we employ a simple specification:

$$\text{WorkerType}_{it} = \alpha + \beta_t \cdot \text{EntrepreneurType}_i + \mu_k + \varepsilon_{it} \tag{1}$$

where i indexes the firm, t the age of the firm (zero indexed), and μ_k represents some level of fixed effects. "WorkerType" measures the fraction of workers who are of a given characteristic (e.g., female) and "EntrepreneurType" is the corresponding measure within the set of founders. To start, we analyze new firms in the first year of their creation. Following Kerr and Kerr (2016), which stresses the importance of geography in determining startup outcomes, we use both detailed industry fixed effects (e.g., six-digit NAICS) and two-digit industry-by-MSA level fixed effects when analyzing the Census sample.

Using this as the building block, we expand along several dimensions. First, we assess the robustness of the results to different weighting schemes (wages of individuals, size of company, etc.). Second, we analyze how β evolves as firms age. We conjecture that differences in hiring practice can be long lasting and persist even as the firm grows larger – for example because initial hires may establish a different culture that may be more or less open to minorities and women. Third, estimate β by the founding year of companies to document trends over time. Fourth, we assess heterogeneity along several firm (e.g., industry, trading vs. non-trading firms) and worker dimensions (e.g., type of role, seniority, wage quintile).

II Hiring Analysis

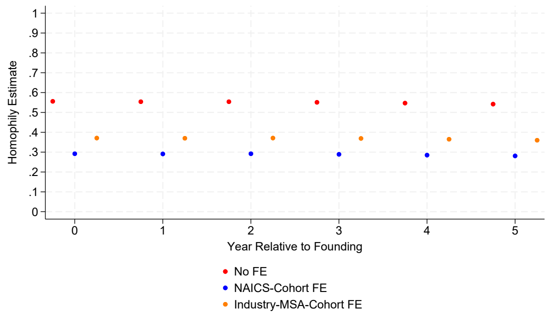
II.A Core Results – LEHD

Figure 1 presents the core estimates of Equation 1 from the Census LEHD balanced sample.

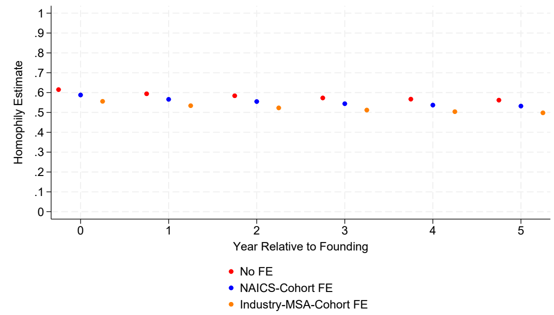
We note three general patterns in the estimates. First, all estimates are quantitatively large when compared to the baseline rate of employees presented in Table 1. The estimate for female founder in year 0 with the most detailed fixed effects is 0.37 on a baseline rate of 0.35. The corresponding numbers for Black founders are 0.556 on a baseline of 0.075 and for Hispanic founders 0.406 on 0.13. We also find large effects for Asian (0.574), age (0.364), and college-educated (0.118). Second, adding detailed controls reduces the quantitative estimates relative to the specification with no controls, but the estimates remain large even with the most detailed comparison. This suggests that a significant part of the raw estimate is driven by differences within very similar firms. Third, we note that the effects are highly persistent over the first few years at a firm.

We assess the robustness of the results in several ways. First, we re-run our analysis on an unbalanced panel of new firms and report the results from the most detailed fixed effect specification in Figure A1. We again find large effects of similar magnitudes (e.g., the female estimate is 0.381 in year 0 of the unbalanced panel, 0.566 for Black and 0.395 for Hispanic). We also find that the effects persist strongly out to the tenth year of the firm, albeit slightly attenuated (e.g., 0.250 in year 9 versus 0.381 in year 0 for female and 0.433 versus 0.566 for Black). Because the results use an unbalanced panel, the decay will reflect both within firm changes and differential attrition across types of organizations.

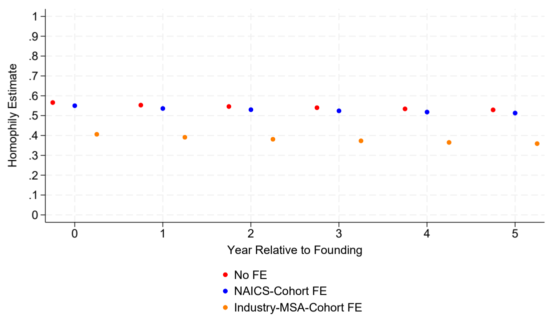
Second, we test the robustness of our estimates to how many top paid employees we count as founders and report the results in Table A1. The estimates become smaller when



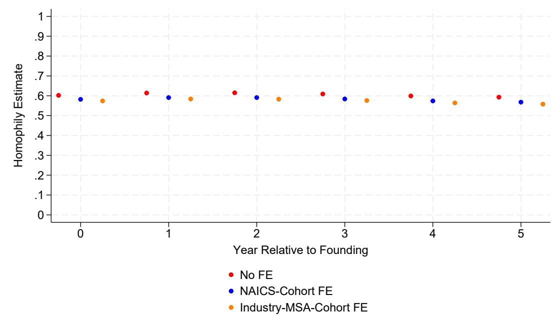
(a) Female



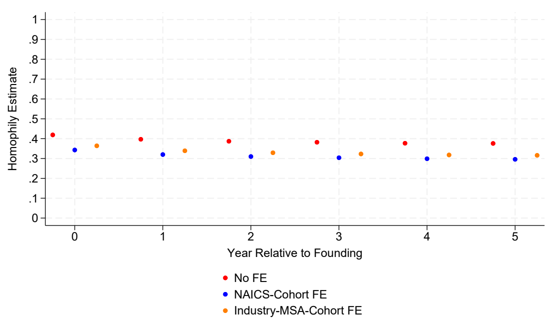
(b) Black



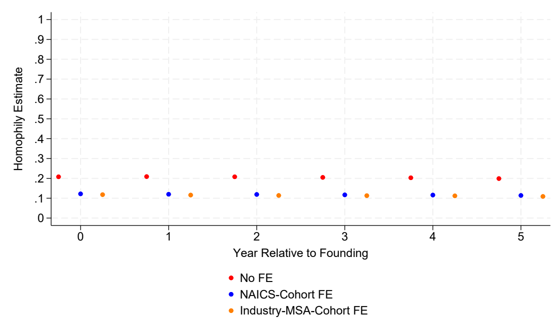
(c) Hispanic



(d) Asian



(e) Age



(f) College Educated

Figure 1: Estimates from Balanced Sample of New Firms

Notes: estimates of Equation 1 using the balanced LEHD sample. All standard errors are negligible relative to the point estimates, so we have not added confidence intervals to the plot. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

assuming that only the top paid individual is the founder, but the point estimates remain quantitatively large. This could be driven by patterns in relative pay across co-founders or our misclassification of early employees as co-founders. Finally, we also add weighting to our regressions to emphasize larger firms and higher-paid individuals. We first compute weighted average employee demographics by weighting by employee income. Then, we weight the core regression by the square root of total employees to give more weight to larger firms. Table A2 provides estimates when applying both weighting schemes. The point estimates are again very similar to the core estimates (0.360 for female, 0.544 for Black, and 0.377 for Hispanic).

II.B Core Results – Crunchbase

Table 3 presents the core results using the balanced panel from the LinkedIn-Crunchbase match. We find quantitatively large differences in early employee composition across firms. In the year of founding, female-founded firms have 17.4pp higher fraction of female employees, on a baseline of 35pp (from Table 2). Likewise, Black-founded companies have an 10.4pp higher fraction of black employees, relative to a baseline of 9pp, and Hispanic-founded companies have a 24.3pp higher fraction of Hispanic workers relative to a baseline of 9pp. The estimated differences persist over several years and remain quantitatively stable over time as startups add more employees, although the baseline rate of female and minority employees increases slightly as firms age.

We again assess the robustness of the core results. First, we provide results weighting by firm size (employment) to make sure that our results are not driven by smaller or slower-growing firms. We show results for two weighting schemes: i) square root of employees; and ii) total employees. Tables A7 and A8 presents the results. We find slightly smaller quantitative differences across the specifications relative to Table 3. For example, the estimates in year 0 are 0.138, 0.107, and 0.130 for female, Black, and Hispanic in the regressions with linear weighting. The estimates with linear weighting do increase as firms age, suggesting bigger differences emerging across growing firms with different founder backgrounds.

Second, we attempt to replicate the industry, geography, and cohort fixed effects within the Crunchbase sample. The challenge is that the Crunchbase data is not as detailed in terms of industry and location, so we use higher-level industry groups (48) and geography variables

Table 3: Entrepreneur-Worker Homophily (Balanced Sample, Crunchbase)

Panel A: Female

Year relative to founding:	0	1	2	3	4	5
Female Founder Fraction	0.174***	0.179***	0.183***	0.181***	0.179***	0.177***
	(0.00491)	(0.00419)	(0.00394)	(0.00388)	(0.00390)	(0.00393)
Constant	0.300***	0.307***	0.316***	0.325***	0.332***	0.337***
	(0.00168)	(0.00138)	(0.00128)	(0.00127)	(0.00128)	(0.00131)
N	44545	44545	44545	44545	44545	44545

Panel B: Black

Year relative to founding:	0	1	2	3	4	5
Black Founder Fraction	0.104***	0.109***	0.109***	0.111***	0.110***	0.112***
	(0.00585)	(0.00535)	(0.00491)	(0.00493)	(0.00499)	(0.00515)
Constant	0.0773***	0.0768***	0.0774***	0.0779***	0.0789***	0.0797***
	(0.000655)	(0.000545)	(0.000502)	(0.000495)	(0.000503)	(0.000521)
N	44545	44545	44545	44545	44545	44545

Panel C: Hispanic

Year relative to founding:	0	1	2	3	4	5
Hispanic Founder Fraction	0.243***	0.241***	0.238***	0.241***	0.244***	0.244***
	(0.00806)	(0.00730)	(0.00705)	(0.00705)	(0.00710)	(0.00724)
Constant	0.0735***	0.0735***	0.0745***	0.0760***	0.0771***	0.0787***
	(0.000766)	(0.000637)	(0.000596)	(0.000594)	(0.000603)	(0.000623)
N	44545	44545	44545	44545	44545	44545

Notes: Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

(state). The specifications we use include: i) industry-by-founding year FEs; ii) industry-by-state-by-founding year FEs. Figure A2 compares the baseline estimates to the two sets of fixed-effects estimates. The estimates again become very slightly smaller in magnitude once we include more detailed controls (0.144, 0.103, and 0.240 in year 0), but the point estimates remain quantitatively large (and statistically significant). We include the most detailed fixed effects for the remaining analysis.

Finally, we also provide longer-run results using an unbalanced sample of firms, corresponding to the unbalanced LEHD sample discussed above. We include the most detailed version of the fixed effects. Table A9 shows quantitatively similar estimates to those in the core sample (0.149, 0.085, 0.216 in year 0 for female, Black, and Hispanic, respectively). The effects again persist strongly over time, and the estimates for year 9 are very similar to the estimates for year 0. The main change over time are the baseline rates, which increase slightly for all three groups (e.g., 0.303 to 0.355 for the constant term), so the relative effect sizes become smaller over time. Another way to view this is that organizations do not diverge over time in their composition.

Overall, the estimates are quantitatively smaller than those derived from the LEHD sample, most likely due to measurement error. As noted above, the Crunchbase sample is likely skewed towards technology-based startups with high growth potential. However, as we show below, homophily estimates remain similar when focusing on the top ten percent of firms by employment growth in the first six years. A major issue with the Crunchbase demographic data is that it is name-based. This may lead to misclassification of gender and race for founders and attenuate our estimates. One potential issue on the LEHD side is that we do not directly measure founders. In the "one founder" robustness check described above, we recover estimates (0.162, 0.309, and 0.194) that are closer to those from the Crunchbase sample, although the estimate for Black homophily is much larger.

III Heterogeneity and Additional Factors

In this section, we assess whether our core estimates remain large within important subgroups of new firms and employees. We also assess additional factors such as production structure and the demographics of the initial (non-founder) workers.

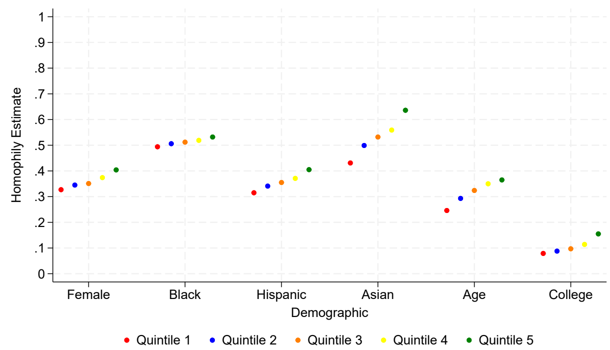


Figure 2: Homophily by Worker Wage Quintile

Notes: estimates of Equation 1 using the balanced LEHD sample and reporting results separately for each demographic and wage quintile (1 is lowest paid, 5 is highest). All standard errors are negligible relative to the point estimates, so we have not added confidence intervals to the plot. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

III.A Heterogeneity By Income Quintile, Occupation, Seniority

Here, we test whether the results hold within important subgroups of workers. We start by presenting results by income quintile from the LEHD, and then move to seniority and occupation in the LinkedIn data. The last two variables are unavailable in the Census LEHD data.

First, we sort employees into quintiles within each firm's pay distribution in any year, and compute average demographics within each quintile for each firm and firm age cell. This then allows us to estimate homophily coefficients by quintile. Figure 2 reports the results for each demographic and worker wage quintile, with 1 being the lowest paid quintile of workers and 5 being the highest. We find large effects across the board, but, for each demographic, we find larger point estimates when analyzing the highest paid workers.

Switching to the Crunchbase sample, we present results related to seniority. Based on role name, Revelio provides an estimate of the seniority of a given role on a scale from 1 (entry-level) to 7 (C-suite). We split the sample of workers into senior (seniority greater than or equal to 4) and junior, and run separate regressions based on seniority. In these regressions, we pool the six years of each firm and cluster standard errors at the firm level. We find similar point estimates across the two groups for each demographic, but female, Black, and Hispanic workers are a smaller fraction of the pool of senior employees, which

Table 4: Entrepreneur-Worker Homophily by Seniority (Crunchbase)

Panel A: Female

	Senior	Junior
Female Founder Fraction	0.142***	0.140***
	(0.00424)	(0.00393)
Constant	0.285***	0.385***
	(0.00124)	(0.00122)
N	407430	407025

Panel B: Black

	Senior	Junior
Black Founder Fraction	0.0805***	0.0785***
	(0.00500)	(0.00535)
Constant	0.0755***	0.0854***
	(0.000492)	(0.000508)
N	412125	411932

Panel C: Hispanic

	Senior	Junior
Hispanic Founder Fraction	0.216***	0.222***
	(0.00744)	(0.00736)
Constant	0.0688***	0.0887***
	(0.000613)	(0.000643)
N	412125	411932

Notes: estimates of Equation 1 using the balanced Crunchbase sample and pooling across years 0–6 and including industry-by-state-by-cohort fixed effects. Standard errors in parentheses and are clustered at the firm level $^*p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

suggests bigger relative effects at more senior levels.

Finally, we present results, splitting by broad occupation categories based on Revelio's processing of raw job titles. The seven broad occupation categories are: administrator, engineer, finance, marketing, operations, sales, and scientist. Table 5 presents the results. We find large effects across all occupation categories for gender and race, with particularly large effects relative to baseline rates within engineers.

III.B Heterogeneity by Sector

Here, we use industry information to assess heterogeneity by sector. We start with the Census sample and present results for service versus non-service industries.[8] Potentially, the nature of production or management might generate differences across these two groups.

[8] We split the sample based on two-digit NAICS codes: service (51 and above); non-service (49 and below).

Table 5: Entrepreneur-Worker Homophily by Occupation (Crunchbase)

Panel A: Female

	Admin	Engineer	Finance	Marketing	Operations	Sales	Scientist
Female Founder Fraction	0.0806***	0.0771***	0.0813***	0.123***	0.120***	0.140***	0.134***
	(0.00638)	(0.00433)	(0.00765)	(0.00496)	(0.00666)	(0.00506)	(0.00906)
Constant	0.567***	0.205***	0.426***	0.479***	0.298***	0.338***	0.424***
	(0.00209)	(0.00118)	(0.00228)	(0.00165)	(0.00202)	(0.00151)	(0.00291)
N	227352	323327	192435	305620	248262	346563	134058

Panel B: Black

	Admin	Engineer	Finance	Marketing	Operations	Sales	Scientist
Black Founder Fraction	0.0861***	0.0831***	0.0723***	0.0753***	0.0867***	0.0681***	0.0688***
	(0.00851)	(0.00675)	(0.00962)	(0.00641)	(0.00847)	(0.00584)	(0.0112)
Constant	0.0918***	0.0714***	0.0830***	0.0853***	0.0809***	0.0839***	0.0886***
	(0.000870)	(0.000602)	(0.000918)	(0.000633)	(0.000858)	(0.000595)	(0.00112)
N	237805	332012	200471	313814	258641	353947	142167

Panel C: Hispanic

	Admin	Engineer	Finance	Marketing	Operations	Sales	Scientist
Hispanic Founder Fraction	0.224***	0.243***	0.221***	0.224***	0.200***	0.196***	0.182***
	(0.0112)	(0.00915)	(0.0122)	(0.00867)	(0.0109)	(0.00786)	(0.0137)
Constant	0.0892***	0.0795***	0.0793***	0.0842***	0.0780***	0.0776***	0.0798***
	(0.00101)	(0.000783)	(0.00107)	(0.000783)	(0.000999)	(0.000701)	(0.00125)
N	237805	332012	200471	313814	258641	353947	142167

Notes: estimates of Equation 1 using the balanced Crunchbase sample and pooling across years 0–6 and including industry-by-state-by-cohort fixed effects. Standard errors in parentheses and are clustered at the firm level $*p < 0.05$, $** \ p < 0.01$, $*** \ p < 0.001$.

Table 6: Entrepreneur-Worker Homophily (Service versus Non-Service)

Panel A: Service

Employee Fraction:	Female	Black	Hispanic	Asian	Age	College
Female Founder Fraction	0.377***					
	(0.002)					
Black Founder Fraction		0.462***				
		(0.004)				
Hispanic Founder Fraction			0.378***			
			(0.002)			
Asian Founder Fraction				0.527***		
				(0.005)		
Average Founder Age					0.266***	
					(0.002)	
College Grad Founder Fraction						0.098***
						(0.001)
Constant	0.208***	0.041***	0.107***	0.019***	22.45***	0.157***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.074)	(0.000)
Observations	1553000	1553000	1553000	1553000	1553000	1553000

Panel B: Non-Service

Employee Fraction:	Female	Black	Hispanic	Asian	Age	College
Female Founder Fraction	0.360***					
	(0.001)					
Black Founder Fraction		0.519***				
		(0.002)				
Hispanic Founder Fraction			0.370***			
			(0.001)			
Asian Founder Fraction				0.568***		
				(0.001)		
Average Founder Age					0.336***	
					(0.001)	
College Grad Founder Fraction						0.114***
						(0.001)
Constant	0.394***	0.049***	0.089***	0.024***	19.17***	0.206***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.040)	(0.000)
Observations	5669000	5669000	5669000	5669000	5669000	5669000

Notes: Regressions pooling all six years within the balanced LEHD sample. All specifications include Industry-MSA-founding year fixed effects. Standard errors in parentheses $^*p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6 presents the results. We find very similar coefficients across the two groups of firms along each demographic dimension.

Next, we use high-level industry groups from Crunchbase to assess heterogeneity across industry groups. We find similar qualitative patterns across four large industry groups (software, hardware, manufacturing, and health care), although the estimates for manufacturing are smaller and imprecisely estimated for two of the demographics (female and Black).

Table 7: Entrepreneur-Worker Homophily by Industry (Crunchbase)

Panel A: Female

	Software	Hardware	Manufacturing	Health Care
Female Founder Fraction	0.0932***	0.111***	0.0615	0.135***
	(0.00888)	(0.0242)	(0.0423)	(0.00991)
Constant	0.296***	0.261***	0.267***	0.408***
	(0.00236)	(0.00587)	(0.00946)	(0.00361)
N	81133	11932	4638	45130

Panel B: Black

	Software	Hardware	Manufacturing	Health Care
Black Founder Fraction	0.0928***	0.123**	0.0512	0.0839***
	(0.0112)	(0.0386)	(0.0377)	(0.0156)
Constant	0.0704***	0.0679***	0.0771***	0.0836***
	(0.000972)	(0.00265)	(0.00370)	(0.00139)
N	81271	11948	4643	45185

Panel C: Hispanic

	Software	Hardware	Manufacturing	Health Care
Hispanic Founder Fraction	0.263***	0.265***	0.229**	0.138***
	(0.0162)	(0.0503)	(0.0806)	(0.0190)
Constant	0.0710***	0.0768***	0.0718***	0.0786***
	(0.00129)	(0.00343)	(0.00553)	(0.00150)
N	81271	11948	4643	45185

Notes: estimates of Equation 1 using the balanced Crunchbase sample and pooling across years 0–6 and including industry-by-state-by-cohort fixed effects. Standard errors in parentheses and are clustered at the firm level *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 8: Entrepreneur-Worker Homophily (Split by Employment Growth)

Panel A: Top Decile Growth

Employee Fraction:	Female	Black	Hispanic	Asian	Age	College
Female Founder Fraction	0.372***					
	(0.003)					
Black Founder Fraction		0.519***				
		(0.005)				
Hispanic Founder Fraction			0.379***			
			(0.004)			
Asian Founder Fraction				0.562***		
				(0.005)		
Average Founder Age					0.299***	
					(0.003)	
College Grad Founder Fraction						0.145***
						(0.003)
Constant	0.332***	0.051***	0.094***	0.022***	21.32***	0.200***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.121)	(0.001)
Observations	469000	469000	469000	469000	469000	469000

Panel B: Remaining Firms

Employee Fraction:	Female	Black	Hispanic	Asian	Age	College
Female Founder Fraction	0.366***					
	(0.001)					
Black Founder Fraction		0.522***				
		(0.002)				
Hispanic Founder Fraction			0.381***			
			(0.001)			
Asian Founder Fraction				0.576***		
				(0.002)		
Average Founder Age					0.333***	
					(0.001)	
College Grad Founder Fraction						0.112***
						(0.001)
Constant	0.354***	0.047***	0.090***	0.022***	20.17***	0.193***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.038)	(0.000)
Observations	4478000	4478000	4478000	4478000	4478000	4478000

Notes: Regressions pooling all six years within the balanced LEHD sample. All specifications include Industry-MSA-founding year fixed effects. Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

III.C Growth and Size

Next, we assess whether faster growing firms or larger firms exhibit different patterns to those found in the main sample. For example, they may have more formalized hiring procedures or be less targeted in their hiring.

In the LEHD sample, we cut firms by realized employment growth over the first six years in the data. We separately estimate homophily within the fastest-growing ten percent of firms versus the remaining ninety percent and report the results in Table 8. We find very similar point estimates across the two groups.

Table 9: Entrepreneur-Worker Homophily by Firm Employment Growth (Crunchbase)

Panel A: Female

	Top Decile	Other
Female Founder Fraction	0.130***	0.144***
	(0.0106)	(0.00386)
Constant	0.353***	0.335***
	(0.00238)	(0.00116)
N	45424	404191

Panel B: Black

	Top Decile	Other
Black Founder Fraction	0.101***	0.0826***
	(0.0267)	(0.00490)
Constant	0.0793***	0.0813***
	(0.00171)	(0.000471)
N	45425	404727

Panel C: Hispanic

	Top Decile	Other
Hispanic Founder Fraction	0.179***	0.223***
	(0.0210)	(0.00715)
Constant	0.0812***	0.0794***
	(0.00148)	(0.000591)
N	45425	404727

Notes: estimates of Equation 1 using the balanced Crunchbase sample and pooling across years 0–6 and including industry-by-state-by-cohort fixed effects. Standard errors in parentheses and are clustered at the firm level $*p < 0.05$, $**\ p < 0.01$, $***\ p < 0.001$.

In the Crunchbase sample, we can also cut the firms by growth rates. As noted above, the Crunchbase sample is already selected for growth potential, but our large sample allows for the same top ten percent cut. Table 9 reports the results.

In addition, we also present results by sorting firm-year data points into size groups and report homophily coefficients by size groups. As noted earlier, firms may formalize HR procedures once they are large enough, in which case we would expect to see smaller homophily coefficients once a firm becomes large enough. Table 10 presents the results by firm size quintile (5 being the largest set of firm-year data points). We find slightly smaller coefficients for the top quintile (within industry-MSA-cohort differences restricted to the largest firms) versus the other four but the point estimates remain quantitatively large.

Table 10: Entrepreneur-Worker Homophily by Size (Crunchbase)

Panel A: Female

	Q1	Q2	Q3	Q4	Q5
Female Founder Fraction	0.150***	0.163***	0.152***	0.136***	0.110***
	(0.0104)	(0.00610)	(0.00504)	(0.00461)	(0.00521)
Constant	0.292***	0.315***	0.336***	0.356***	0.389***
	(0.00295)	(0.00176)	(0.00142)	(0.00125)	(0.00135)
N	94433	91629	82180	88180	93193

Panel B: Black

	Q1	Q2	Q3	Q4	Q5
Female Founder Fraction	0.0964***	0.0805***	0.0881***	0.0597***	0.0838***
	(0.0115)	(0.00689)	(0.00733)	(0.00570)	(0.0119)
Constant	0.0813***	0.0805***	0.0800***	0.0811***	0.0834***
	(0.00125)	(0.000680)	(0.000593)	(0.000502)	(0.000802)
N	94957	91642	82180	88180	93193

Panel C: Hispanic

	Q1	Q2	Q3	Q4	Q5
Female Founder Fraction	0.197***	0.210***	0.240***	0.244***	0.199***
	(0.0140)	(0.00957)	(0.0106)	(0.0110)	(0.0136)
Constant	0.0750***	0.0772***	0.0791***	0.0804***	0.0869***
	(0.00139)	(0.000855)	(0.000812)	(0.000795)	(0.000931)
N	94957	91642	82180	88180	93193

Notes: estimates of Equation 1 using the balanced Crunchbase sample and pooling across years 0–6 and including industry-by-state-by-cohort fixed effects. Q5 represents the set of largest firm-years. Standard errors in parentheses and are clustered at the firm level: *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

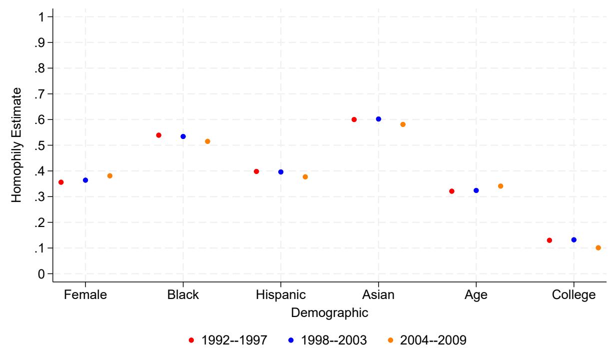


Figure 3: Homophily by Founding Year Cohorts

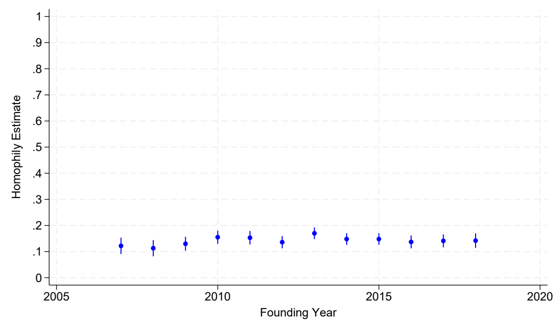
Notes: estimates of Equation 1 using the balanced LEHD sample and reporting results separately for each founding cohort group (1992–1997, 1998–2003, 2004–2009). All standard errors are negligible relative to the point estimates, so we have not added confidence intervals to the plot. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

III.D Trend Across Cohorts

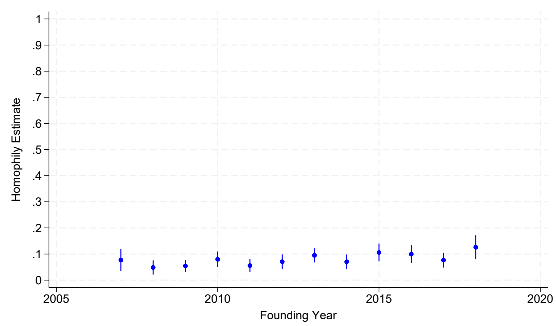
So far, we have pooled our estimates across firms founded in different years. Potentially, homophily could be shifting over time due to changes in labor markets, production structure, and social factors. To assess this, we estimate homophily coefficients within different founding year groups for both the LEHD and Crunchbase sample.

Figure 3 presents results from the LEHD sample. We break the firms into three founding group cohorts (1992–1997, 1998–2003, 2004–2009) so that the final cohort still has six years of data available, and report estimates within each group. We find minimal quantitative differences across the three groups, with some coefficients increasing over time (female and age) and some decreasing (Black, Hispanic, and college-educated). The changes are not quantitatively large.

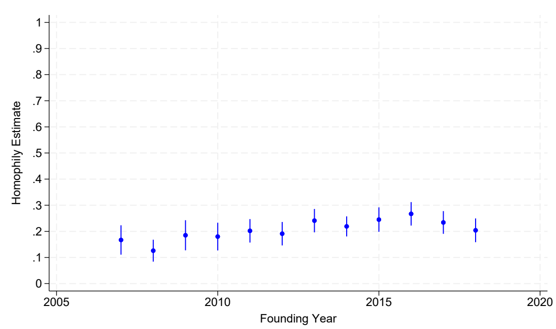
We also report estimates by detailed cohorts in the Crunchbase sample. For this sample, we are not restricted in terms of volume of output, so we report estimates by founding year. Figure 4 reports the estimates. We again find fairly stable estimates for all three demographic variables, with a slight upwards trend rather than the opposite.



(a) Female



(b) Black



(c) Hispanic

Figure 4: Estimates by Founding Year (Crunchbase)

Notes: estimates of Equation 1 using the balanced Crunchbase sample. Standard errors are clustered at the firm level and the figure reports 95 percent confidence intervals.

III.E Assessing Additional Factors Correlated with Founder Characteristics

Finally, we conclude by considering whether factors correlated with founder demographics might account for some of the observed homophily estimates.

First, we use Census data to measure the production structure of new firms. We test whether founders from different backgrounds create firms with different average wages, labor share, and trading behavior. To do so, we focus on the 2007 and 2012 Economic Census, the last two during our sample period, and compute statistics for each new firm in the first Economic Census they are surveyed, restricted to firms within 5 years of founding. We again compare within detailed industry-MSA-cohort groups and find significant differences in terms of average wage: female, Black, and Hispanic founded firms pay significantly lower average wages ($15k, $11k, and $6.5k, respectively), whereas college-educated founders pay more ($19.8k). We find smaller differences in terms of labor share and trading intensity, although some significant differences are present. Appendix Tables A3, A4, A5, and A6 provide detailed estimates.

We then control for these characteristics and re-run the homophily analysis. Table A10 presents the estimates. Because the sample is different to the main balanced sample (restricted to startups founded between 2002 and 2012 and to firms with wage and labor share information), we report specifications with and without controls. We find very minimal differences in homophily when controlling and not controlling for production structure characteristics, other than for college-founded startups. This suggests that the set of production structure measures available to us are not driving the observed homophily coefficients.

Second, we analyze whether founder demographics or initial employee demographics have a larger long-run effect on employee demographics. For example, it is possible that organizations evolve based more on initial employees because they are the ones in charge of future hiring. We focus on the Crunchbase sample because we can reliably separate founders and employees. Table 11 reports the estimates from a horserace regression incorporating both founder characteristics and initial employee characteristics in the founding year. We find that initial worker demographics is strongly predictive of subsequent worker demographics, which is unsurprising given that some initial employees will remain at the firm, but the effect

Table 11: Entrepreneur-Worker Homophily Horserace (Balanced Sample, Crunchbase)

Panel A: Female

Year relative to founding:	1	2	3	4	5
Female Founder Fraction	0.0608***	0.0878***	0.0948***	0.0983***	0.0970***
	(0.00321)	(0.00357)	(0.00374)	(0.00392)	(0.00401)
Initial Female Employee Fraction	0.602***	0.441***	0.378***	0.344***	0.332***
	(0.00410)	(0.00431)	(0.00440)	(0.00453)	(0.00461)
N	45360	45360	45360	45360	45360

Panel B: Black

Year relative to founding:	1	2	3	4	5
Black Founder Fraction	0.0366***	0.0503***	0.0557***	0.0582***	0.0599***
	(0.00372)	(0.00425)	(0.00447)	(0.00479)	(0.00497)
Initial Black Employee Fraction	0.594***	0.428***	0.360***	0.334***	0.328***
	(0.00861)	(0.00872)	(0.00847)	(0.00853)	(0.00891)
N	45360	45360	45360	45360	45360

Panel C: Hispanic

Year relative to founding:	1	2	3	4	5
Hispanic Founder Fraction	0.0813***	0.106***	0.121***	0.130***	0.132***
	(0.00461)	(0.00524)	(0.00570)	(0.00591)	(0.00617)
Initial Hispanic Employee Fraction	0.634***	0.490***	0.434***	0.407***	0.398***
	(0.00764)	(0.00796)	(0.00810)	(0.00819)	(0.00839)
N	45360	45360	45360	45360	45360

Notes: Standard errors in parentheses $^{*}p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

fades quickly as firms age. In contrast, the founder coefficient strengthens significantly as firms age, suggesting that the founder has greater influence in the long run.

IV Wage Analysis

Given our core results, we turn to assessing whether the differences in relative quantity are driven by labor supply or labor demand factors.

To do so, we make use of wage data from the LEHD and employ a two-step approach. First, we compute AKM regressions across all employees and firms. Because of computational limitations, we do so state-by-state. The difference from a standard approach is that we compute firm and individual fixed effects for individuals from each demographic group (and the complementary group of individuals). Formally, we run the following regression for each state s:

$$\log(Wage_{i(ds)t}) = \alpha_i + \gamma_{jsd} + \epsilon_{it}$$

where γ_{jsd} is firm j's fixed effect for a particular demographic group d based on data from state s, α_i is the individual's fixed effect, and the outcome is log wages. This allows us to collect a set of estimates at the firm-demographic-state level. As noted above, for each demographic group d, we also compute the firm fixed effect for the complementary set of workers (d').

For the second step, we compute the differential in firm FEs for each demographic group d (versus the complement d') and regress the difference against founder fraction from that group. Unlike the first step, we now focus on new firms (same sample of firms as the unbalanced LEHD sample). Formally, we run the following regression:

$$\gamma_{jsd} - \gamma_{jsd'} = \beta_0 + \beta_1 \times FounderFraction_{jd} + \nu_{jds} \tag{2}$$

where d is a specific demographic group. We cluster standard errors at the firm level to account for firms having establishments in multiple states.

Table 12 presents the estimates from Equation 2. Across demographic groups, we find sizable positive coefficients range from 9 percent to 19.7 percent, with the largest effects for female workers. This means that workers who share a demographic with the founders earn wages that are higher than expected relative to their pay at other firms, relative to the same differential for individuals who do not share a demographic. More broadly, this suggests that labor demand effects dominate labor supply effects, in a relative sense. The coefficients would be negative if workers prefer working for founders from a similar background. The positive coefficients could be the result of: i) limited ability of workers from the preferred backgrounds and/or rent sharing; ii) non-pecuniary value to the entrepreneur from managing people from similar backgrounds.

Our analysis has a few limitations. First, the LEHD data do not contain information about non-wage compensation (e.g., options). However, there would need to be large within-firm across-demographic differences in non-wage compensation to offset the observed estimates. Second, some firms do not hire any individuals from certain backgrounds, so we

Table 12: AKM Decomposition Regressions

	Firm Fixed-Effect Differential				
Female Founder Fraction	0.197*** (0.002)				
Black Founder Fraction		0.157*** (0.004)			
Hispanic Founder Fraction			0.101*** (0.002)		
Asian Founder Fraction				0.091*** (0.003)	
College Grad Founder Fraction					0.093*** (0.003)
Observations	1,794,000	798,000	1,822,000	1,116,000	1,125,000

Notes: Standard errors in parentheses and are clustered at the firm level. *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

cannot include them in the corresponding analysis. Finally, due to computational constraints, we run the model in a state-by-state manner. Therefore, the estimates will not reflect across-state movers, thereby reducing their precision.

V Model and Counterfactuals

V.A Framework

Finally, we use our empirical estimates to calibrate an equilibrium model that allows for the production function of a firm to vary based on entrepreneur characteristics. Our goal is to provide estimates of the implications of unequal access to entrepreneurship for inequality across workers with different socio-economic characteristics. The starting point is a simple CES production function framework that focuses on heterogeneity in labor demand, which helps map observed equilibrium labor quantities to wage ratios.

We consider a firm with a CES production function over male and female workers (this framework would also apply to other groups, e.g. by race):

$$Y = \left((A_W L_W)^{\frac{\sigma-1}{\sigma}} + (A_M L_M)^{\frac{\sigma-1}{\sigma}}\right)^{\sigma/(\sigma-1)}$$

The firm solves the profit-maximization problem

$$\text{Max}\, Y - w_W L_W - w_M L_M$$

The FOCs yield:

$$\frac{w_M}{w_W} = \left(\frac{A_M}{A_W}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{L_M}{L_W}\right)^{-1/\sigma}$$

Using this equation, we can run a simple counterfactual, asking what would happen to equilibrium wages if the economy experienced a change in the "tastes" for male vs. female workers. This change in parameters is a simple way to model broader access to entrepreneurship, which can have implications for wage inequality through employer-employee homophily.

To get a sense for order of magnitudes, we compare the ratio of male-to-female workers at firms founded by males and females:

$$\Delta\log\left(\frac{L_W}{L_M}\right) \equiv \log\left(\frac{L_W}{L_M}\right)_{\text{FemaleStartup}} - \log\left(\frac{L_W}{L_M}\right)_{\text{MaleStartup}}$$

Then, assuming that labor supply is fixed, using the equations above we get the corresponding change in equilibrium wages:

$$\Delta\log\left(\frac{w_W}{w_M}\right) = \frac{1}{\sigma} \cdot \Delta\log\left(\frac{L_W}{L_M}\right) \tag{3}$$

where σ is the elasticity of substitution between male and female workers. As an illustrative example, with $\sigma = 2$, if there are 20 percent more women in female-founded startups, we expect to end up with a 10 percent higher wage for women if we went from an equilibrium with only male startups to an equilibrium with only female startups.

We then consider less extreme counterfactuals, e.g., going to an equilibrium with equal representation of male and female startups (relative to the worker pool at startups), compared to the current equilibrium with unequal access. In this case we still use equation but we input a different expression on the RHS:

$$\Delta\log\left(\frac{L_W}{L_M}\right) \equiv \log\left(s'_M\left(\frac{L_W}{L_M}\right)_{\text{MaleStartup}} + s'_W\left(\frac{L_W}{L_M}\right)_{\text{FemaleStartup}}\right) - \log\left(s_M\left(\frac{L_W}{L_M}\right)_{\text{MaleStartup}} + s_W\left(\frac{L_W}{L_M}\right)_{\text{Fem}}\right. \tag{4}$$

where s_M and s_W denote the shares of female and male startups observed today, with $s_M + s_W = 1$. This expression contains quantities we can estimate using the data and, when combined with estimates of the elasticity of substitution, can be used to estimate changes in the wage ratio using Equation 3.

Table 13 presents calculations using our summary stats and estimates. We project sizable increases in relative wage for women (2.77%), Blacks (4.36%), and Hispanics (2.26%). The estimate for non-college workers (0.90%) is smaller because the difference across founders

Table 13: Back-of-the-envelope calculations, $\sigma = 2$

	Founder (s_W)	Employee (s'_W)	Intercept	Slope	$\frac{L_W}{L_M}$ (group)	$\frac{L_W}{L_M}$(other)	ΔRelative
Female	0.408	0.501	0.353	0.364	2.53	0.55	2.77%
Black	0.054	0.075	0.047	0.512	1.27	0.05	4.36%
Hispanic	0.105	0.131	0.092	0.373	0.87	0.10	2.26%
Non-college	0.719	0.774	0.805	0.113	11.20	4.13	0.90%

Notes: back-of-the-envelope calculations using Equations 3 and 4 and plugging in estimates from the summary stats table (Table 1) and the pooled LEHD balanced sample regression (Table A1, Panel C).

is smaller relative to the baseline rate. Of course, these estimates will vary if we focus on high-growth startups, where women appear to have greater access barriers.

In ongoing work, we are exploring several extensions of this initial framework. First, we want to estimate elasticities of substitution using our sample by leveraging shocks at the MSA level. Second, we want to incorporate the relative wage estimates from Section IV into the model. Finally, we want to incorporate entrepreneurial ability distributions to assess gains from greater exposure (Einio et al., 2023; Chiplunkar et al., 2024). Our current projections are all in relative terms, and, under the assumption that exposure is partly driving access differences, increases in exposure for productive entrepreneurs would increase wages for all workers.

VI Conclusion

As noted above, our primary contribution to policy is to document the role played by entrepreneurship in creating labor market opportunities for different groups of individuals. In particular, the combination of our empirical results and quantitative framework can help evaluate the gains from increased participation in entrepreneurship by individuals from underrepresented groups.

There are several policies and initiatives that aim to increase access to entrepreneurship. The Small Business Innovation Research program already aim to promote entrepreneurship from women, underrepresented minorities, and individuals from "HUBZones," historically underutilized business zones. Other policy initiatives include the SUCCESS Act of 2018 that directed the USPTO to make recommendations for promoting the participation of women, minorities, and veterans in entrepreneurship and innovation. In addition, venture capital funds have recently increased their hiring of black and female venture capital partners, which research has shown to increase funding for female and black entrepreneurs (Calder-Wang et al., 2021; Cook et al., 2022). The evidence presented adds to our understanding of the distributional effects of policies aimed at broadening access to innovation and entrepreneurship careers. For example, facilitating careers in entrepreneurship for women might provide an important avenue to reduce the gender wage gap.

References

Abowd, John M., Francis Kramarz, and David N. Margolis, "High wage workers and high wage firms," *Econometrica*, 1999, *67*, 251–333.

Bell, Alexander, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," 2017.

Bento, Pedro, "Female entrepreneurship in the U.S. 1982â2012: Implications for welfare and aggregate output," *Journal of Monetary Economics*, 1 2024.

_ **and Sunju Hwang**, "Barriers to black entrepreneurship: Implications for welfare and aggregate output over time," *Journal of Monetary Economics*, 3 2023, *134*, 16–34.

Bessen, James, Felix Poege, Ronja RÃ¶ttger, and (Tpri, "Competing for Talent: Large Firms and Startup Growth," 2023.

Bryan, Kevin A, Mitchell Hoffman, and Amir Sariri, "INFORMATION FRICTIONS AND EMPLOYEE SORTING BETWEEN STARTUPS," 2022.

Calder-Wang, Sophie, Paul Gompers, and Patrick Sweeney, "Venture Capital's "Me Too" Moment," 2021.

Card, David, Ana Rute Cardoso, and Patrick Kline, "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women," *Quarterly Journal of Economics*, 5 2016, *131*, 633–686.

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter, "Race and economic opportunity in the United States: An intergenerational perspective," 5 2020.

Chiplunkar, Gaurav, Pinelopi Koujianou Goldberg, Oriana Bandiera, Latika Chaudhary, Rafael Dix-Carneiro, Erik Hurst, Asif Islam, Rachel Ngai, Rohini Pande, Michael Peters, Imran Rasul, and Gabriel Ulyssea, "Aggregate Implications Of Barriers To Female Entrepreneurship *," *Econometrica*, 2024.

Choi, Joonkyu, Nathan Goldschlag, John Haltiwanger, and J Daniel Kim, "Early Joiners and Startup Performance," *The Review of Economics and Statistics*, 2023.

Cook, Lisa D, Matt Marx, and Emmanuel Yimfor, "FUNDING BLACK HIGH-GROWTH STARTUPS," 2022.

Einio, Elias, Josh Feng, and Xavier Jaravel, "Social Push and the Direction of Innovation," 2023.

Goldin, Claudia and Cecilia Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 2000, *90*, 715–741.

Gozen, Ruveyda, Richard Hornbeck, Anders Humlum, and Martin Rotemberg, "Historical Differences in Female-Owned Manufacturing Establishments: The United States, 1850â1880," 2024.

Guzman, Jorge and Aleksandra Kacperczyk, "Gender Gap in Entrepreneurship," 2018.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda, "WHO CREATES JOBS? SMALL VERSUS LARGE VERSUS YOUNG," *The Review of Economics and Statistics*, 2013, *95*, 347–361.

Kerr, Sari Pekkala and William R Kerr, *Immigrant Entrepreneurship*

Kim, J. Daniel, "Is there a startup wage premium? Evidence from MIT graduates," *Research Policy*, 4 2018, *47*, 637–649.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson, "Who do we invent for? Patents by women focus more on women's health, but few women get to invent," *Science*, 2021, *372*, 1345–1348.

Marinoni, Astrid and John Voorheis, "Who Gains from Creative Destruction? Evidence from High-Quality Entrepreneurship in the United States," 2019.

Roach, Michael and Henry Sauermann, "Can Technology Startups Hire Talented Early Employees? Ability, Preferences, and Employee First Job Choice," *Management Science*, 8 2023.

Ronchi, Maddalena and Nina Smith, "Daddy's Girl: Daughters, Managerial Decisions, and Gender Inequality," 2024.

Tate, Geoffrey and Liu Yang, "Female leadership and gender equity: Evidence from plant closure," *Journal of Financial Economics*, 7 2015, *117*, 77–97.