Who Hires Whom? Entrepreneurial Backgrounds and Labor Market Opportunities*

Josh Feng, *University of Utah – David Eccles School of Business*
Xavier Jaravel, *London School of Economics* and *CEPR*

April 2026

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 several demographic dimensions. 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, we present evidence that these differences persist out to at least 20 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 model of entrepreneurship with heterogeneous ability and production functions using our estimates, and assess the impacts on relative wage from reducing access barriers to entrepreneurship.

*We thank Naomi Hausman, Ben Jones, Tim Simcoe, and Elena Simintzi for providing detailed feedback. We thank participants at the Banff Innovation, Productivity, and Growth Conference, the NBER Productivity Lunch, the NBER Summer Institute, and the Wasatch Front RDC seminar for thoughtful comments. 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 (disclosure numbers CBDRB-FY25-P2229-R11972 and CBDRB-FY26-P2229-R12962). 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., 2019; Guzman and Kacperczyk, 2018; Cook et al., 2022; Bento and Hwang, 2023; Bento, 2024; Glaeser and Hausman, 2020). 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 these underrepresented 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 that incorporates entrepreneurs creating firms with heterogeneous production functions in terms of labor inputs in order to assess the implications of our estimates for

[1]The closest work to ours is Chiplunkar and Goldberg (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.

labor market opportunities.

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).[2] 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 2.9 percent to 12.9 percent across demographic variables. This suggests that the relative demand dominates the relative supply in terms of generating the observed equilibrium quantities. In

[2]To focus on hiring, we exclude founders from this analysis.

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.[3] 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., 2019; 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.[4] 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.[5]

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; Sorenson et al., 2021; 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; Balsmeier et al., 2025). Our work aims to add additional evidence to this line of work by highlighting the workers that gain from new firm creation using administrative data.

Finally, we will also address the literature on labor market opportunities and outcomes.

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

[4]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/`.

[5]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.

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 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 (Campero and Kacperczyk, 2020; Chiplunkar and Goldberg, 2024; Gozen et al., 2024; Chen et al., 2025; Ashraf et al., 2025). Beyond gender, Miller and Schmutte (2021) study the mechanisms behind racial co-hiring using administrative data from Brazil, Lerner et al. (2025) studies how impact investors influence the wages of minority employees, and Hwang and Wallskog (2026) analyze the impact of entrepreneurship on the labor market opportunities of "marginalized workers" and find that successful young firms do have a positive impact on the careers of these workers. 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 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.[6] We use this as the core sample throughout.[7]

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 log points in terms of 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.[8] 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.

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

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

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

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 include the demographics of the founders (based on the top three earner in the firm's first year) and 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).

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.[9]

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, focusing on firms founded after 2007, roughly the time Crunchbase starts tracking startups. We also construct an unbalanced panel of all Crunchbase firms, removing the founding year filter. Table 2 reports summary statistics corresponding to those in Table 1 for the balanced sample. We again find that founder shares are smaller than average employee 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.

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

[9]Hyatt et al. (2021) show that the standard approach for identifying entrepreneurs in the LEHD data is unreliable. The approach does not work for sole proprietors and partnerships. In addition, for S-corporations, selecting the top three earners does reliably cover business owners, but a majority of these individuals are not business owners.

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		44,545	

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		267,270	

Notes: Panel A presents statistics at the new firm level for the balanced sample of firms founded in or after 2007 that have non-founder employees with LinkedIn profiles in each of the years 0 to 5 after founding. These include the demographics of the founders (as identified from raw LinkedIn job titles) and 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 merged to LinkedIn (Revelio Labs).

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 (e.g., six-digit NAICS) and two-digit industry-by-MSA level and interact them with founding year (cohort) to facilitate the comparison of similar firms.

We note two issues surrounding the scale of β. First, the estimate is mechanically constrained to range from -1 to 1. Second, it is not constrained by the underlying population shares, as long as some individuals from each group are present in the data.

Using the basic "homophily" regression 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

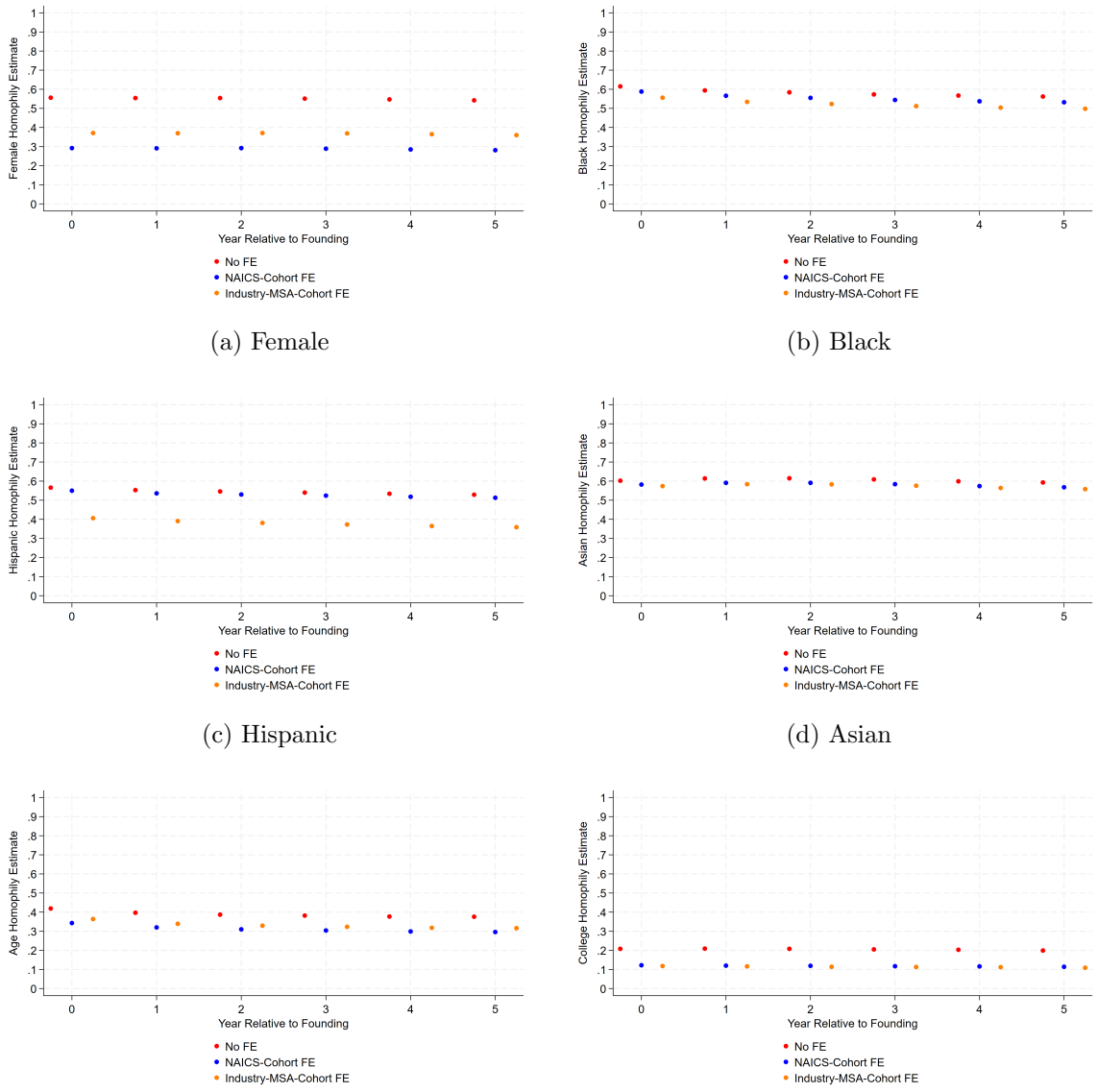
II.A Core Results – LEHD

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 at firms founded by the complementary group (i.e., the regression intercept). 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 test the robustness of our estimates to how many top paid employees we count as founders and report the results in Table A2. The estimates become smaller when 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.

Second, 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 A3 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).



(a) Female

(b) Black

(c) Hispanic

(d) Asian

(e) Age

(f) College Educated

Figure 1: Estimates from Balanced Sample of New Firms

Notes: Each panel plots the coefficient $\hat{\beta}_t$ from Equation 1 using a different demogrpahic variable. The three series correspond to specifications with: (i) no fixed effects; ii) 6-digit NAICS-by-founding year fixed effects; iii) 2-digit industry-by-MSA-by-founding year fixed effects. All standard errors are clustered at the firm level and 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 numbers: CBDRB-FY25-P2229-R11972 and CBDRB-FY26-P2229-R12962).

Third, we also report results using zip code instead of MSA in the very detailed fixed effects specification. Zip code level analysis helps account for more granular labor markets, but comes at the cost of less reliable estimates when comparing within industry by geography by founding year cohorts. Table A1 shows that the estimates become slightly smaller across all demographic characteristics, but remain quantitatively large.

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 A5 and A6 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 (state). The specifications we use include: i) industry-by-founding year FEs; ii) industry-by-state-by-founding year FEs. Figure A7 compares the baseline estimates to the two sets of fixed-effects estimates. The estimates again become very slightly smaller in magnitude once

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	44,545	44,545	44,545	44,545	44,545	44,545

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	44,545	44,545	44,545	44,545	44,545	44,545

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	44,545	44,545	44,545	44,545	44,545	44,545

Notes: Each column reports the coefficient from a separate regression of employee demographic share on founder demographic share for the indicated year relative to founding (Equation 1), estimated using the balanced Crunchbase sample. No fixed effects are included. Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: Crunchbase merged to LinkedIn (Revelio Labs).

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.

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.

II.C Persistence of the associations

To provide more evidence on long-run associations, we make use of the unbalanced panels in each setting.

First, we re-run our analysis on an unbalanced panel of new firms from the LEHD and report the results from the most detailed fixed effect specification in Figure A4. We find similar initial effects (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 also provide longer-run results using an unbalanced sample of firms from the LinkedIn-Crunchbase match. We include the most detailed version of the fixed effects. Figure A5 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 across all three

demographics. Even out to twenty years, the estimates are still quantitatively large (0.135, 0.062, 0.174). The main change over time are the baseline rates, which increase slightly for all three groups (e.g., 0.303 to 0.389 for the constant term in the female regressions), so the relative effect sizes become smaller over time.

In addition, we also use the LinkedIn data to provide a comparison between startups and publicly traded companies, as another way to proxy for long-run effects.[10] To do so, we collect data on active C-suite employees for each publicly traded firm and year between 2010 and 2019. For the 2010–2019 period, we then run employee homophily regressions using founder demographics for the startup sample (restricting to years 0–5 of a startup) and C-suite demographics for publicly-traded companies. We incorporate year fixed effects to account for changes in composition. Table A4 provides the estimates from this comparison. We find similar quantitative estimates across the two settings: larger homophily estimates for gender and smaller for race in the startup sample versus the public company sample. We also find that the publicly traded firms have a larger share of female and minority employees, consistent with the intercepts in the long-run estimates discussed above.

III Heterogeneity and Additional Results

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.

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

[10] We thank our discussant Elena Simintzi for this suggestion.

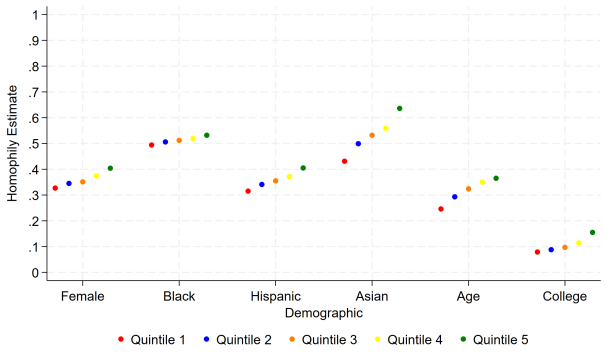


Figure 2: Homophily by Worker Wage Quintile

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). Workers are sorted into quintiles based on within-firm pay distribution in each year, and average demographics are computed within each quintile for each firm-year cell. 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).

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 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.

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	407,430	407,025

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	412,125	411,932

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	412,125	411,932

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 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

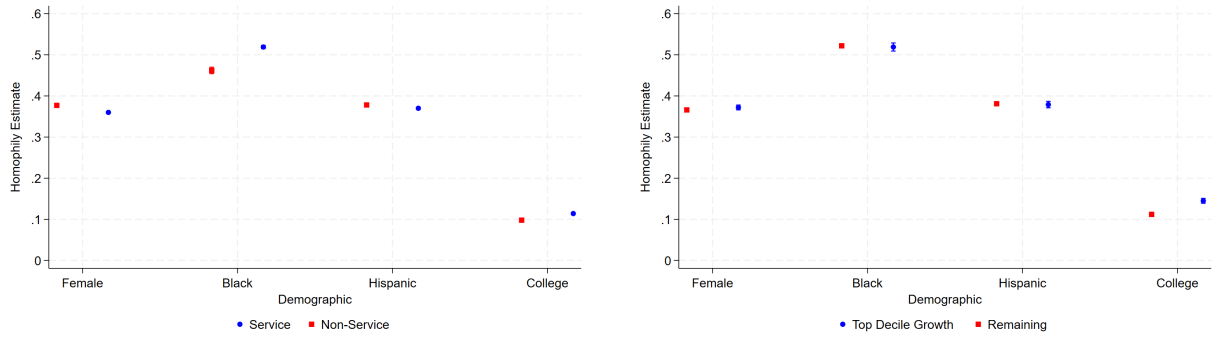
Panel B: Black

	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 *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.



(a) Service versus Non-Service

(b) Top Decile Growth versus Remaining

Figure 3: Homophily by Firm Characteristics (LEHD)

Notes: Each panel plots pooled homophily coefficients from the balanced LEHD sample (Equation 1), separately for subsamples defined by firm characteristics. Panel A splits firms into service (2-digit NAICS $\geq$ 51) and non-service ($\leq$ 49) industries. Panel B splits firms by employment growth from year 0 to year 5 (top decile versus remaining 90 percent). All specifications include industry-MSA-founding year fixed effects. Bars show 95% confidence intervals. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, DRB approval numbers: CBDRB-FY25-P2229-R11972 and CBDRB-FY26-P2229-R12962).

III.B Heterogeneity by Sector and Growth

Here, we use industry and growth information to assess heterogeneity by firm characteristics. We start with the Census sample and present results for service versus non-service industries.[11] Potentially, the nature of production or management might generate differences across these two groups. We also cut firms by realized employment growth over the first six years in the data, separately estimating homophily within the fastest-growing ten percent of firms versus the remaining ninety percent.

Figure 3 summarizes the results. Panel A shows very similar coefficients across service and non-service firms along each demographic dimension. Panel B shows that faster growing firms exhibit nearly identical homophily estimates as the remaining firms.

We also use high-level industry groups from the Crunchbase sample to assess heterogeneity across industry groups, finding similar qualitative patterns across four large industry groups (software, hardware, manufacturing, and health care); see Appendix Table A7.

In addition, we also estimate the core homophily regressions (with MSA-by-cohort fixed effects) for each six-digit NAICS industry. Table 6 reports the distribution across all industries with any variation and for the top quartile of industries in terms of new firms. The

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

Table 6: Distribution of Homophily Estimates at the Industry Level

Panel A – All Industries

Percentile	Homophily Coefficient					
	Female	Black	Hispanic	Asian	Age	College
5	-0.2173	-0.1285	-0.2085	-0.02077	-0.4312	-0.2179
10	-0.04522	-0.02846	-0.0146	0.05343	-0.09821	-0.09125
25	0.1197	0.1895	0.2113	0.2836	0.1554	0.02385
50	0.2489	0.3725	0.3435	0.4926	0.2879	0.09186
75	0.3795	0.5019	0.4591	0.6697	0.4144	0.1921
90	0.5328	0.6409	0.6461	0.8196	0.624	0.3288
95	0.7178	0.7694	0.8454	0.9538	0.9075	0.4969

Panel B – Top Quartile (Number of New Firms)

Percentile	Homophily Coefficient					
	Female	Black	Hispanic	Asian	Age	College
5	0.06349	0.1399	0.1431	0.07074	0.1191	-0.008247
10	0.09168	0.2145	0.1845	0.1721	0.1575	0.01526
25	0.1496	0.3159	0.2659	0.3397	0.223	0.04129
50	0.241	0.4145	0.3418	0.4773	0.291	0.07855
75	0.3385	0.5005	0.4042	0.5935	0.3658	0.1285
90	0.4109	0.5665	0.4816	0.6965	0.4196	0.1915
95	0.4549	0.6407	0.5357	0.7443	0.4579	0.2274

Notes: Distribution of homophily estimates using the balanced LEHD sample, with coefficients estimated industry-by-industry (NAICS six-digit). All specifications include Industry-MSA-founding year fixed effects. Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, DRB approval number: CBDRB-FY26-P2229-R12962).

latter sample likely produces more reliable estimates, especially with the very detailed fixed effects.

We confirm these patterns in the Crunchbase sample. Appendix Table A8 shows similar estimates when splitting the Crunchbase sample by employment growth, and Appendix Table A9 shows slightly smaller coefficients for the largest quintile of firms by size but estimates that remain quantitatively large. The evidence suggests that homophily is not just a small firm phenomenon and persists even as firms grow and hiring processes become formalized.

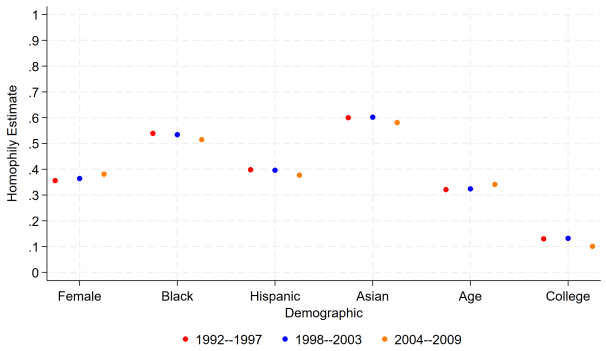


Figure 4: Homophily by Founding Year Cohorts

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.C 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 4 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 founding year cohorts in the Crunchbase sample, which confirm fairly stable estimates across cohorts with a slight upwards trend (see Appendix Figure A8).

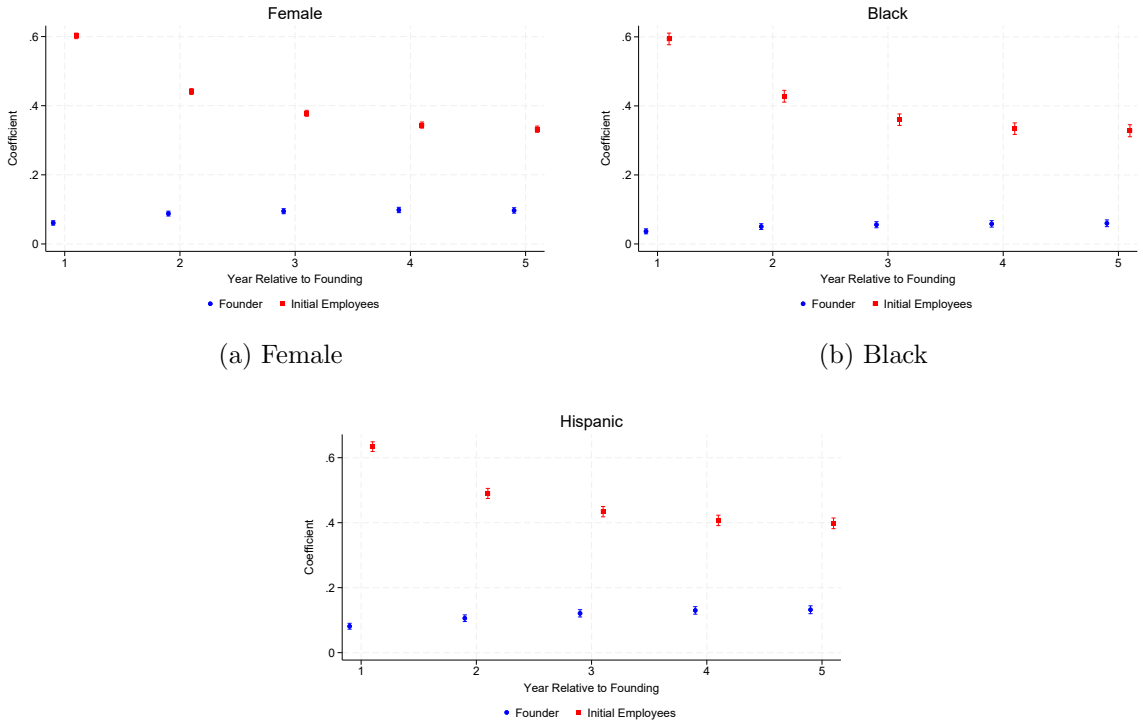
III.D Assessing Additional Factors Correlated with Founder Characteristics

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 A10, A11, A12, and A13 provide detailed estimates.

We then control for these characteristics and re-run the homophily analysis. Table A14 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. Figure 5 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



(a) Female

(b) Black

(c) Hispanic

Figure 5: Founder versus Initial Employee Demographics Horserace (Crunchbase)

Notes: Each panel plots coefficients from a regression of employee demographic share on both founder demographic share and initial (year 0) non-founder employee demographic share, estimated separately for years 1 through 5 relative to founding. Year 0 is excluded as the dependent variable because it mechanically overlaps with the initial employee share regressor. All specifications include industry-by-state-by-founding year fixed effects. Standard errors are clustered at the firm level; shaded areas show 95% confidence intervals. Source: Crunchbase merged to LinkedIn (Revelio Labs).

effect 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. We also report corresponding estimates from the Census sample in Appendix B, which show similar qualitative patterns.

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 estimate regressions in the style of Abowd et al. (1999) (hereafter, AKM) across all employees and firms, except founders at

new 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), in the style of Card et al. (2016). Formally, we run the following regression for each state s:

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

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'). The additional controls include age, age squared, and tenure at the current firm.

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 7 presents the estimates from Equation 2. Across demographic groups, we find sizable positive coefficients range from 2.9 percent to 12.9 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) rent sharing; ii) non-pecuniary value to the entrepreneur from managing people from similar backgrounds.[12]

[12]Our results stand in contrast to work by Campero and Kacperczyk (2020), who find that labor supply

Table 7: AKM Decomposition Regressions

	Firm Fixed-Effect Differential				
Female Founder Fraction	0.129***				
	(0.002)				
Black Founder Fraction		0.068***			
		(0.004)			
Hispanic Founder Fraction			0.057***		
			(0.002)		
Asian Founder Fraction				0.029***	
				(0.003)	
College Grad Founder Fraction					0.049***
					(0.005)
Constant	0.062***	0.117***	-0.316***	0.128***	-0.140***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Observations	1,775,000	791,000	1,802,000	1,104,000	584,000

Notes: The dependent variable is the difference in firm fixed effects between the indicated demographic group and its complement $\hat{\gamma}_{jsd} - \hat{\gamma}_{jsd'}$. Firm fixed effects are estimated separately by state and demographic group following Equation 2, controlling for age, age squared, and firm tenure. 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-FY26-P2229-R12962).

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 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.

Appendix B.J presents additional evidence supporting the validity of the AKM-by-group approach. First, we show estimates using firm effects estimated using a wage regression without any controls. Second, we follow Bonhomme et al. (2019) and Lerner et al. (2024) in allowing for non-linear match effects. We report statistics related to model fit, averaged across all states (Table A17, Panel A). The approach also sorts firms into ten bins, and ends up estimating an assortative matching parameter for each bin. We find statistically significant but quantitatively small differences in slope between new firms and the remaining firms in the sample (Table A17, Panel B). Third, we also validate the firm-by-demographics

effects dominate, using job application data from high-tech startups. The discrepancy could be due to differences in sample or due to the fact that only applications and interviews are observed.

estimates, by showing descriptive wage trends in the style of Card et al. (2013). Figure A9 presents the results for movers to quintile five (highest firm effect) firms, breaking movers down by their pre-move firm's quintile. We find minimal pre-trends, and larger jumps for individuals moving from lower quintile firms.[13]

V A Model of Entrepreneurs with Heterogeneous Production Functions

Our evidence suggests that labor demand effects drive the observed employee composition differences and that founder effects persist over time. Using this as motivation, we build an equilibrium model where entrepreneurs create firms with heterogeneous production functions, calibrate the model using the summary stats and estimates presented above, and run counterfactuals to assess the impacts of reducing access barriers to entrepreneurship.

V.A Model

Our model follows the general structure of Chiplunkar and Goldberg (2024), hereafter CG, but simplifies away some aspects that are more important in a developing economy context (e.g., informal businesses).

The economy is populated by two demographic groups, majority (m) and minority (f), with population shares N_m and N_f. Each individual draws ability z from a log-normal distribution and chooses whether to become an entrepreneur or supply labor at the market wage.[14]

Production. An entrepreneur of type $g \in m, f$ with ability z produces output using a composite labor bundle:

$$y_g(z) = z \cdot \ell_g(z)^\rho, \quad \rho \in (0, 1)$$

The composite labor bundle aggregates workers of both types via a CES technology.

[13]Disclosure constraints prevent us from showing the full set of pairwise moves.

[14]This follows the structure of the Chiplunkar and Goldberg (2024) model, but removes self-employment and non-participation as choices for individuals.

$$\ell_g = \left[A_{gm}^{1/\gamma} \, \ell_{gm}^{\frac{\gamma-1}{\gamma}} + A_{gf}^{1/\gamma} \, \ell_{gf}^{\frac{\gamma-1}{\gamma}} \right]^{\frac{\gamma}{\gamma-1}}$$

where ℓ_{gh} is the quantity of type-h workers hired, $\gamma > 1$ is the elasticity of substitution between worker types, and the parameters $A_{gh} \geq 0$ govern each founder type's relative productivity from different worker types. This specification makes explicit that entrepreneurs of different backgrounds g create different production functions: the weights $\{A_{gm}, A_{gf}\}$ vary across founder types, so majority and minority founders value the two worker types differently. Higher values of A_{gg} relative to A_{gh} ($h \neq g$) capture in-group hiring bias (homophily): founders are able to elicit more productivity from workers from similar backgrounds (e.g., better management, different processes, etc.).

The key distinction between our model and CG's model is related to the qualitative predictions on potential rent sharing. CG generates homophily via different hiring costs by entrepreneur and employee backgrounds, while keeping the production function fixed across entrepreneurs. Female entrepreneurs end up hiring more female workers because the hiring costs are lower, but productivity is the same conditional on hiring. Under the CG scenario, entrepreneurs would likely share more rents with opposite group workers, because they are equally valuable in production and harder to replace. Under our model, we would expect rent sharing to be higher for own-group workers, because they are relatively more productive when working for entrepreneurs from similar backgrounds.

Entry into entrepreneurship. Entry is governed by two conditions. First, the individual must have access to entrepreneurial opportunities: a fraction $\phi_g \in (0, 1]$ of each group is exposed, where $\phi_m = 1$ (the majority group is normalized to face no access frictions). Second, conditional on access, entry must be profitable: variable profit must exceed the outside option,

$$(1 - \rho), z, \ell_g(z)^\rho \geq w_g$$

This implies an endogenous ability cutoff z_g^* proportional to ω_g. Individuals with $z < z_g^*$ or without access supply labor at the market wage.

Equilibrium. A competitive equilibrium is a pair of wages (w_m, w_f) such that labor markets for both worker types clear simultaneously:

$$\sum_{g \in m,f} N_g \int_{z_g^*}^{\infty} \ell_{gh}(z) \cdot \phi_g \cdot dF(z) = N_h \left(1 - \phi_h \left[1 - F(z_h^*)\right]\right), \quad h \in m, f$$

where the left-hand side is aggregate demand for type-h workers from all active entrepreneurs and the right-hand side is the labor supply of type-h workers (population minus entrepreneurs). The optimal demands $\ell_{gh}(z)$ and the cutoffs z_g^* are derived from cost minimization and free entry, as detailed in Appendix C.

V.B Calibration and Counterfactuals

For our calibration, we impose a row-sum normalization on the preference matrix A, setting $A_{fm} = A_{mm} + A_{mf} - A_{ff}$. This aims to ensure that founders of similar ability z have similar productivity multipliers and costs per implied unit of labor. With $A_{ff} = 1$ as a normalization, the model has three free parameters: A_{mm}, A_{mf}, and ϕ_f. These are then calibrated to four empirical moments: the minority worker share at minority-founded firms, the minority worker share at majority-founded firms, the minority founder share, and the minority-to-majority wage ratio. The system is over-identified by one moment.

We calibrate the model to match observed moments, and then compute counterfactual equilibrium outcomes when setting $\phi = 1$. We follow Chiplunkar and Goldberg (2024) and set $\rho = 0.75$ (decreasing returns to effective labor) and $\gamma = 2$ (elasticity of substitution between majority and minority workers). Population shares N_m, N_f are taken from Table 1, Panel B. For each demographic group, we target four moments: the minority worker share at minority-founded firms, the minority worker share at majority-founded firms, the minority founder share, and the minority-to-majority wage ratio.[15] Despite the over-identified setup, the model generally has good fit.

The main estimates also help us contextualize the AKM estimates from Table 7. Under a hypothetical scenario where a worker is currently working at a firm run by an entrepreneur from the same background and her outside option is joining a firm founded by an opposite background entrepreneur, the additional relative surplus will be $log(A_{ff}) - log(A_{fm}) -$

[15]For example, in the female calibration, the target moments are: female share at female-founded firms (0.717), female share at male-founded firms (0.353), female founder share (0.408), and the raw female-to-male wage ratio (0.80, from the U.S. Bureau of Labor Statistics). See Appendix Table A18 for the full set of moments.

Table 8: Model Calibration and Counterfactuals

	Calibration						Counterfactuals	
Group	A_{mm}	A_{mf}	A_{fm}	A_{ff}	ϕ	w_f/w_m	w'_f/w'_m	ΔRelative Wages
Female	1.16	0.41	0.57	1.00	0.52	0.80	0.89	8.95pp
Black	2.37	0.06	1.44	1.00	0.47	0.75	0.83	7.90pp
Hispanic	2.83	0.15	1.98	1.00	0.53	0.73	0.80	6.97pp
Non-college	0.90	0.74	0.63	1.00	0.47	0.60	0.63	3.22pp

Notes: Each row calibrates the model separately for the indicated group (f) versus all other individuals (m). A_{gh} are the CES productivity weights governing demand composition across worker types. ϕ is the share of the minority group with access to entrepreneurship ($\phi_m = 1$ by normalization). w_f/w_m is the calibrated equilibrium wage ratio. Counterfactual columns report the equilibrium when $\phi_f = 1$ (equal access). Parameters: $\rho = 0.75$, $\gamma = 2$, $A_{ff} = 1$ (normalization), $A_{fm} = A_{mm} + A_{mf} - A_{ff}$ (row-sum normalization).

$log(A_{mf}) + log(A_{mm})$, which is equal to about 1.60 log points for female. The corresponding AKM estimate is 0.129, suggesting a rent sharing rate of 8.1 percent. The corresponding calculations for the other groups range from 2 percent ot 8 percent. These are generally in line with the estimates in Card et al. (2016).

For our main counterfactuals, we remove access barriers (set $\phi = 1$) and recompute equilibrium outcomes for each demographic group comparison. We find significant relative wage effects, ranging from 3.22pp for non-college individuals to 8.95pp for women. Of course, this counterfactual does not incorporate dynamics. If new firms today were formed in a world where $\phi = 1$, it would still take a significant period for these firms to have significant labor market impacts, because they need to replace enough of the existing pool of firms.

VI Conclusion

In this article, we have provide systematic evidence of large and persistent differences in the employee composition across similar new firms with different founder backgrounds. In addition, we have provided evidence that founders pay employees more than expected, suggesting that labor demand dominates labor supply in driving the relative equilibrium employment patterns. To capture these patterns, we build a stylized model of entrepreneurship where founders create heterogeneous production functions, and use it to assess the implications of removing access barriers to entrepreneurship for specific demographic groups. We find large and significant effects.

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.

Ashraf, Nava, Alexia Delfino, Edward L. Glaeser, and Irene Solmone, "The Inverted U-Shaped Relationship between Female Entrepreneurship and Economic Development," *AEA Papers and Proceedings*, 5 2025, *115*, 500–508.

Balsmeier, Benjamin, Lee Fleming, Matt Marx, and Seungryul Ryan Shin, "Startups, Unicorns, and the Local Influx of Inventors," *Review of Economics and Statistics*, 2025.

Bell, Alexander, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *The Quarterly Journal of Economics*, 2019, *134*, 647–713.

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.

Bonhomme, StÃ©phane, Thibaut Lamadon, and Elena Manresa, "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 2019, *87*, 699–739.

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," Technical Report 2021.

Campero, Santiago and Aleksandra Kacperczyk, "Asymmetric Gender Homophily in the Startup Labor Market," *Advances in Strategic Management*, 2020, *41*, 329359.

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.

_ , JÃ¶rg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 2013, *128*.

Chen, Jun, Song Ma, and Feng Zhang, "Anti-Harassment Policy and the Startup Labor Market," Technical Report 2025.

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 and Pinelopi Koujianou Goldberg, "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," Technical Report 2023.

Glaeser, Edward L. and Naomi Hausman, *The Spatial Mismatch between Innovation and Joblessness*, UNIV OF CHICAGO Press, 2020.

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.

Hwang, Kylie and Melanie Wallskog, "The Role of Young Firms as Employers of Marginalized Workers," Technical Report 2026.

Hyatt, Henry, Seth Murray, and L. Kristin Sandusky, "Business Income Dynamics and Labor Market Fluidity," *IZA Journal of Labor Economics*, 1 2021, *10*.

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.

Lerner, Josh, Henry Manley, Carolyn Stein, and Heidi Williams, "The Wandering Scholars: Understanding the Heterogeneity of University Commercialization," Technical Report 2024.

_ , Markus Lithell, and Gordon M Phillips, "Impact Investing and Worker Outcomes," 2025.

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

Miller, Conrad and Ian M Schmutte, "The Dynamics of Racial Co-Hiring," 2021.

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.

Sorenson, Olav, Michael S. Dahl, Rodrigo Canales, and M. Diane Burton, "Do startup employees earn more in the long run?," *Organization Science*, 7 2021, *32*, 587–604.

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

For Online Publication

Appendix to "Who Hires Whom? Entrepreneurial Backgrounds and Labor Market Opportunities"

Josh Feng, *University of Utah – David Eccles School of Business*
Xavier Jaravel, *London School of Economics*
April 2026

A Data Construction

A.A LEHD

We follow the methodology in Kerr and Kerr (2016) to identify new firms and their founders in the LEHD data. Specifically, we use the Longitudinal Business Database (LBD) to identify firms in their first year of existence. We filter out companies that had earlier entries in the LBD, as these represent existing firms that appear as new in the LEHD only because of state coverage limitations.

For each new firm, we flag the top three earners in the company's first year in the LEHD data and classify them as founders. Next, we compute average demographics of non-founder employees in each year relative to the firm's founding. Demographic information comes from the Decennial Census files linked to the LEHD.

For our primary analysis, we construct a balanced panel of firms that have at least one non-founder employee with valid MSA information in each of years 0 through 5 relative to founding. This yields approximately 825,000 firms founded between 1992 and 2014. We also construct an unbalanced panel that relaxes the survival requirement, which we use to study longer-run dynamics.

A.B Crunchbase

Our second sample merges startup information from Crunchbase with employee work histories from LinkedIn, as compiled by Revelio Labs.

We focus on startups in Crunchbase that were founded in or after 2007, roughly the time Crunchbase began systematically tracking new firms. Our Crunchbase data comes from a December 2020 snapshot, while the LinkedIn data comes from Revelio Labs' April 2023 snapshot.

We use a crosswalk provided by Revelio Labs to link Crunchbase firm identifiers to Revelio company IDs (RCIDs). Founders are identified based on raw job titles in LinkedIn profiles (e.g., "Founder," "Co-Founder," "Founding CEO"). Non-founder employees working at each firm in a given year are identified based on the start and end dates reported in their LinkedIn work histories.

Unlike the LEHD, demographic variables are constructed from names rather than self-reported data. As discussed in the main text, name-based classification introduces measurement error that likely attenuates our estimates relative to the Census sample, where demographics are directly reported.

For our primary analysis, we construct a balanced sample of 44,545 firms that have non-founder employees with LinkedIn profiles in each of years 0 through 5 after founding. We also construct an unbalanced panel to study longer-run dynamics out to 20 years after founding. We remove the founding year filter for the unbalanced sample.

A.C LEHD Sample Sizes

Table A1 reports rounded firm counts for the primary LEHD samples and subsamples used in the heterogeneity analyses.

B Additional Results

This section presents additional results that complement the main analysis.

Table A1: LEHD Sample Sizes

Sample	Subsample	Firms
Balanced Sample		824,000
	Service	649,000
	Trade in first year	34,500
	High Growth	78,000
	Immigrant Founded	241,000
	Economic Census Sample	336,000
Unbalanced Sample		3,582,000

Notes: Rounded firm counts for LEHD samples and subsamples used in the analysis. The balanced sample requires firms to have at least one non-founder employee in each of years 0 through 5 relative to founding. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

B.A Robustness to Founder Definition and Weighting (LEHD)

We assess robustness to how we measure founders and to the importance we put on different firms.

Number of Founders. As noted in the main text, a limitation of the LEHD data is that we cannot measure founder directly. Our baseline specification identifies founders as the top three earners in a new firm's first year. Here, we assess sensitivity to this definition.

Table A2 reports estimates from the balanced LEHD sample under three alternative founder definitions: one, two, and three top earners. For parsimony, we pool all six years together in the reported estimates instead of reporting estimates by age and demographic. Panel A restricts founders to the single highest earner, Panel B uses the top two, and Panel C (our default) uses the top three.

The estimates increase monotonically with the number of individuals classified as founders. For example, the female homophily coefficient rises from 0.162 (one founder) to 0.277 (two founders) to 0.364 (three founders). One potential explanation is that with more individuals classified as founders, the founder demographic share captures a broader set of the firm's initial leadership, which strengthens the measured association with employee composition. Even under the most conservative one-founder definition, all estimates remain quantitatively large and highly significant.

Twice-Weighted Estimates. Table A3 reports estimates from specifications that apply

Table A2: Entrepreneur-Worker Homophily (Robustness to Founder Number)

Panel A: One Founder

Employee Fraction:	Female	Black	Hispanic	Asian	Age	College
Female Founder Fraction	0.162***					
	(0.001)					
Black Founder Fraction		0.309***				
		(0.002)				
Hispanic Founder Fraction			0.194***			
			(0.001)			
Asian Founder Fraction				0.363***		
				(0.001)		
Average Founder Age					0.156***	
					(0.001)	
College Grad Founder Fraction						0.040***
						(0.000)
Constant	0.446***	0.060***	0.114***	0.034***	25.98***	0.214***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.029)	(0.000)
Observations	7222000	7222000	7222000	7222000	7222000	7222000

Panel B: Two Founders

Employee Fraction:	Female	Black	Hispanic	Asian	Age	College
Female Founder Fraction	0.277***					
	(0.001)					
Black Founder Fraction		0.434***				
		(0.002)				
Hispanic Founder Fraction			0.298***			
			(0.001)			
Asian Founder Fraction				0.484***		
				(0.001)		
Average Founder Age					0.251***	
					(0.001)	
College Grad Founder Fraction						0.078***
						(0.001)
Constant	0.396***	0.053***	0.102***	0.027***	22.39***	0.204***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.034)	(0.000)
Observations	7222000	7222000	7222000	7222000	7222000	7222000

Panel C: Three Founders (Default)

Employee Fraction:	Female	Black	Hispanic	Asian	Age	College
Female Founder Fraction	0.364***					
	(0.001)					
Black Founder Fraction		0.512***				
		(0.002)				
Hispanic Founder Fraction			0.373***			
			(0.001)			
Asian Founder Fraction				0.565***		
				(0.001)		
Average Founder Age					0.324***	
					(0.001)	
College Grad Founder Fraction						0.113***
						(0.001)
Constant	0.353***	0.047***	0.092***	0.023***	19.76***	0.195***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.035)	(0.000)
Observations	7222000	7222000	7222000	7222000	7222000	7222000

Notes: Regressions pooling all six years within the balanced LEHD sample. All specifications include Industry-MSA-founding year fixed effects. Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

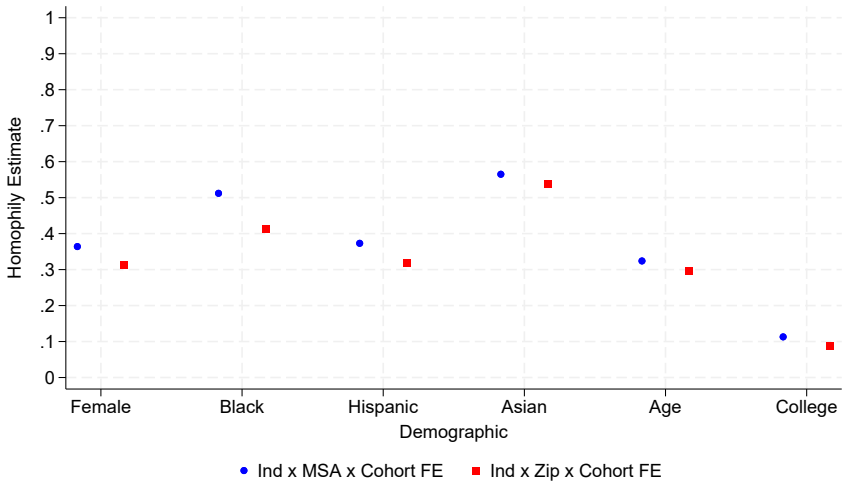


Figure A1: Comparison of Results at Different Levels of Geographic Controls

Notes: estimates of Equation 1 using the balanced LEHD sample. We compare estimates using two-digit industry by MSA by founding year fixed effects versus two-digit industry by zip code by founding year fixed effects . Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval numbers: CBDRB-FY25-P2229-R11972 and CBDRB-FY26-P2229-R1296).

two layers of weighting. First, we compute employee demographic shares using individual quarterly earnings as weights, so that higher-paid employees receive more weight. Second, we weight the regression by the square root of total employees, giving more influence to larger firms. The estimates are quantitatively very similar to the unweighted pooled baseline: 0.360 for female (versus 0.364), 0.544 for Black (versus 0.512), and 0.377 for Hispanic (versus 0.373). This confirms that our core findings are not driven by small or low-wage workers only.

B.B Geographic Control Level

In the core results, we compare startups within two-digit by MSA by founding year cells. MSA may be too coarse to control for the labor market conditions facing new firms. Figure A1 reports a comparison of the estimates using MSA versus zip code as the geographic cell. We find slightly smaller estimates across the demographic variables when analyzing within zip codes, but the estimates remain quantitatively large.

Table A3: Entrepreneur-Worker Homophily Twice-Weighted (Balanced Sample, LEHD)

Panel A: Female

Year relative to founding:	0	1	2	3	4	5
Female Founder Fraction	0.360***	0.359***	0.363***	0.360***	0.357***	0.353***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.352***	0.334***	0.327***	0.326***	0.326***	0.328***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	825000	825000	825000	825000	825000	825000

Panel B: Black

Year relative to founding:	0	1	2	3	4	5
Black Founder Fraction	0.544***	0.517***	0.506***	0.495***	0.486***	0.478***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.056***	0.054***	0.053***	0.053***	0.054***	0.054***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	825000	825000	825000	825000	825000	825000

Panel C: Hispanic

Year relative to founding:	0	1	2	3	4	5
Hispanic Founder Fraction	0.377***	0.361***	0.352***	0.344***	0.337***	0.331***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.094***	0.095***	0.095***	0.097***	0.099***	0.100***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	825000	825000	825000	825000	825000	825000

Panel D: College-Educated

Year relative to founding:	0	1	2	3	4	5
College-Educated	0.133***	0.136***	0.137***	0.137***	0.136***	0.134***
Founder Fraction	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.184***	0.194***	0.197***	0.199***	0.201***	0.203***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	825000	825000	825000	825000	825000	825000

Notes: All specifications include Industry-MSA-founding year fixed effects. Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

B.C Hiring versus Retention

Next, we present evidence comparing homophily estimates when we use all employees versus just new hires. Figure A2 shows that, for each demographic, the homophily coefficient for new hires decreases significantly more than for the overall homophily coefficient, albeit still remaining quantitatively large over time. This suggests that retention also plays a large role in generating the persistent homophily over time.

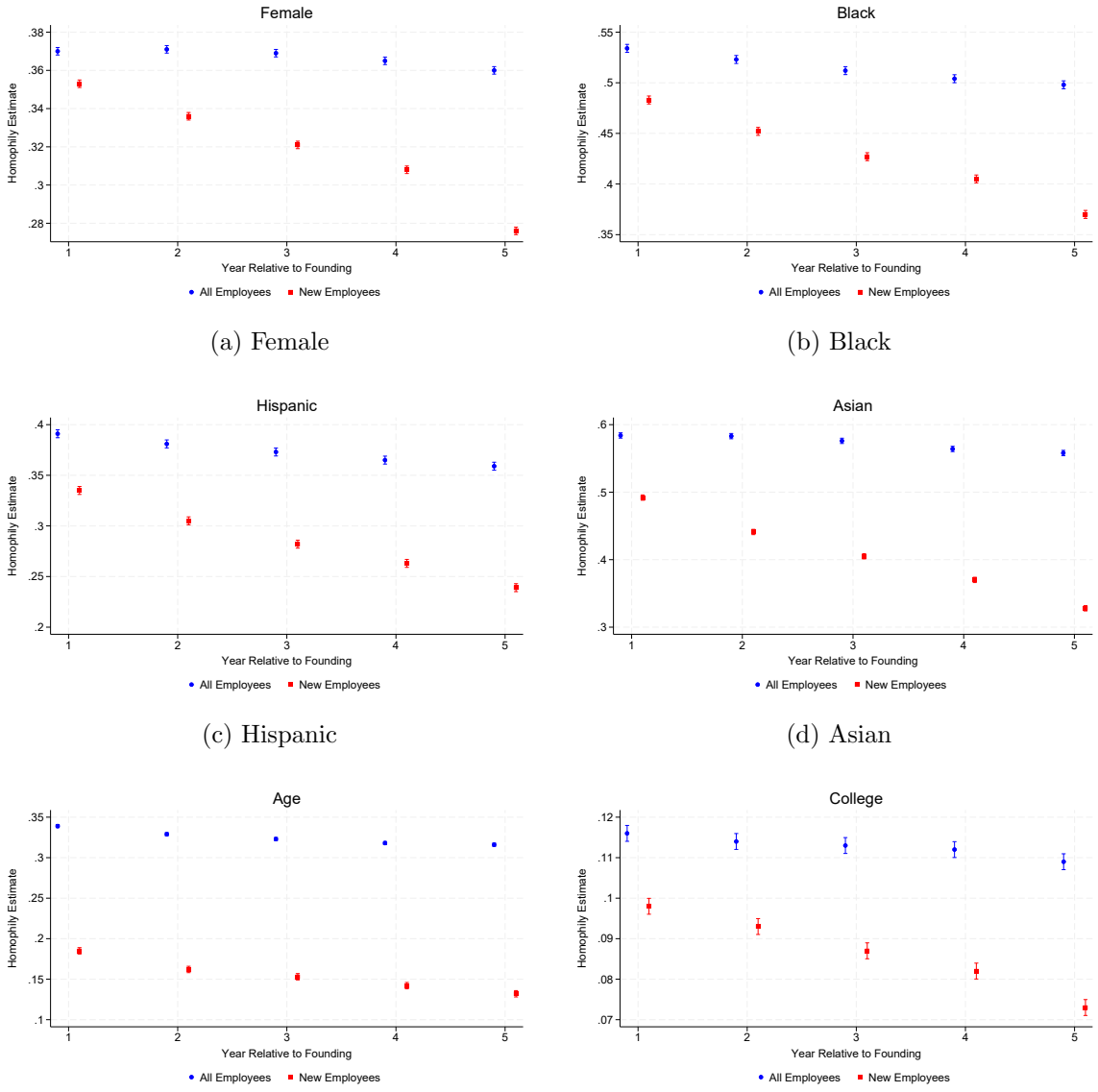
We also report the analogous comparison using the Crunchbase sample. Figure A3 shows the overall homophily estimates alongside new-hire-specific estimates across years relative to founding. The patterns are broadly consistent with the LEHD results: new-hire homophily declines more over time than overall homophily, reinforcing the role of retention in generating persistent compositional differences.

B.D Unbalanced Panel Estimates

The balanced panel ensures that our estimates are not confounded by differential firm exit, but it restricts the sample to relatively successful firms. Here, we relax the survival requirement and report estimates from unbalanced panels in both the LEHD and Crunchbase samples, which also allow us to track homophily over longer horizons.

LEHD Unbalanced Panel. Figure A4 reports estimates from the unbalanced LEHD sample, extending to 9 years after founding. The initial effects are very similar to the balanced panel (e.g., 0.381versus 0.364 for female in year 0). The effects attenuate gradually over time but remain quantitatively large even at year 9 (e.g., 0.250 for female, 0.433 for Black). Because the sample is unbalanced, the decay reflects both within-firm convergence and differential attrition across firm types.

Crunchbase Unbalanced Panel. Figure A5 reports estimates from the unbalanced Crunchbase sample, covering years 0 through 20. The LinkedIn data allow us to track homophily patterns over a substantially longer horizon than the LEHD. The initial estimates are very similar to the balanced sample (0.149 versus 0.174 for female, 0.085 versus 0.104 for Black, and 0.216 versus 0.243 for Hispanic in year 0). We also find that the effects persist out to 20 years after founding: the year 20 estimates are 0.135 for female, 0.062 for Black, and 0.174 for Hispanic.



(a) Female

(b) Black

(c) Hispanic

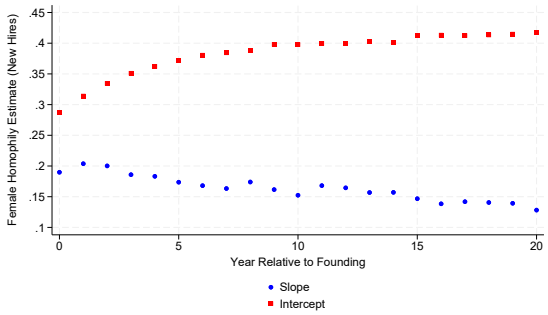
(d) Asian

(e) Age

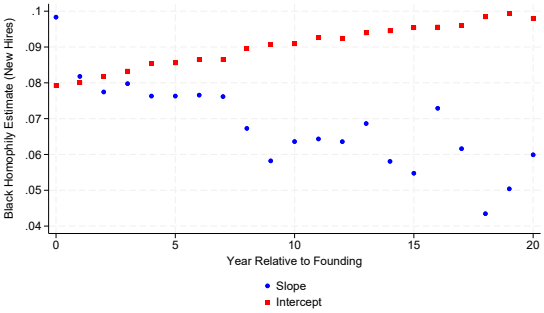
(f) College Educated

Figure A2: Comparison of Overall Homophily Estimates versus New Hire Homophily Estimates

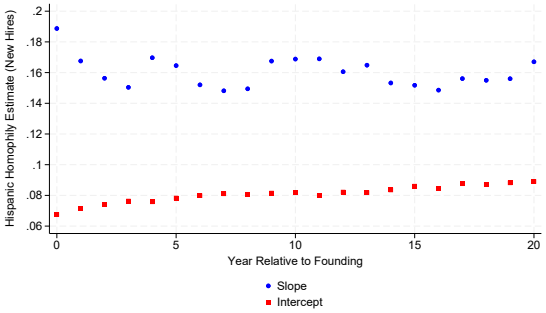
Notes: estimates of Equation 1 using the balanced LEHD sample. The two outcomes that are compared are overall demographic composition of workers and the demographic composition of new hires (the coefficients represent the difference across firms with different founder backgrounds). All standard errors 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 numbers: CBDRB-FY25-P2229-R11972 and CBDRB-FY26-P2229-R1296).



(a) Female



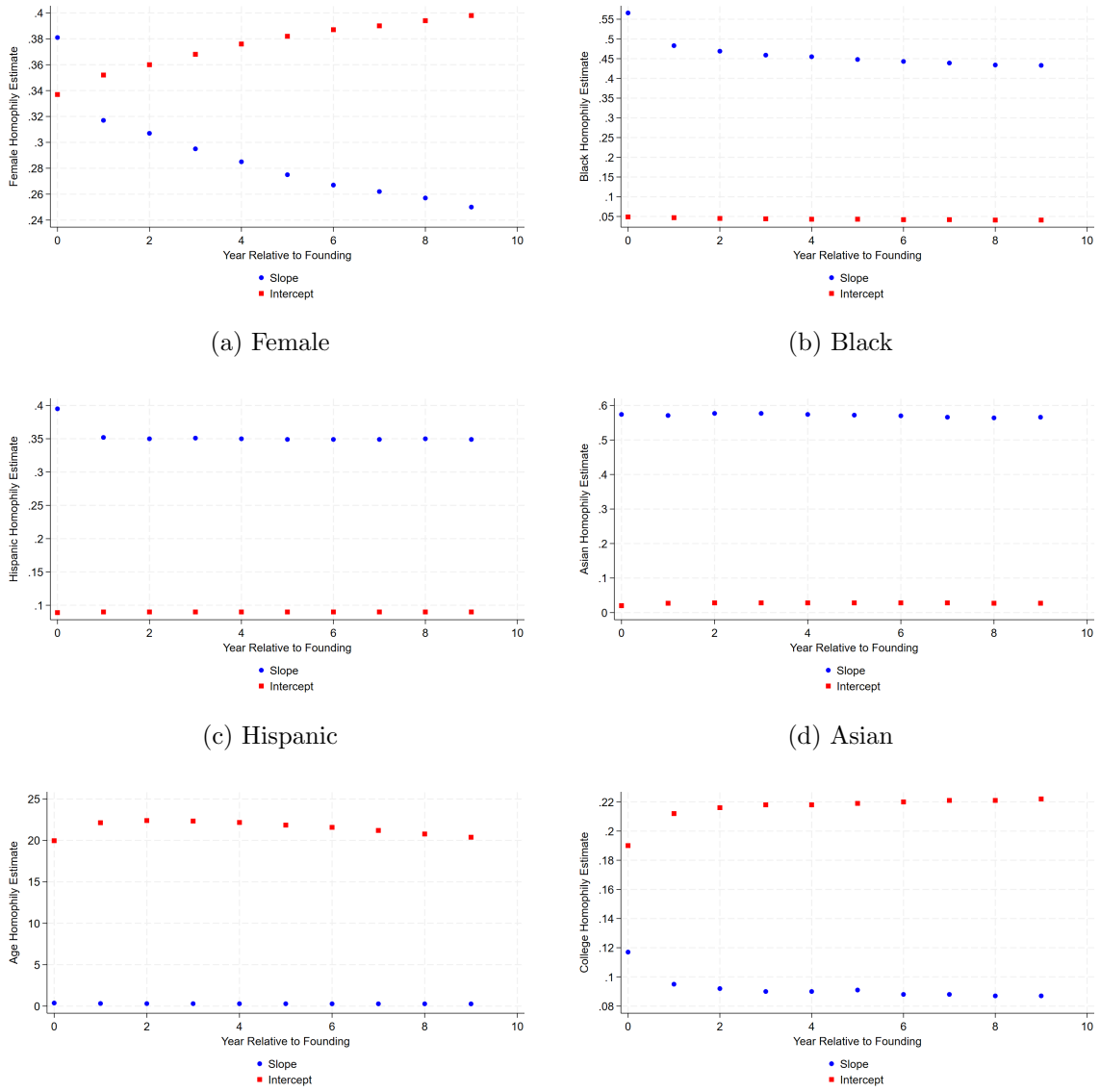
(b) Black



(c) Hispanic

Figure A3: Comparison of Overall versus New Hire Homophily Estimates (Crunchbase)

Notes: estimates of Equation 1 using the unbalanced Crunchbase sample, focusing on the demographic composition of new hires. Standard errors are clustered at the firm level and the figure reports 95 percent confidence intervals.



(a) Female

(b) Black

(c) Hispanic

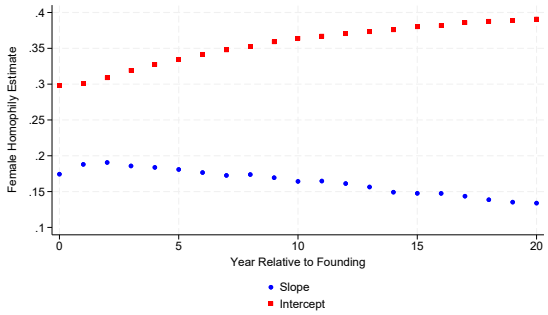
(d) Asian

(e) Age

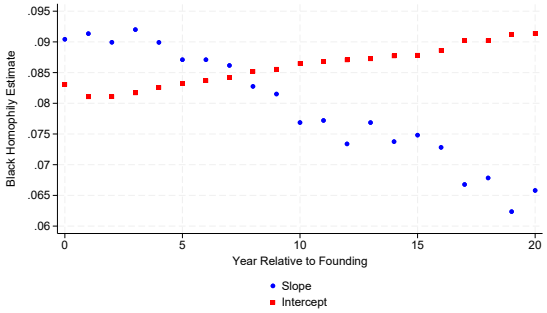
(f) College Educated

Figure A4: Estimates from Unbalanced Sample of New Firms

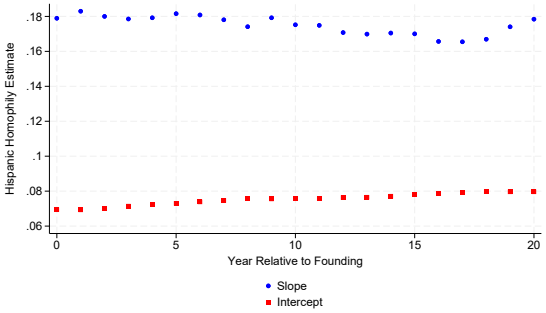
Notes: estimates of Equation 1 using the unbalanced LEHD sample. All standard errors are negligible relative to the point estimates, so we have not added confidence intervals to the plot. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB)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



(a) Female



(b) Black



(c) Hispanic

Figure A5: Entrepreneur-Worker Homophily (Unbalanced Sample, Crunchbase)

Notes: estimates of Equation 1 using the unbalanced Crunchbase sample, covering years 0 through 20 relative to founding. Standard errors are clustered at the firm level and the figure reports 95 percent confidence intervals.

B.E Founder versus Initial Employees – LEHD Sample

Here, we report the horserace estimates using the LEHD sample. In the main text, we report estimates using the Crunchbase-LinkedIn sample, given its greater reliability in measuring founders. Consistent with the results from the other sample, we find that founder effects persist over time, whereas the initial employee composition decays over time.

B.F Startups Versus Public Traded Companies

As noted in the main text, a different way to assess potential long-run impacts is to study publicly traded companies. To do so, we again lean on the Revelio Labs data. We focus on the 2010–2019 period, and extract average C-suite demographics and employee demographics at the firm-by-year level. For comparison, we also keep data on startups in aged 0–5 (matching our core balanced sample estimates) in this period.

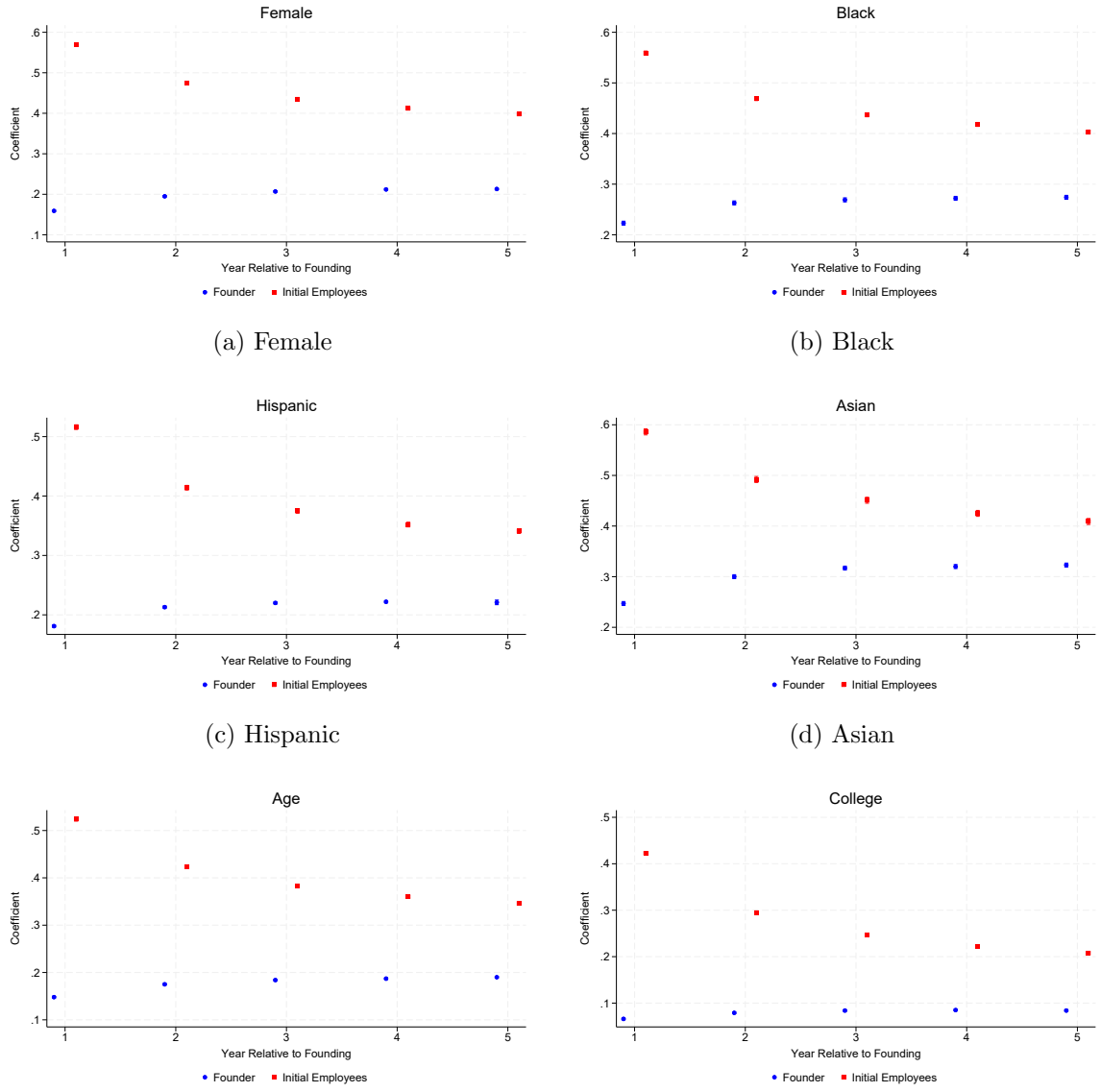
Table A4 compares homophily patterns at startups (using founder demographics) with those at publicly traded companies (using C-suite demographics). We find quantitatively similar homophily estimates across the two settings, with somewhat larger gender homophily at startups (0.201 versus 0.145) and similar or larger racial homophily at public firms. The publicly traded firms also have higher baseline shares of female and minority employees, consistent with the intercept trends observed in the long-run unbalanced panel estimates.

B.G Crunchbase Robustness Checks

We conduct several robustness checks within the Crunchbase sample, paralleling those performed for the LEHD.

Alternative Fixed Effects. Figure A7 compares baseline estimates (no fixed effects) to specifications with industry-by-founding year and industry-by-state-by-founding year fixed effects. The estimates become slightly smaller with more detailed controls but remain quantitatively large.

Weighting by Firm Size. Tables A5 and A6 present estimates from the balanced Crunchbase sample under two alternative weighting schemes: square root of employees and total employees, respectively. The estimates are somewhat smaller under employee weighting (e.g., 0.138 versus 0.174 for female in year 0 with linear weights), consistent with slightly



(a) Female (b) Black

(c) Hispanic (d) Asian

(e) Age (f) College Educated

Figure A6: Horserace Estimates Using the LEHD Sample

Notes: estimates of Equation 1 using the balanced LEHD sample, with an additional variable measuring the initial employee composition. 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-FY26-P2229-R1296).

Table A4: C-Suite versus Founder

Panel A: Startups

	Female Employee Fraction	Black Employee Fraction	Hispanic Employee Fraction
Female Founder Fraction	0.201***		
	(0.00388)		
Black Founder Fraction		0.0962***	
		(0.00412)	
Hispanic Founder Fraction			0.180***
			(0.00602)
Constant	0.301***	0.0785***	0.0665***
	(0.00301)	(0.00114)	(0.00129)
Observations	212,478	224,762	224,762

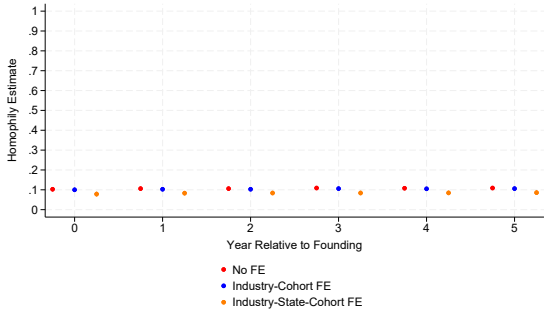
Panel B: Publicly Traded Companies

	Female Employee Fraction	Black Employee Fraction	Hispanic Employee Fraction
Female C-Suite Fraction	0.145***		
	(0.0111)		
Black C-Suite Fraction		0.147***	
		(0.0138)	
Hispanic C-Suite Fraction			0.243***
			(0.0307)
Constant	0.364***	0.0811***	0.0710***
	(0.00369)	(0.00131)	(0.00171)
Observations	11,621	11,735	11,735

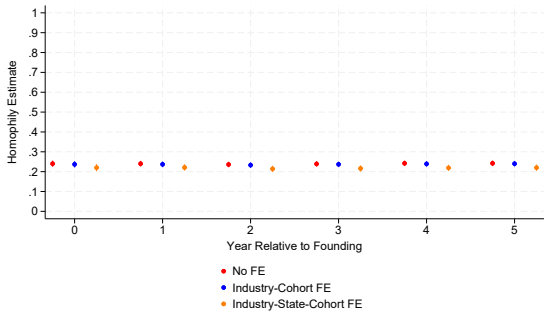
Notes: Pooled cross-sectional regressions comparing homophily for publicly-traded companies versus startups that are within 5 years of founding in the 2010–2019 period, with year FEs included throughout to account for secular trends. Standard errors in parentheses and are clustered at the firm level. *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: Crunchbase merged to LinkedIn (Revelio Labs).



(a) Female



(b) Black



(c) Hispanic

Figure A7: Estimates from Balanced Sample of New Firms (Crunchbase)

Notes: estimates of Equation 1 using the balanced Crunchbase sample.

Table A5: Entrepreneur-Worker Homophily Weighting by Square Root of Employees (Balanced Sample, Crunchbase)

Panel A: Female

Year relative to founding:	0	1	2	3	4	5
Female Founder Fraction	0.164***	0.173***	0.176***	0.175***	0.171***	0.166***
	(0.00561)	(0.00492)	(0.00471)	(0.00469)	(0.00475)	(0.00478)
Constant	0.333***	0.334***	0.341***	0.349***	0.356***	0.363***
	(0.00188)	(0.00157)	(0.00145)	(0.00139)	(0.00138)	(0.00139)
N	44545	44545	44545	44545	44545	44545

Panel B: Black

Year relative to founding:	0	1	2	3	4	5
Black Founder Fraction	0.0985***	0.104***	0.107***	0.109***	0.108***	0.109***
	(0.00681)	(0.00646)	(0.00621)	(0.00621)	(0.00614)	(0.00643)
Constant	0.0793***	0.0783***	0.0784***	0.0791***	0.0799***	0.0807***
	(0.000675)	(0.000581)	(0.000538)	(0.000522)	(0.000513)	(0.000526)
N	44545	44545	44545	44545	44545	44545

Panel C: Hispanic

Year relative to founding:	0	1	2	3	4	5
Hispanic Founder Fraction	0.226***	0.231***	0.229***	0.229***	0.227***	0.229***
	(0.0102)	(0.00909)	(0.00881)	(0.00868)	(0.00846)	(0.00849)
Constant	0.0783***	0.0775***	0.0783***	0.0797***	0.0809***	0.0822***
	(0.000915)	(0.000770)	(0.000713)	(0.000690)	(0.000681)	(0.000683)
N	44545	44545	44545	44545	44545	44545

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

lower homophily among the largest firms. However, the estimates under linear weighting increase as firms age, suggesting that growing firms with different founder backgrounds maintain compositional differences over time.

B.H Additional Crunchbase Heterogeneity Results

Industry Subsamples. Table A7 reports homophily estimates across broad Crunchbase industry groups (Software, Hardware, Manufacturing, Health Care). The results show consistent positive homophily across industries, though the magnitudes vary by industry and demographic group, with some estimates imprecisely estimated in smaller industries.

Employment Growth. Table A8 splits the Crunchbase sample by employment growth, comparing the fastest-growing decile to the remaining firms. The estimates are similar across both groups.

Firm Size. Table A9 presents estimates by firm size quintile. We find slightly smaller

Table A6: Entrepreneur-Worker Homophily Weighting by Employees (Balanced Sample, Crunchbase)

Panel A: Female

Year relative to founding:	0	1	2	3	4	5
Female Founder Fraction	0.138***	0.147***	0.153***	0.157***	0.159***	0.154***
	(0.0285)	(0.0245)	(0.0220)	(0.0205)	(0.0200)	(0.0193)
Constant	0.369***	0.367***	0.369***	0.373***	0.377***	0.382***
	(0.0138)	(0.0116)	(0.0101)	(0.00888)	(0.00812)	(0.00781)
N	44545	44545	44545	44545	44545	44545

Panel B: Black

Year relative to founding:	0	1	2	3	4	5
Black Founder Fraction	0.107***	0.112***	0.114***	0.116***	0.118***	0.123***
	(0.0238)	(0.0225)	(0.0214)	(0.0206)	(0.0204)	(0.0210)
Constant	0.0738***	0.0739***	0.0745***	0.0758***	0.0761***	0.0762***
	(0.00422)	(0.00392)	(0.00363)	(0.00328)	(0.00320)	(0.00330)
N	44545	44545	44545	44545	44545	44545

Panel C: Hispanic

Year relative to founding:	0	1	2	3	4	5
Hispanic Founder Fraction	0.130***	0.143***	0.153***	0.156***	0.154***	0.158***
	(0.0388)	(0.0316)	(0.0280)	(0.0249)	(0.0223)	(0.0208)
Constant	0.0875***	0.0866***	0.0856***	0.0856***	0.0858***	0.0861***
	(0.00393)	(0.00327)	(0.00277)	(0.00238)	(0.00213)	(0.00199)
N	44545	44545	44545	44545	44545	44545

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

Table A7: 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 A8: 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 *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A9: 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
Black 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
Hispanic 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: $^*p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

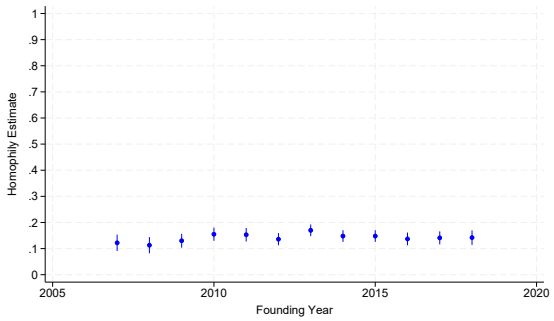
coefficients for the top quintile but estimates that remain quantitatively large across all size groups.

Cohort Trends. Figure A8 reports homophily estimates by detailed founding year in the Crunchbase sample. We find fairly stable estimates with a slight upwards trend.

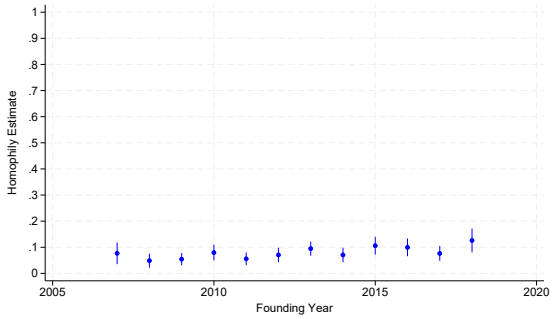
B.I Production Structure and Additional Controls

A potential concern with our core results is that founders from different backgrounds may create firms with systematically different production structures, and that these differences are driving the observed employee composition patterns. We address this concern using data from the Economic Census.

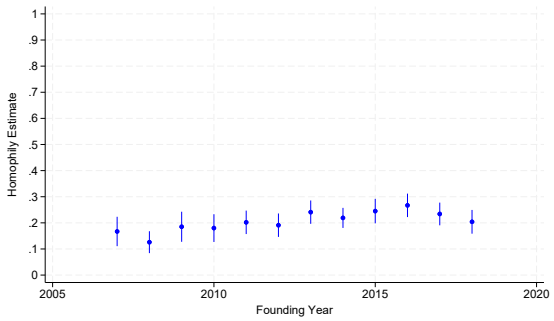
Founder Characteristics and Production Structure. Tables A10, A11, A12, and A13 report associations between founder demographics and four measures of firm production



(a) Female



(b) Black



(c) Hispanic

Figure A8: 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.

Table A10: Association between Founder Characteristics and Average Wage

	Average Wage					
Female Founder Fraction	-15.27***					
	(0.220)					
Black Founder Fraction		-11.63***				
		(0.371)				
Hispanic Founder Fraction			-6.467***			
			(0.277)			
Asian Founder Fraction				-0.406		
				(0.402)		
Average Founder Age					0.282***	
					(0.008)	
College Grad Founder Fraction						19.76***
						(0.300)
Constant	35.33***	28.69***	28.71***	28.01***	17.09***	22.44***
	(0.134)	(0.057)	(0.067)	(0.056)	(0.294)	(0.072)
Observations	337000	337000	337000	337000	337000	337000

Notes: Regressions at the firm level, based on characteristics measured during the first Economic Census available for the firm (2007 or 2012). All specifications include Industry-MSA-founding year fixed effects. Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

structure: average wage, labor share, import intensity, and export intensity. These are the primary measures available to us in our Census project. These regressions use firm-level data from the 2007 and 2012 Economic Census, restricted to new firms within 5 years of founding.

We find meaningful differences in average wages: female-founded firms pay $15.3k less on average, Black-founded firms $11.6k less, and Hispanic-founded firms $6.5k less, while college-educated founders pay $19.8k more. Differences in labor share, import intensity, and export intensity are statistically significant in some cases but economically small.

Homophily Controlling for Production Characteristics. Given these observed differences, we want to test whether they absorb some of the homophily estimates. Table A14 presents homophily estimates with and without controls for the production structure variables described above. The sample is restricted to firms founded between 2002 and 2012 that can be matched to the Economic Census, which is why the observation count differs from the main balanced sample.

Across all demographic groups except college education, the changes in homophily coefficients after including production structure controls is quantitatively small. For example, the female coefficient moves from 0.385 to 0.382, and the Black coefficient from 0.519 to

Table A11: Association between Founder Characteristics and Labor Share

	Labor Share					
Female Founder Fraction	-0.020*					
	(0.009)					
Black Founder Fraction		0.021***				
		(0.005)				
Hispanic Founder Fraction			-0.020***			
			(0.004)			
Asian Founder Fraction				-0.035***		
				(0.007)		
Average Founder Age					0.001***	
					(0.000)	
College Grad Founder Fraction						0.051***
						(0.011)
Constant	0.323***	0.312***	0.315***	0.316***	0.268***	0.299***
	(0.006)	(0.002)	(0.003)	(0.003)	(0.005)	(0.002)
Observations	337000	337000	337000	337000	337000	337000

Notes: Regressions at the firm level, based on characteristics measured during the first Economic Census available for the firm (2007 or 2012). All specifications include Industry-MSA-founding year fixed effects. Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

Table A12: Association between Founder Characteristics and Import Intensity

	Import Intensity					
Female Founder Fraction	-0.001					
	(0.000)					
Black Founder Fraction		-0.002***				
		(0.000)				
Hispanic Founder Fraction			-0.004***			
			(0.000)			
Asian Founder Fraction				0.006***		
				(0.001)		
Average Founder Age					0.000*	
					(0.000)	
College Grad Founder Fraction						0.006***
						(0.000)
Constant	0.007***	0.007***	0.007***	0.006***	0.006***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	337000	337000	337000	337000	337000	337000

Notes: Regressions at the firm level, based on characteristics measured during the first Economic Census available for the firm (2007 or 2012). All specifications include Industry-MSA-founding year fixed effects. Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

Table A13: Association between Founder Characteristics and Export Intensity

	Export Intensity					
Female Founder Fraction	-0.001**					
	(0.000)					
Black Founder Fraction		-0.000				
		(0.001)				
Hispanic Founder Fraction			-0.001*			
			(0.001)			
Asian Founder Fraction				0.002***		
				(0.001)		
Average Founder Age					0.000	
					(0.000)	
College Grad Founder Fraction						0.003***
						(0.001)
Constant	0.007***	0.006***	0.006***	0.006***	0.005***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Observations	337000	337000	337000	337000	337000	337000

Notes: Regressions at the firm level, based on characteristics measured during the first Economic Census available for the firm (2007 or 2012). All specifications include Industry-MSA-founding year fixed effects. Standard errors in parentheses *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: authors' calculations based on U.S. Census data (FSRDC Project Number 2229, Disclosure Review Board (DRB) approval number: CBDRB-FY25-P2229-R11972).

0.517. The college-educated coefficient shows a more meaningful decline (0.092 to 0.079), suggesting that some of the education-based sorting reflects differences in firm type. Overall, these results suggest that the production structure measures available to us do not account for most of the observed homophily patterns.

B.J AKM Robustness Checks and Additional Results

First, Table A15 presents the corresponding estimates when we compute firm fixed effects without any additional controls (the X's in the first-stage AKM wage equation). We find qualitatively similar but larger effects relative to the estimates presented in the main text. This suggests that there are significant observable differences across individuals hired by entrepreneurs from different backgrounds.

Next, we present regressions of the raw demographic fixed effect versus founder characteristics in Table A16. The slope captures the difference from having more founders from the same demographic background, and the intercept reflects how the full sample of startups compare to the broader set of firms. We generally find that female and minority founded firms pay less than other startups. College-educated founders appear to pay college-educated individuals more than both startups and the broader sample of firms.

Table A14: Homophily Estimates Controlling for Production Characteristics

Employee Fraction:	Female		Black		Hispanic		Asian		Age		College	
Female Found. Fr.	0.385***	0.382***										
	(0.001)	(0.001)										
Black Found. Fr.			0.519***	0.517***								
			(0.002)	(0.002)								
Hispanic Found. Fr.					0.358***	0.357***						
					(0.002)	(0.002)						
Asian Founder Fr.							0.564***	0.564***				
							(0.002)	(0.002)				
Founder Age									0.347***	0.343***		
									(0.001)	(0.002)		
College Found. Fr.											0.092***	0.079***
											(0.001)	(0.003)
Constant	0.385***	0.391***	0.049***	0.053***	0.094***	0.096***	0.026***	0.026***	19.66***	19.47***	0.201***	0.187***
	(0.001)	(0.002)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.054)	(0.070)	(0.000)	(0.003)
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Observations	2021000	2021000	2021000	2021000	2021000	2021000	2021000	2021000	2021000	2021000	2021000	2021000

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

Table A15: AKM Decomposition Regressions (No Controls)

	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).

Table A16: AKM Decomposition Regressions, Raw Firm Effects

	Female FE	Black FE	Hispanic FE	Asian FE	College FE
Female Founder Fraction	-0.256***				
	(0.002)				
Black Founder Fraction		-0.315***			
		(0.005)			
Hispanic Founder Fraction			-0.189***		
			(0.003)		
Asian Founder Fraction				-0.098***	
				(0.004)	
College Grad Founder Fraction					0.596***
					(0.003)
Constant	-0.155***	-0.137***	-0.684***	-0.142***	-0.341***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Observations	1,775,000	791,000	1,802,000	1,104,000	584,000

Notes: The dependent variable is the raw firm fixed effect for the indicated demographic group $\hat{\gamma}_{jsd}$. Firm fixed effects are estimated separately by state and demographic group, controlling for age, age squared, and firm tenure. 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-FY26-P2229-R12962).

Third, we implement the approach in Bonhomme et al. (2019) and Lerner et al. (2024) to test for the importance of complementarities between worker fixed effects and firms in our setting. The two models we compare are simplified linear and interacted versions of the core AKM approach:

$$Y_{it} = \alpha_i + \psi_{g(j,t)} + \varepsilon_{it}$$

$$Y_{it} = \sum_{k=1}^{K} \beta_k \, \alpha_i \, \mathbf{1}\{g(j,t) = k\} + \sum_{k=2}^{K} \gamma_k \, \mathbf{1}\{g(j,t) = k\} + \varepsilon_{it}$$

Essentially, the first is a simplified version of AKM where firms are clustered into K groups indexed by g (instead of firm-specific fixed effects). The second is a model that allows for flexible complementarities (β_k) that are specific to each group. The specific steps for the implementation are listed below (done state by state due to computational limitations):

1. Clustering
    (a) We first purge the wage data of state-year mean effects to make the data comparable across years.
    (b) Using a set of "stayers" pooled across all years in the sample, we summarize each firm based on 21 quantiles (p1, p5, p10, p15,..., p95, p99).
    (c) We then cluster the firms into 10 clusters via K-means (500 iterations, 17 random starts)
2. Creating the estimation sample
    (a) We focus on a set of movers who change firms for the first time, and have earnings data in periods $t-1$ and $t+1$.
    (b) We then randomly sample non-movers with the same data coverage in the same years (each control has a pseudo "move year")
3. Estimating the models
    (a) Using the estimation sample, we then start by estimating the linear model.

Table A17: Non-linear versus linear models

Panel A: Model Fit Statistics

	R^2	$\frac{Var(\hat{\alpha})}{Var(y)}$	$\frac{Var(\hat{\gamma})}{Var(y)}$	$\frac{2 \cdot Cov(\hat{\alpha},\hat{\gamma})}{Var(y)}$	$\frac{Var(\epsilon)}{Var(y)}$	$Corr(\hat{\alpha},\hat{\gamma})$
Interacted Model	0.8063	0.5466	0.09191	0.1723	0.1891	0.3867
Linear Model	0.8106	0.5425	0.09687	0.1713	0.1894	0.3758

Panel B: Comparison of Match Coefficients

	Estimated Slope for Firm's Cluster
New Firm	0.021***
	(0.000)
Constant	0.985***
	(0.000)
Observations	5,594,000

Notes: Panel A follows the methodology in Bonhomme et al. (2019) to compare linear versus interacted models, using ten clusters of firms for each state. Panel B compares the group-level estimated slope for new firms in our sample versus the broader sample of firms. 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-FY26-P2229-R12962).

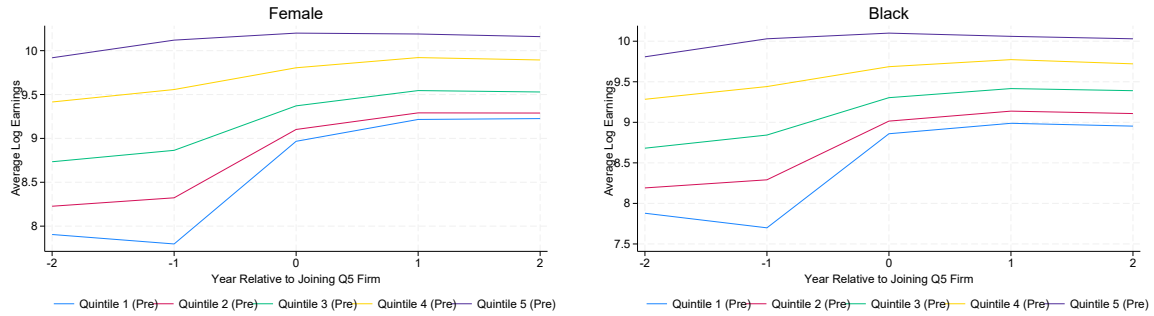
(b) We then estimate the interacted model:

i. The worker covariates α_i are constructed using the estimates from the linear model.

ii. We then instrument for the interacted covariates using a LIML instrumental variables approach. The instruments are constructed based on pre-move log earnings interacted with the group indicators.

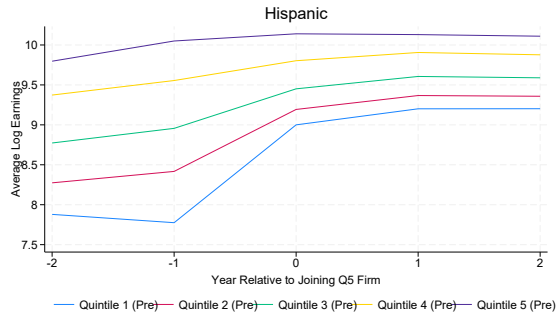
Table A17, Panel A reports statistics related to the fit of the two models. We average the stats across all the states for which we have data access. The models have similar fit statistics, which suggests that complementarities are not playing a large role in our setting. To focus on new firms in our sample, we also compare the average β_k (based on the clusters into which each firm is sorted) for new firms and the other firms in the overall LEHD sample. We find that new firms exhibit higher complementarities with worker abilities, a difference that is statistically significant but not quantitatively large.

Finally, we validate the group-specific firm fixed effects estimates using descriptive event studies, following Card et al. (2013). Each firm in our sample is classified into quintiles

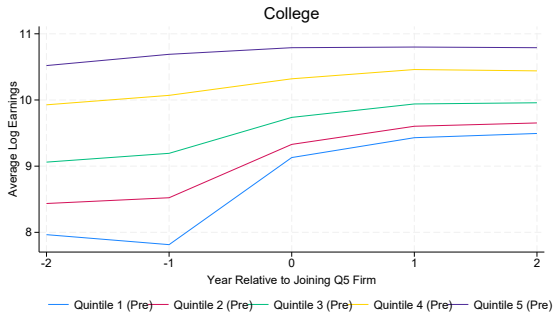


(a) Female

(b) Black



(c) Hispanic



(d) College

Figure A9: Descriptive Evidence for Movers Moving to Quintile Five Firms

Notes: descriptive event studies showing average log earnings for different mover groups and different demographics. 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-FY26-P2229-R12962).

for each worker demographic (female, Black, etc.), with quintile 5 representing firms with the largest estimated effects. Because of disclosure limits, we focus on workers who move to quintile 5 firms in each demographic group. Figure A9 presents the results. We see relatively flat trends in the two years before the move, and larger jumps for movers who move from low quintile firms. This provides some support for the reliability of the firm-demographic fixed effects approach.

C Extra Model Details

C.A Model Derivations

We present the cost-minimization problem and derive closed-form expressions for factor demands, the unit cost of the composite labor bundle, and the entrepreneurial entry threshold.

C.A.1 Cost Minimization and Factor Demands

An entrepreneur of type g with ability z chooses labor inputs $\{\ell_{gm}, \ell_{gf}\}$ to minimize total labor cost subject to producing a given level of the composite labor bundle:

$$\min_{\ell_{gm},\ell_{gf}} w_m \ell_{gm} + w_f \ell_{gf} \quad \text{s.t.} \quad \left[A_{gm}^{1/\gamma} \, \ell_{gm}^{\frac{\gamma-1}{\gamma}} + A_{gf}^{1/\gamma} \, \ell_{gf}^{\frac{\gamma-1}{\gamma}} \right]^{\frac{\gamma}{\gamma-1}} \geq \ell_g$$

The first-order conditions yield the conditional factor demand for type-h workers:

$$\ell_{gh}(z) = A_{gh} \left(\frac{w_h}{\omega_g} \right)^{-\gamma} \ell_g(z)$$

where ω_g is the unit cost (cost index) of the composite labor bundle for a type-g founder

$$\omega_g = \left[A_{gm} \, w_m^{1-\gamma} + A_{gf} \, w_f^{1-\gamma} \right]^{\frac{1}{1-\gamma}}$$

C.A.2 Optimal Scale and Profits

Given the unit cost ω_g, the entrepreneur's profit-maximization problem reduces to a single-variable choice:

$$\max_{\ell_g} z \, \ell_g^{\rho} - \omega_g \, \ell_g$$

The optimal composite labor input is:

$$\ell_g(z) = \left(\frac{\rho \, z}{\omega_g} \right)^{\frac{1}{1-\rho}}$$

and the resulting variable profit is:

$$\pi_g(z) = (1-\rho) \, z \, \ell_g(z)^{\rho} = (1-\rho) \, z^{\frac{1}{1-\rho}} \left(\frac{\rho}{\omega_g} \right)^{\frac{\rho}{1-\rho}}$$

C.A.3 Entry Threshold

An individual of type g with access to entrepreneurship enters if and only if $\pi_g(z) \geq w_g$, i.e.,

$$(1-\rho) \, z^{\frac{1}{1-\rho}} \left(\frac{\rho}{\omega_g} \right)^{\frac{\rho}{1-\rho}} \geq w_g$$

Solving for the cutoff ability level:

$$z_g^* = \left[\frac{w_g}{1-\rho}\right]^{1-\rho} \cdot \left(\frac{\omega_g}{\rho}\right)^{\rho}$$

Individuals with $z \geq z_g^*$ and access (ϕ_g) become entrepreneurs; all others supply labor at the market wage w_g.

C.A.4 Market Clearing

Equilibrium wages (w_m, w_f) satisfy two market clearing conditions (one for each worker group):

$$\sum_{g \in \{m,f\}} N_g \, \phi_g \int_{z_g^*}^{\infty} A_{gh} \left(\frac{w_h}{\omega_g}\right)^{-\gamma} \left(\frac{\rho \, z}{\omega_g}\right)^{\frac{1}{1-\rho}} dF(z) = N_h \left(1 - \phi_h \left[1 - F(z_h^*)\right]\right), \quad h \in \{m, f\}$$

C.B Details on Model Computation

We provide additional notes on how we solve for the equilibrium under any given parameter guess:

1. Guess market wages w_m, w_f and compute excess labor demand at these candidate wages:
   (a) For each agent, figure out their career choice, by computing expected entrepreneurial profits as a function of ability z and comparing them to w
   (b) Then compute aggregate supply and demand for each group of workers
2. Find market wages to ensure that the absolute value of the difference between labor demand and labor supply is less than 10^{-10}.
   (a) At the equilibrium wages, compute the four moments

We then wrap an fminsearch call around this routine to find the optimal parameter values. The optimization uses MATLAB's Nelder-Mead simplex algorithm with default tolerances.

Table A18 summarizes the input data moments for all of the different runs.

Table A18: Moments used in the calibration

Moment	Female	Black	Hispanic	Non-College
Minority worker share at minority-founded firms	0.717	0.559	0.465	0.805
Minority worker share at majority-founded firms	0.353	0.047	0.092	0.692
Minority founder share	0.408	0.054	0.105	0.719
Group size ratio	1.000	0.081	0.151	3.42
Minority-to-majority wage ratio	0.80	0.75	0.73	0.60

Notes: The first two rows are based on the pooled estimates from the balanced LEHD sample in Panel C of Table A2. The next two rows are based on the summary statistics for founders and workers in Table 1. The final row contains approximate wage ratios based on various government reports from the sample period.

C.C Back-of-the-Envelope Calculations

In the main text, we use a full equilibrium model to assess the implications of our empirical findings for workers. Here, we provide a simpler linear approximation approach. 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:

$$\Delta \log\left(\frac{w_W}{w_M}\right) = \frac{1}{\sigma} \cdot \Delta \log\left(\frac{L_W}{L_M}\right) \tag{A1}$$

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:

$$\Delta \log\left(\frac{L_W}{L_M}\right) \equiv \log\left(s'_M \left(\frac{L_W}{L_M}\right)_{\text{MaleStartup}} + s'_W \left(\frac{L_W}{L_M}\right)_{\text{FemaleStartup}}\right) \tag{A2}$$

$$- \log\left(s_M \left(\frac{L_W}{L_M}\right)_{\text{MaleStartup}} + s_W \left(\frac{L_W}{L_M}\right)_{\text{FemaleStartup}}\right), \tag{A3}$$

where s_M and s_W denote the shares of female and male startups observed today, with $s_M + s_W = 1$. Note that this expression computes a weighted average of firm-level demographic ratios rather than the ratio of aggregate worker levels. Because the ratio L_W/L_M is a convex function of the underlying worker counts, this approximation may overstate the aggregate ratio for firm types where the denominator group is small. Nonetheless, this simplified expression contains quantities we can estimate directly from the data and, when combined with estimates of the elasticity of substitution, can be used to obtain approximate changes in the wage ratio using Equation A1.

Table A19 presents calculations using our summary stats and estimates. We project sizable increases in relative wage for women (6.37%), Blacks (9.98%), and Hispanics (5.23%). The estimate for non-college workers (1.42%) is smaller because the difference across founders is smaller relative to the baseline rate. These estimates are broadly comparable to those implied by the counterfactual in the full model, though differences arise due to non-linearities and general equilibrium effects that the linear approximation does not capture.

Table A19: 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 Wages
Female	0.408	0.501	0.353	0.364	2.53	0.55	6.37%
Black	0.054	0.075	0.047	0.512	1.27	0.05	9.98%
Hispanic	0.105	0.131	0.092	0.373	0.87	0.10	5.23%
Non-college	0.719	0.774	0.692	0.113	4.13	2.25	1.42%

Notes: Back-of-the-envelope calculations using Equations A1 and A2,and plugging in estimates from the summary stats table (Table 1, Panel B for employee shares, Panel A for founder shares) and the pooled LEHD balanced sample regression (Table A1, Panel C for Intercept and Slope). The L_W/L_M ratios are computed as (Intercept + Slope) / (1 - Intercept - Slope) for the "group" column and Intercept / (1 - Intercept) for the "other" column. ΔRelative Wages are computed using $\sigma = 2$. The counterfactual sets the founder share equal to the employee share.