The Invisible Majority: Selection Bias in Self-Reported Data

Emmanuel Yimfor*

December 1, 2025

Abstract

Self-reported diversity data are increasingly used to study disparities in entrepreneurship, yet selection into self-reporting may bias conclusions. Using Crunchbase Diversity Spotlight, the largest voluntary diversity database for startups, I provide the first systematic analysis of this selection bias. Algorithmic classification identifies 8,564 Black founders; only 12.6% appear in self-reported data. The 87% who do not self-report have weaker credentials and raise less capital, but achieve better exit outcomes per dollar invested. This selection reverses conclusions about funding gaps: with comprehensive data, Black founders are 8.7 percentage points less likely to raise VC; with self-reported data, they appear 19 percentage points more likely. Selection is driven by investor networks: backing by a launch partner increases self-identification probability six-fold. These findings have direct implications for California's SB 54, which requires VCs to collect voluntary founder diversity data starting in 2026: without accounting for selection, conclusions drawn from such data may be misleading.

Keywords: Venture Capital, Race, Selection Bias, Entrepreneurship, Self-Reported Data
JEL Classification: G24, J15, L26, M13

*Columbia Business School, Columbia University; emmanuel.yimfor@columbia.edu

1. Introduction

Self-reported diversity data are increasingly central to research and policy on entrepreneurship. Prominent studies of funding gaps by race and gender rely on the Kauffman Firm Survey, where owner race and gender are self-reported (Robert Fairlie, Alicia Robb and David T Robinson, 2022; Robert Fairlie and David T Robinson, 2023; Marta Morazzoni and Andrea Sy, 2022). The Federal Reserve's Small Business Credit Survey, an opt-in survey, uses self-reported owner race to measure disparities in credit access (Brett Barkley and Mark E Schweitzer, 2022). The Census Bureau's Annual Business Survey collects self-reported owner demographics to study innovation and venture capital outcomes (Timothy R Wojan, 2024). Yet selection into self-reporting is unlikely to be random. If founders who opt in differ systematically from those who do not, conclusions drawn from self-reported data could be misleading or even directionally wrong. This paper provides a systematic analysis of selection bias in voluntary self-reported founder race in venture capital. I focus on a setting where I can approximate the full population of Black startup founders by applying algorithmic race classification to the universe of founders. Comparing this population to the subset who voluntarily self-report reveals severe selection bias that reverses key conclusions about funding disparities.

The policy stakes are immediate. California's SB 54, signed in 2023, requires venture capital firms to collect and report voluntary founder diversity data starting in 2026.[1] The law explicitly makes founder participation optional and prohibits VCs from incentivizing responses.

I study Crunchbase Diversity Spotlight, the largest existing voluntary diversity database for startups. Launched in August 2020, just weeks after George Floyd's murder, the program allows companies to disclose diversity characteristics of their leadership. Using algorithmic race classification following Lisa D Cook, Matt Marx and Emmanuel Yimfor (2025), I identify 8,564 Black founders in Crunchbase; only 1,077 (12.6%) appear in Diversity Spotlight. The 87% who do not self-report are invisible to any analysis relying on these data. Media outlets have widely cited Diversity Spotlight statistics, reporting that "Black founders receive less than 1% of venture capital."[2] Academic researchers have been more cautious, with most studies explicitly avoiding self-reported data (Cook, Marx and

[1]See California Senate Bill 54, signed October 2023, with amendments in SB 164 signed June 2024. The law requires covered VC firms to survey portfolio company founders on demographics including race, gender, and veteran status, and report aggregated results to the California Department of Financial Protection and Innovation. For details, see https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB54.

[2]See, e.g., "Funding to Black founders was down in 2023 for the third year in a row," *TechCrunch*, January 2024; "Venture capital for Black entrepreneurs plummeted 45% in 2022," *CNBC*, February 2023.

Yimfor, 2025; Matt Marx, Qian Wang and Emmanuel Yimfor, 2025), but the systematic consequences of selection into self-reporting have not been documented.

The founders who self-identify differ dramatically from those who do not. Self-identified founders have stronger credentials: they are 14 percentage points more likely to have attended a top school, hold 65% more prior startup experience, and are far more likely to have raised capital (73% vs. 25%). But conditional on funding raised, they achieve *worse* exit outcomes, with a 1.8 percentage point lower probability of a good exit. This selection is severe enough to reverse conclusions about funding gaps: with comprehensive data, Black founders are 8.7 percentage points less likely to raise VC; with self-reported data, they appear 19 percentage points *more* likely.

These patterns likely generalize beyond Crunchbase. Any voluntary diversity disclosure system faces the same fundamental challenge: participation requires awareness of the program, resources to complete disclosure, and a belief that disclosure will be beneficial. The direction of selection is not obvious ex ante: if disclosing race invites discrimination, less sophisticated founders might be more likely to disclose, generating negative selection. Indeed, Raffi E García and William A Darity Jr (2022) find exactly this pattern in PPP loans, where Black owners who disclosed race received 52% less funding. In venture capital, however, these requirements correlate with founder quality and connectedness, generating positive selection. The institutional origins of Diversity Spotlight amplify this dynamic: the program launched with diversity-focused VCs as official partners, who encouraged their portfolio companies to register. Backing by a launch partner increases self-identification probability by 67.5 percentage points, a six-fold increase relative to baseline. Founders outside these networks, particularly those at earlier stages or in geographies without partner presence, had little reason to participate.

I organize my findings around three questions. First, by how much do self-reported data understate funding to Black founders? Second, what would a researcher conclude about racial funding gaps using self-reported data versus comprehensive data? Third, who selects into self-reporting?

On the first question, self-reported data dramatically understate how much capital flows to Black founders. Black founders receive 2.4% of total VC funding; self-reported data capture only 0.9%. The undercount varies systematically: self-reported data capture 37% of seed-stage funding to Black founders but only 26% at late stage. Geographic variation is similarly stark: in Texas, where Black founders receive 3.9% of VC funding, self-reported data capture just 0.1%.

On the second question, a researcher estimating funding gaps with self-reported data would reach fundamentally incorrect conclusions. With comprehensive data, I find that

Black founders are 8.7 percentage points less likely to raise VC and receive approximately 70% less funding conditional on raising. With self-reported data, the extensive margin coefficient *flips sign*: self-identified Black founders appear *more* likely to raise funding than non-Black founders, and the intensive margin gap shrinks from 70% to 23%. The pattern holds across funding stages: at seed, the true gap is −1.9 percentage points on the extensive margin and −58% on the intensive margin; with self-reported data, the extensive margin becomes +29 percentage points and the intensive margin shrinks to −20%.

Why do self-reported data produce such different results? Self-identified founders are far more likely to have raised capital: 73% versus 24% for any venture funding, and 11% versus 2% for late-stage funding. Average funding raised is $11.7 million for self-identified founders versus $1.9 million for those who do not self-identify. Comparing this elite subset to all non-Black founders naturally attenuates, or reverses, estimated funding gaps.

But are the self-identified founders genuinely higher quality, or are they simply better networked and more visible to investors? If self-identified founders are truly stronger, they should outperform conditional on funding, since they have better credentials and more experience. If they are merely better connected, their outcomes should be no better, or perhaps worse, than other Black founders who raise similar amounts. The evidence points decisively to the latter. Conditional on total funding raised, self-identified Black founders are 1.8 percentage points *less* likely to achieve a good exit (IPO or acquisition at ≥2x funding) and 2.2 percentage points less likely to achieve any exit. These are economically large effects: relative to base rates of 1% for good exits and 5% for any exit among Black founders, self-identification is associated with a 180% reduction in the probability of a good exit and a 44% reduction in the probability of any exit, holding funding constant.

This finding has a striking implication: self-identified Black founders are *overfunded* relative to their outcomes. Investors pour more capital into founders who are visible and well-credentialed, but these founders do not deliver better results per dollar invested. Self-reported data therefore not only misrepresent the size of funding gaps but also mischaracterize which Black founders are underfunded.

These findings have implications for how we measure and address racial disparities in venture capital. Studies relying on Crunchbase Diversity Spotlight or similar voluntary disclosures will systematically overstate funding gaps on the extensive margin, because they capture only a selected subset of Black founders, while understating gaps on the intensive margin, because the founders who self-report are disproportionately those who have already raised substantial capital. Overstating the funding gap may also discourage entry: experimental evidence shows that activating negative stereotypes about a group's likelihood of success in entrepreneurship reduces that group's entrepreneurial intentions (Vishal K Gupta, Daniel B Turban and Nachiket M Bhawe, 2008). If potential Black founders believe the odds are worse than they actually are, they may self-select out of entrepreneurship entirely. The funding response following George Floyd's murder illustrates a related concern: self-identified Black founders captured 52% of post-George Floyd funding despite representing only 12.6% of Black founders. If diversity-motivated capital flows primarily to founders who are already visible and already overfunded relative to outcomes, such initiatives may be less effective at expanding opportunity than their proponents hope.

This paper makes three contributions. First, I provide the first large-scale characterization of who selects into voluntary diversity disclosure and demonstrate that this selection reverses conclusions about racial funding gaps: researchers using self-reported data will overstate funding gaps on the extensive margin, understate them on the intensive margin, and misidentify which founders are underfunded. Second, I show that self-identified Black founders underperform conditional on funding, implying that visibility and credentials are imperfect proxies for quality. Third, I provide comprehensive data on Black founders to enable future research. The Black Founder Dataset, available at https://github.com/eyimfor/race_classifier_fbhgs, contains founder-level information on U.S.-based startups from 2009–2020. A companion dataset on minority venture capital partners is available at https://www.eyimfor.com/data/MinorityGroupList.xlsx (Johan Cassel, Josh Lerner and Emmanuel Yimfor, 2023).

The rest of the paper proceeds as follows. Section 2 describes sample construction and race classification. Section 3 presents summary statistics comparing self-identified and non-self-identified Black founders. Section 4 describes the empirical strategy. Section 5 presents findings on funding gaps, exit outcomes, and selection into self-identification. Section 6 discusses limitations. Section 7 concludes.

2. Data

2.1. Sample Construction

I construct a dataset of startup founders from Crunchbase, focusing on U.S.-based companies founded between 2000 and 2024. Crunchbase provides comprehensive coverage of venture-backed startups and is widely used in entrepreneurship research (Andre Retterath and Reiner Braun, 2020). The sample includes all individuals listed as founders, co-founders, or CEOs of U.S. startups in the Crunchbase database. Funding flow analyses

(Figures 1–4) use funding rounds from 2015–2024; regression analyses use the full founder sample.

A key step is matching founders to profiles that contain photos. I match founders to LinkedIn profiles using URLs provided in Crunchbase, supplemented by web searches for founders without direct links. For founders with LinkedIn profiles but no photo, I search company websites and social media for profile images. This process yields 174,347 unique founders with identifiable profiles, representing approximately 92% of the founder universe in Crunchbase during this period.

2.2. Race Classification

I classify founder race using the methodology developed in Cook, Marx and Yimfor (2025), which combines image-based classification with manual review. This methodology has been applied in several studies of race in entrepreneurship and finance, including Marx, Wang and Yimfor (2025) on venture funding responses to social movements, Cassel, Lerner and Yimfor (2023) on racial diversity in private capital fundraising, and Johan Cassel, James Weston and Emmanuel Yimfor (2023) on startup board diversity. A key challenge is that name-based algorithms alone perform poorly for identifying Black founders: due to historical naming patterns in the United States, surname-based classifiers struggle to distinguish Black from White Americans and have Type II error rates exceeding 70% for Black individuals (Daniel L Greenwald, Sabrina T Howell, Cangyuan Li and Emmanuel Yimfor, 2024). Image-based classification addresses this limitation by directly capturing perceived race from facial features.

The classification process has three stages. First, I collect profile photos from matched LinkedIn profiles, company websites, and social media accounts. Second, I apply DeepFace, an open-source facial analysis algorithm that infers demographic characteristics from facial features (Sefik Ilkin Serengil and Alper Ozpinar, 2020), combined with NamePrism for name-based probability estimates. The combination helps resolve ambiguous cases: for example, a founder with an Indian surname but Black-presenting photo would be flagged for manual review rather than automatically classified. This initial algorithmic pass identifies approximately 6,400 potential Black founders from the 174,347 founders with identifiable profiles.

Third, all algorithmically-flagged founders undergo manual review by trained research assistants to remove false positives. The most common source of false positives is dark-skinned South Asian founders, who are sometimes misclassified by the image algorithm despite having names that suggest South Asian origin. Additionally, all founders *not* flagged as Black undergo a secondary review to identify false negatives, particularly light-skinned Black founders whom the algorithm missed. This secondary review draws on multiple sources: membership in affinity groups such as the Nigerian Leadership Initiative, news reports identifying founders as Black, crowd-sourced lists of Black founders, and HBCU attendance records. The secondary review identifies approximately 2,100 additional Black founders, representing roughly 25% of the final sample. This labor-intensive process yields a final classification of 8,564 Black founders, representing 4.9% of the sample.

2.3. Validation

An important feature of this methodology is that it captures *perceived* race rather than self-reported race. This distinction matters for studying investor behavior, since investment decisions presumably respond to how investors perceive founders rather than how founders identify themselves. To validate the methodology, I apply the classification algorithm to the Chicago Face Database (Debbie S Ma, Joshua Correll and Bernd Wittenbrink, 2015), which contains 197 Black and 183 White faces with both actual race labels and perceived race ratings collected from approximately 43 participants per image.

The validation reveals that image-based classification correlates more strongly with third-party perceived race than with self-reported race. For Black faces, the correlation between algorithmic classification and perceived race is 0.94, compared to 0.86 for self-reported race. For White faces, the corresponding correlations are 0.88 and 0.79. This pattern confirms that the methodology captures perceived race, the characteristic most relevant for studying differential treatment in funding markets, rather than self-reported identity.

2.4. Crunchbase Diversity Spotlight

Crunchbase launched Diversity Spotlight on August 11, 2020, just weeks after George Floyd's murder intensified attention on racial equity. The program allows companies to voluntarily disclose diversity characteristics of their leadership teams, indicating whether leadership includes individuals who are Black or African American, Hispanic or Latino, women, veterans, or members of the LGBTQ community. Only verified company employees can add diversity tags, and participation is entirely voluntary.

The program's coverage remains limited. Crunchbase's methodology documentation explicitly acknowledges "voluntary or selective reporting bias," noting that "this may paint a rosier picture of a company or industry sector." The program launched with five official partners: Backstage Capital, Harlem Capital Partners, BLCK VC, All Raise, and Precursor Ventures.

I match the 8,564 Black founders identified via algorithmic classification to Diversity Spotlight using organization identifiers. A founder is classified as "self-identified" if (1) they are Black according to my algorithmic classification *and* (2) their organization appears in Diversity Spotlight with Black leadership indicated. This matching identifies 1,077 self-identified Black founders at 892 organizations.[3] The remaining 7,487 Black founders at 6,721 organizations are classified as non-self-identified. These founders are invisible to any analysis relying solely on self-reported diversity data.

The 12.6% self-identification rate raises an immediate question: who are the founders that self-identify, and how do they differ from those who do not? There are three possibilities. First, self-identified founders could be a representative sample of all Black founders, in which case self-reported data would yield unbiased estimates of funding gaps. Second, weaker founders might be more likely to self-identify if they seek visibility or believe that disclosing race will help them access diversity-focused capital. Third, stronger founders might be more likely to self-identify if participation requires awareness, resources, or connections that correlate with founder quality.

2.5. Outcomes and Controls

I measure funding outcomes using Crunchbase data on venture capital and angel investments, aggregated to the startup level. Key variables are total funding raised, number of funding rounds, indicators for raising any VC and for late-stage funding (Series B or later), and funding raised in specific periods. The post-George Floyd variable captures funding between June 2020 and December 2022.

I measure exit outcomes using Crunchbase acquisition and IPO records. A "good exit" is an IPO or acquisition at ≥2x total funding raised, capturing exits where equity holders likely earned positive returns. "Any exit" includes all IPOs and acquisitions regardless of price.

I measure founder characteristics using LinkedIn profiles. Education controls include degree attainment (bachelor's, master's, MBA, PhD/JD/MD, CS or engineering) and a "top school" indicator for universities in the top 20 for producing VC-backed founders (Cook, Marx and Yimfor, 2025). Experience controls include years of work experience,

[3]Of the U.S. companies that appear in Diversity Spotlight with Black leadership and have founder records in Crunchbase, 100% are captured in my sample. The remaining Diversity Spotlight registrants either lack founder data in Crunchbase or are investor organizations rather than startups.

senior roles (director or above), and prior startups founded. I infer age from undergraduate graduation dates. I winsorize all continuous variables at the 1st and 99th percentiles.

3. Summary Statistics

Of the 8,564 Black founders in my sample, 1,077 (12.6%) appear in Diversity Spotlight and 7,487 (87.4%) do not. Table 1 compares these two groups.

[Insert Table 1 About Here.]

Panel A reports founder characteristics from LinkedIn profiles. On some dimensions, the two groups look similar: both average about 15 years of work experience and have similar rates of master's degrees (18%). But significant differences emerge on markers of elite credentials. Self-identified founders are more likely to have bachelor's degrees (75% vs. 66%), MBAs (18% vs. 13%), and CS or engineering degrees (25% vs. 18%). The gap in top school attendance is particularly stark: 33% of self-identified founders attended elite universities versus just 19% of non-self-identified founders. Self-identified founders also have more professional experience, holding more senior roles (1.65 vs. 1.39) and having founded more prior startups (0.53 vs. 0.32). Interestingly, self-identified founders are younger on average (39.7 vs. 42.3 years), suggesting that selection into self-identification is not simply about having more time to accumulate credentials.

Panel B reports startup outcomes. Self-identified founders are three times more likely to have raised any funding (73% vs. 25%). Conditional on raising, they secured six times more capital ($11.7 million vs. $1.9 million) over nearly five times as many rounds (3.5 vs. 0.7). Self-identified founders are five times more likely to reach late-stage funding (11% vs. 2%). Exit rates are also higher for self-identified founders (11% vs. 4% for any exit), though the gap in good exits is smaller. Whether these differences survive multivariate controls is the subject of the next sections.

4. Empirical Strategy

All analyses estimate descriptive associations using OLS. The goal is to document how conclusions about funding gaps differ when using algorithmic classification versus self-reported data, and to characterize the selection process into self-identification.

4.1. Funding Gap Regressions

The first set of analyses examines funding gaps at the startup level. I estimate:

$$y_i = \alpha + \beta \cdot \text{P(Black)}_i + X_i'\gamma + \delta_s + \tau_t + \eta_j + \varepsilon_i \tag{1}$$

where i indexes startups, y_i is the funding outcome (either an indicator for raising any VC or log total funding conditional on raising), P(Black)_i is the proportion of founders who are Black (ranging from 0 to 1), X_i is a vector of founder controls, and δ_s, τ_t, η_j are fixed effects for state, founding year, and industry respectively.

The unit of observation is the startup ($N = 174,347$ for the extensive margin; $N = 46,431$ for the intensive margin, restricted to startups that raised funding). The key coefficient β measures the association between having Black founders and funding outcomes, conditional on founder characteristics and fixed effects. For the extensive margin (linear probability model), moving from no Black founders to all Black founders (P(Black) = 0 to P(Black) = 1) is associated with a $100 \times \beta$ percentage point change in the probability of raising VC. For the intensive margin (log funding), moving from no Black founders to all Black founders is associated with a $100 \times (e^{\beta} - 1)$% change in funding raised.

Founder controls include education (bachelor's degree, master's degree, MBA, PhD/JD/MD, CS or engineering degree, top school attendance) and experience (years of experience, senior roles, prior startups). Standard errors are clustered at the startup level.

I run this specification twice: first using algorithmic classification to identify Black founders, then using only self-reported Diversity Spotlight data. Comparing the two estimates reveals how selection into self-reporting is related to conclusions about funding gaps.

4.2. Exit Outcome Regressions

The second set of analyses examines exit outcomes among Black-founded startups only. I estimate:

$$\text{Exit}_i = \alpha + \beta \cdot \mathbb{1}[\text{Self-ID}]_i + \phi \cdot \ln(\text{Raised})_i + X_i'\gamma + \delta_s + \tau_t + \eta_j + \varepsilon_i \tag{2}$$

where the sample is restricted to startups with at least one Black founder identified via algorithmic classification ($N = 8,717$). The dependent variable Exit_i is an indicator for achieving a good exit (IPO or acquisition $\geq 2\times$ funding), any exit, or IPO. The key regressor $\mathbb{1}[\text{Self-ID}]_i$ indicates whether the startup appears in Diversity Spotlight with Black leadership indicated. Controlling for $\ln(\text{Raised})_i$ isolates whether self-identified founders outperform conditional on funding raised.

The coefficient β measures the association between self-identification and exit outcomes. A positive β would suggest self-identified founders are higher quality; a negative β (conditional on funding) would suggest they are overfunded relative to outcomes.

4.3. Selection into Self-Identification

The third set of analyses examines what predicts appearing in Diversity Spotlight. I estimate:

$$\mathbb{1}[\text{Self-ID}]_i = \alpha + \beta \cdot \mathbb{1}[\text{Launch Partner}]_i + X_i'\gamma + \delta_s + \tau_t + \eta_j + \varepsilon_i \tag{3}$$

where $\mathbb{1}[\text{Launch Partner}]_i$ indicates whether the startup received investment from a Diversity Spotlight launch partner (Backstage Capital, Harlem Capital Partners, BLCK VC, All Raise, or Precursor Ventures). The sample is again restricted to Black-founded startups ($N = 8,717$).

The coefficient β measures the association between launch partner backing and self-identification in percentage points. A large positive β that is stable across specifications would suggest the institutional origins of Diversity Spotlight drive selection into the database.

4.4. Interpretation

These regressions document associations, not causal effects. The funding gap estimates in Equation 1 should be interpreted as conditional correlations: Black founders raise less capital than observably similar non-Black founders, but unobserved differences (e.g., business quality, networks, discrimination) could drive this gap. The exit regressions in Equation 2 condition on funding raised, which is itself endogenous. The selection regressions in Equation 3 document a strong correlation but do not establish that launch partner backing *causes* self-identification.

The key contribution is comparative: showing that conclusions differ dramatically depending on whether race is measured via algorithmic classification or self-reported data, and documenting the selection process that generates this divergence.

5. Results

5.1. The Funding Gap

Figure 1 presents the central finding. Black founders receive 2.4% of total venture capital funding—more than twice the 0.9% captured by self-reported Diversity Spotlight data. This 2.4% figure is consistent with media reports, industry surveys, and previous academic work that have documented Black founders' underrepresentation relative to their 13% population share (Cook, Marx and Yimfor, 2025). Self-reported data, by contrast, suggest Black founders receive less than 1% of VC funding, understating the true share by more than 60%.

[Insert Figure 1 About Here.]

The gap between algorithmic and self-reported estimates varies by funding stage. At seed stage, Black founders receive 3.8% of funding—their highest share—but self-reported data capture only 1.4%. At later stages, the undercount worsens: self-reported data capture just 26% of Black founder funding at Series D and beyond (0.6% vs. 2.4%). The systematic undercount reflects the 87% of Black founders invisible to self-reported data.

5.2. Time-Series Variation

Figure 2 plots Black founders' share of VC funding annually from 2015 to 2024. The gap between algorithmic classification and self-reported data is substantial in every year, but the time series reveals important dynamics.

[Insert Figure 2 About Here.]

Black founders' funding share spiked dramatically in 2020, the year of George Floyd's murder, reaching 9% by algorithmic classification. This spike was short-lived: by 2021, the share had fallen to under 2%, and it has remained between 1–2% through 2024. Self-reported data show a much more muted pattern, never exceeding 1% even in 2020.

5.3. Geographic Variation

Figure 3 compares Black founder funding shares across the top 10 states by VC activity. Texas leads with 3.9% of VC funding going to Black founders, followed by New York (3.6%) and Illinois (3.1%). California, despite being the largest VC market, is fourth at (2.5%).

[Insert Figure 3 About Here.]

The undercount in self-reported data is severe across all geographies. In Texas, self-reported data capture only 3% of Black founder funding (0.1% vs. 3.9%). The geographic pattern suggests that self-reported data are most incomplete outside coastal VC hubs, precisely where Black founders are most likely to operate.

5.4. Industry Variation

Figure 4 presents funding shares by industry. Black founders receive their highest share in AI and machine learning (4.4%) and apps (4.3%), followed by hardware (1.9%), commerce and shopping (1.8%), and financial services (1.5%). The lowest shares are in consumer electronics (0.3%) and energy (0.7%).

[Insert Figure 4 About Here.]

Self-reported data systematically undercount Black founder funding across all industries. The gap is particularly large in biotech, where self-reported data capture only 8% of Black founder funding (0.1% vs. 1.3%), and in hardware (0.2% vs. 1.9%).

5.5. Regression Evidence: Algorithmic Classification vs. Self-Reported Data

The preceding figures documented the extensive margin: Black founders' share of total VC funding. We now turn to the intensive margin, examining funding gaps conditional on founder race using multivariate regressions that compare algorithmic classification with self-reported data.

Table 2 presents regression estimates of Equation 1.

[Insert Table 2 About Here.]

Columns (1) and (2) use algorithmic classification to identify Black founders. Column (1) examines the extensive margin: whether the startup raised any VC funding. Moving from zero to 100% Black founders reduces the probability of raising VC by 8.7 percentage points, a 33% reduction relative to the baseline funding rate of 26%. Column (2) examines the intensive margin: log funding conditional on raising at least $1. The coefficient of −1.205 implies that Black-founded startups raise approximately 70% less than comparable non-Black startups ($100 \times (e^{-1.205} - 1) = -70\%$), controlling for founder characteristics and fixed effects.

Columns (3) and (4) repeat the analysis using only Diversity Spotlight data, treating startups as Black-founded only if they appear in Diversity Spotlight with Black leadership indicated. The results are strikingly different. The extensive margin coefficient *flips sign*, from −8.7 percentage points to +19.3 percentage points. Using self-reported data, a researcher would conclude that Black founders are *more* likely to raise funding than non-Black founders. The intensive margin gap shrinks from 70% to 23%, a three-fold reduction.

5.6. Funding Gaps by Stage

Table 3 estimates Equation 1 separately by funding stage: seed, early (Series A–B), and later (Series C+). The comparison between algorithmic classification and self-reported data is even more stark when examined stage by stage.

Panel A uses algorithmic classification data. Black founders face significant gaps at every stage on both margins. On the extensive margin, Black founders are 1.9 percentage points less likely to raise seed funding, 5.6 percentage points less likely to raise early-stage funding, and 1.4 percentage points less likely to raise late-stage funding. On the intensive margin, the log coefficients of −0.861 and −0.444 imply they receive approximately 58% less at seed ($100\times(e^{-0.861}-1) = -58\%$) and 36% less at early stage ($100\times(e^{-0.444}-1) = -36\%$). The late-stage intensive margin gap is small and insignificant, suggesting that conditional on reaching late-stage, Black founders raise similar amounts. But reaching late-stage is itself substantially less likely.

[Insert Table 3 About Here.]

Panel B uses self-reported data only. The pattern reverses at seed and early stages. Self-identified founders appear 29 percentage points *more* likely to raise seed funding and 4.7 percentage points more likely to raise early-stage funding. Only at the intensive margin do gaps persist, and even these are substantially attenuated: the seed-stage intensive margin gap shrinks from 58% to 20%, and the early-stage gap shrinks from 36% to 24%. At late-stage, both extensive and intensive margin effects are small and insignificant with self-reported data.

The stage-by-stage comparison reinforces the central finding: self-reported data systematically misrepresent Black founders' experience in venture capital. The bias is largest at early stages, where the extensive margin coefficient not only shrinks but reverses sign entirely.

5.7. Exit Outcomes: Are Self-Identified Founders Higher Quality?

The previous results establish that self-identified founders are more successful at raising capital. But are they genuinely higher quality, or simply better networked and more visible? If self-identified founders are truly stronger, they should outperform conditional on funding. If they are merely better connected, their outcomes should be no better, or perhaps worse, than other Black founders who raise similar amounts.

Table 4 presents estimates of Equation 2. The sample is restricted to startups with at least one Black founder identified via algorithmic classification ($N = 8,717$), and the key explanatory variable is an indicator for whether the startup appears in Diversity Spotlight. This within-Black-founder comparison isolates the association between self-identification and outcomes, holding race constant.

[Insert Table 4 About Here.]

Columns (1) and (3) do not control for funding. Without this control, self-identified founders show slightly higher exit rates, but the differences are small and insignificant.

Columns (2), (4), and (5) control for log total funding raised. The pattern reverses sharply. Conditional on funding, self-identified founders are 1.8 percentage points *less* likely to achieve a good exit, 2.2 percentage points less likely to achieve any exit, and 1.5 percentage points less likely to IPO. These effects are economically large relative to base rates: the 1.8 percentage point reduction in good exits is more than twice the 0.78% base rate for good exits among Black founders who raised funding.

The finding has a striking implication: self-identified Black founders are overfunded relative to their outcomes. Investors pour more capital into founders who are visible and well-credentialed, but these founders do not deliver better results per dollar invested. Self-reported data therefore not only misrepresent the size of funding gaps but also mischaracterize which Black founders are underfunded.

5.8. The George Floyd Response

George Floyd's murder in May 2020 prompted unprecedented attention to racial inequality, including in venture capital. Emmanuel Yimfor, Matt Marx and Qian Wang (2025) document that VCs responded swiftly to the social movement, with investment in Black-founded startups increasing substantially in the immediate aftermath. However, this response proved temporary: funding reverted to prior levels within two years, and the increase was concentrated among investors who had never previously backed a Black founder. These "newcomer" investors typically made only one investment, were less likely to take board seats, and failed to attract the highest-quality Black entrepreneurs, who appeared to interpret the response as tokenism.

The data in this paper reveal a complementary pattern. Self-identified founders captured 52% of post-George Floyd funding (June 2020–December 2022) despite comprising only 12.6% of Black founders. This 4-fold overrepresentation suggests that diversity-motivated capital flowed disproportionately to founders who had already made their race known to investors.

These findings add a new dimension to the theoretical framework of Stephen Coate and Glenn C Loury (1993), who show that diversity initiatives can backfire by confirming rather than dispelling negative stereotypes. In their model, the question is whether initiatives change employers' beliefs about group productivity. Selection into self-reported data introduces an additional channel: if diversity-motivated capital flows to founders who are visible in the data, and if those founders systematically underperform relative to funding (as the exit analysis suggests), then investors observing poor outcomes may conclude that Black founders are weak investments. Yet the 87% of Black founders invisible to self-reported data might deliver comparable or better returns per dollar. Selection into data may thus determine not just who receives diversity-motivated funding, but whether such initiatives ultimately reinforce or correct negative stereotypes.

5.9. Why Do Some Black Founders Self-Identify?

The preceding analysis establishes that self-identified founders differ systematically from other Black founders: they have stronger credentials, raise more funding, but deliver worse outcomes per dollar invested. What explains selection into self-reported data? One possibility is founder characteristics. Perhaps more credentialed founders are more likely to register. But the stability of outcome differences after controlling for observables suggests something else may be at work.

Figure 5 explores whether investor networks drive selection. The figure plots each investor's share of deals going to self-identified Black founders (x-axis) against their share going to non-self-identified Black founders (y-axis). Investors above the 45-degree line fund more "invisible" Black founders; those below fund more self-identified founders.

[Insert Figure 5 About Here.]

The figure reveals a striking pattern: many of the investors clustering below the 45-degree line, those whose Black founder portfolios skew toward self-identified founders, are themselves Black-led VC firms. This suggests that selection into Diversity Spotlight may be driven by investor networks rather than founder characteristics alone.

To investigate systematically, I use the list of 113 self-identified Black-led VC firms from Diversity Spotlight, which overlaps substantially with the minority investor data from Cassel, Lerner and Yimfor (2023). Table 5 reports the investment activity of the 25 most active of these firms. Even among investors who self-identified as Black-led, only 9.5% of their Black founder deals go to startups that also appear in Diversity Spotlight. The remaining 90.5% are "invisible" Black founders. But there is substantial variation across investors: Backstage Capital, a Diversity Spotlight launch partner, has the highest self-ID rate at 47.4%, while firms like GE Ventures and Riverwood Capital have close to 0% despite investing in over 100 Black-founded startups each. What explains this variation?

[Insert Table 5 About Here.]

The concentration among launch partners suggests a network-driven selection mechanism. When Crunchbase launched Diversity Spotlight in August 2020, the program explicitly partnered with five diversity-focused VCs: Backstage Capital, Harlem Capital Partners, BLCK VC, All Raise, and Precursor Ventures. These partners likely encouraged their portfolio companies to register, creating an initial sample disproportionately drawn from their networks.

Table 6 tests this directly. Column (1) shows that launch partner backing is associated with a 67.5 percentage point increase in the probability of appearing in Diversity Spotlight. Relative to the baseline self-identification rate of 11.9%, this represents roughly a six-fold increase in visibility. But is this effect driven by launch partners specifically, or by diversity-focused investors more broadly? Columns (3)–(5) add a control for backing by any of the 113 self-identified Black-led investors. The launch partner effect drops to 24.6 percentage points but remains highly significant; the diversity investor effect is 45.6 percentage points. Both effects survive controls for founder characteristics and funding. The results suggest that while diversity-focused investor networks broadly predict self-identification, the launch partners retain an independent effect, consistent with their direct role in seeding the Diversity Spotlight database.

[Insert Table 6 About Here.]

6. Limitations

A key strength of this analysis is that it nests self-reported data within comprehensive data. The sample includes 100% of the startups that self-identified as Black-founded in Diversity Spotlight. The comparison between self-reported and algorithmic data is therefore not a comparison of two different samples; it is a comparison between a selected subset and the full population from which that subset was drawn. When I estimate funding gaps using "self-reported data," I am using the exact same Diversity Spotlight data that a researcher relying on voluntary disclosure would use. The difference is that I can also identify the 7,487 Black-founded startups that do not appear in Diversity Spotlight, allowing direct measurement of selection bias.

Several limitations remain. First, the algorithmic classification methodology, while validated against perceived race in the Chicago Face Database, may contain classification errors. Light-skinned Black founders may be misclassified as non-Black, and some ambiguous cases may be resolved incorrectly despite manual review. To the extent that misclassification is random, it would attenuate estimated funding gaps; systematic misclassification could bias results in either direction. Second, the sample is restricted to founders with identifiable profile photos, excluding approximately 8% of the Crunchbase founder universe. Third, the outcome analysis conditions on funding raised, which is itself endogenous to founder characteristics and investor decisions. Fourth, Crunchbase coverage, while comprehensive for venture-backed startups, may underrepresent bootstrapped companies and those outside traditional VC networks. Finally, the analysis focuses on Black founders and may not generalize to other demographic groups where name-based algorithms perform differently or where self-reporting patterns differ.

7. Conclusion

Self-reported diversity data miss 87% of Black founders, and the founders who do self-report are not representative. They have stronger credentials, higher funding rates, and, paradoxically, worse outcomes per dollar invested than those who remain invisible. Research relying on such data will reach systematically incorrect conclusions about funding disparities for Black entrepreneurs.

The magnitude of the bias is substantial. Using comprehensive data from algorithmic classification, I find that Black founders receive 2.4% of venture capital funding; self-reported data capture only 0.9%. The true funding gap is also larger: an 8.7 percentage point reduction in funding probability and 70% less funding conditional on raising, compared to a *positive* extensive margin coefficient and only a 23% intensive margin gap when using self-reported data. At the stage level, self-reported data not only understate the gap but reverse its sign entirely at seed and early stages.

Perhaps most surprising, self-identified founders are overfunded relative to their outcomes. Conditional on funding raised, they are less likely to achieve good exits, any exits, or IPOs than other Black founders. This pattern suggests that visibility and credentials, the characteristics that predict self-identification, are imperfect signals of founder quality.

These findings have implications for diversity initiatives in venture capital. The funding surge following George Floyd's murder concentrated among already-visible founders: 52% of post-George Floyd funding went to the 12.6% of Black founders who appear in Diversity Spotlight. If diversity-motivated capital flows primarily to founders who are already visible and already overfunded relative to outcomes, such initiatives may reinforce existing disparities rather than expanding opportunity. The 87% of Black founders invisible to self-reported data may be precisely those who would benefit most from additional capital.

Why does self-reported data exhibit such severe selection bias? The institutional origins of Diversity Spotlight explain much of the pattern. Black founders backed by diversity-focused investors, particularly the five launch partners who seeded the database, are far more likely to appear. This network-driven selection suggests that expanding diversity initiatives beyond existing networks may be necessary to reach the invisible majority.

To support future research, I provide two datasets. The Black Founder Dataset, available at https://github.com/eyimfor/race_classifier_fbhgs, contains founder-level information for U.S.-based startups. A companion dataset on minority venture capital partners is available at https://www.eyimfor.com/data/MinorityGroupList.xlsx. I hope these resources enable more accurate measurement of funding disparities and more effective interventions to address them.

References

Barkley, Brett, and Mark E Schweitzer. 2022. "Credit Availability for Minority Business Owners in an Evolving Credit Environment: Before and During the COVID-19 Pandemic." Federal Reserve Bank of Cleveland Working Paper 22-18.

Cassel, Johan, James Weston, and Emmanuel Yimfor. 2023. "Venturing into Racial Diversity on Startup Boards." Columbia Business School Research Paper 4622698.

Cassel, Johan, Josh Lerner, and Emmanuel Yimfor. 2023. "Racial Diversity in Private Capital Fundraising." Working Paper.

Coate, Stephen, and Glenn C Loury. 1993. "Will affirmative-action policies eliminate negative stereotypes?" *American Economic Review*, 83(5): 1220–1240.

Cook, Lisa D, Matt Marx, and Emmanuel Yimfor. 2025. "Funding Black High-Growth Startups." Forthcoming, Journal of Finance.

Fairlie, Robert, Alicia Robb, and David T Robinson. 2022. "Black and white: Access to capital among minority-owned start-ups." *Management Science*, 68(4): 2377–2400.

Fairlie, Robert, and David T Robinson. 2023. "Racial Differences in Access to Capital for Innovative Start-Ups." *Entrepreneurship and Innovation Policy and the Economy*, 2(1): 149–166.

García, Raffi E, and William A Darity Jr. 2022. "Self-Reporting Race in Small Business Loans: A Game-Theoretic Analysis of Evidence from PPP Loans in Durham, NC." Vol. 112, 299–302.

Greenwald, Daniel L, Sabrina T Howell, Cangyuan Li, and Emmanuel Yimfor. 2024. "Regulatory Arbitrage or Random Errors? Implications of Race Prediction Algorithms in Fair Lending Analysis." *Journal of Financial Economics*, 157: 103857.

Gupta, Vishal K, Daniel B Turban, and Nachiket M Bhawe. 2008. "The effect of gender stereotype activation on entrepreneurial intentions." *Journal of applied psychology*, 93(5): 1053.

Ma, Debbie S, Joshua Correll, and Bernd Wittenbrink. 2015. "The Chicago face database: A free stimulus set of faces and norming data." *Behavior research methods*, 47(4): 1122–1135.

Marx, Matt, Qian Wang, and Emmanuel Yimfor. 2025. "Minimum viable signal: Venture funding, social movements, and race." *Management Science*.

Morazzoni, Marta, and Andrea Sy. 2022. "Female Entrepreneurship, Financial Frictions and Capital Misallocation in the US." *Journal of Monetary Economics*, 129: 93–118.

Retterath, Andre, and Reiner Braun. 2020. "Benchmarking venture capital databases." *Available at SSRN 3706108*.

Serengil, Sefik Ilkin, and Alper Ozpinar. 2020. "LightFace: A Hybrid Deep Face Recognition Framework." 23–27, IEEE.

Wojan, Timothy R. 2024. "Exploratory Report: Annual Business Survey Ownership Diversity and Its Association with Patenting and Venture Capital Success." U.S. Census Bureau, Center for Economic Studies Working Paper 24-62.

Yimfor, Emmanuel, Matt Marx, and Qian Wang. 2025. "Minimum Viable Signal: Venture Funding, Social Movements, and Race." *Management Science*.

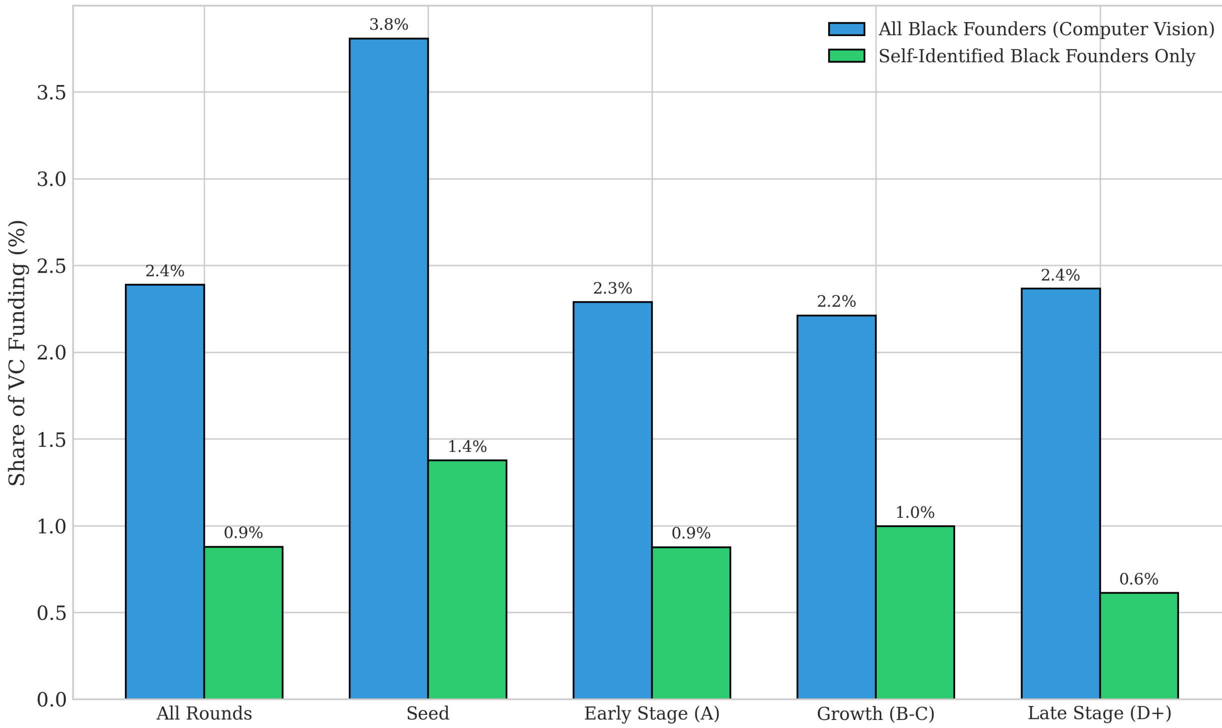

Figure 1: **Share of Venture Funding to Black Founders by Stage**

This figure compares the share of venture capital funding raised by Black-founded companies across funding stages. Blue bars show Black founders identified via algorithmic classification; green bars show only those at companies that self-reported to Crunchbase Diversity Spotlight. The sample includes U.S. VC deals from 2015–2024.

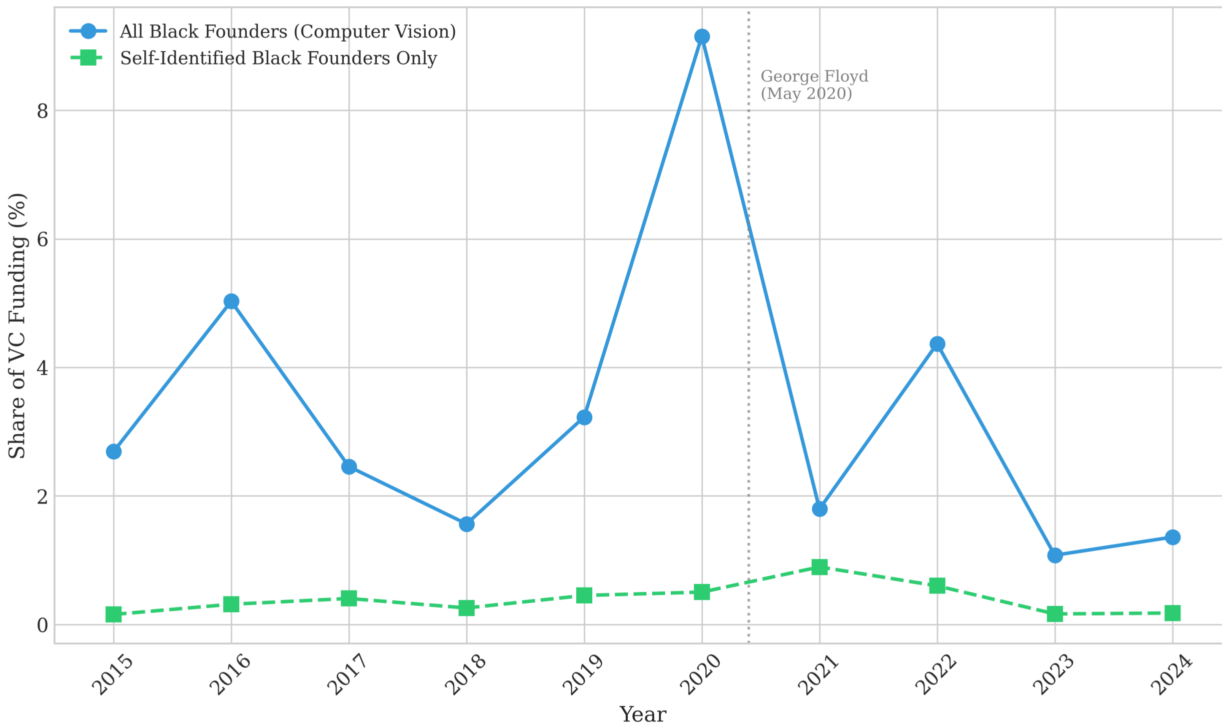

Figure 2: **Share of Venture Funding to Black Founders Over Time**

This figure plots the annual share of venture capital funding raised by Black-founded companies from 2015–2024. The blue line shows Black founders identified via algorithmic classification; the green line shows only those at companies that self-reported to Crunchbase Diversity Spotlight. The vertical dashed line marks George Floyd's murder in May 2020.

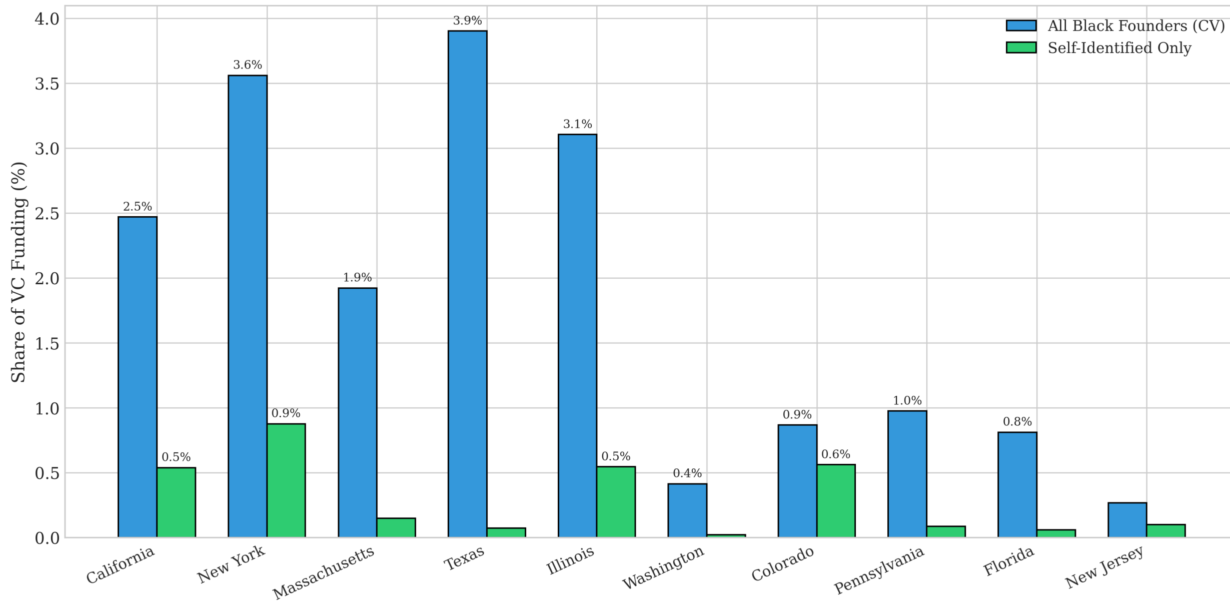

Figure 3: **Share of Venture Funding to Black Founders by State**

This figure compares the share of venture capital funding raised by Black-founded companies across the top 10 states by VC activity. Blue bars show Black founders identified via algorithmic classification; green bars show only those at companies that self-reported to Crunchbase Diversity Spotlight.

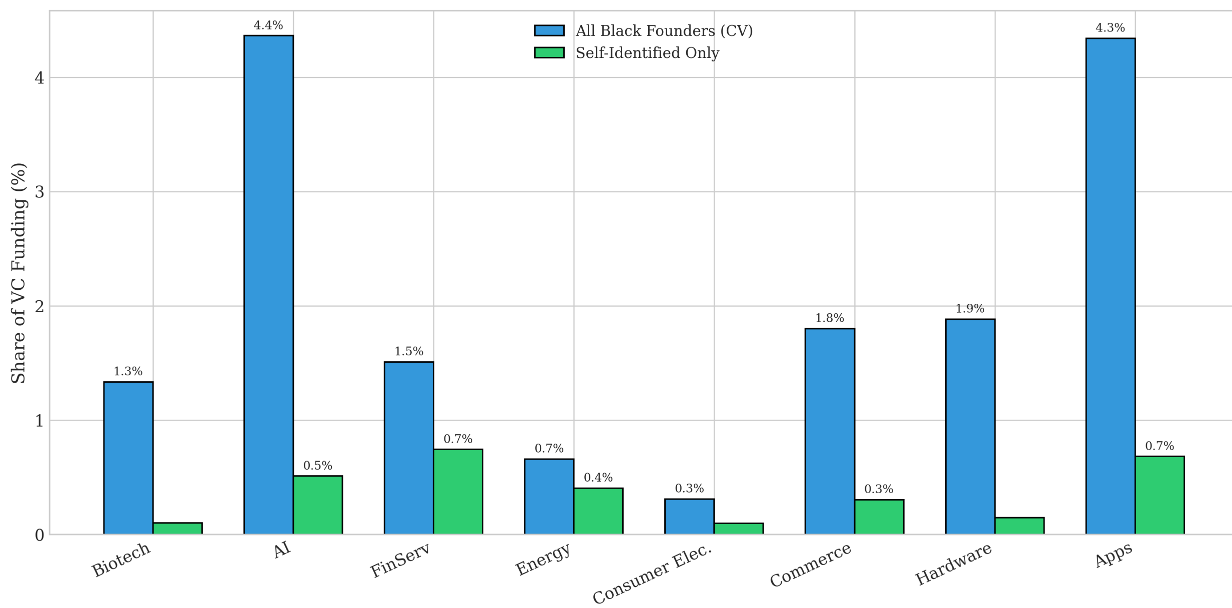

Figure 4: **Share of Venture Funding to Black Founders by Industry**

This figure compares the share of venture capital funding raised by Black-founded companies across major industry categories. Blue bars show Black founders identified via algorithmic classification; green bars show only those at companies that self-reported to Crunchbase Diversity Spotlight.

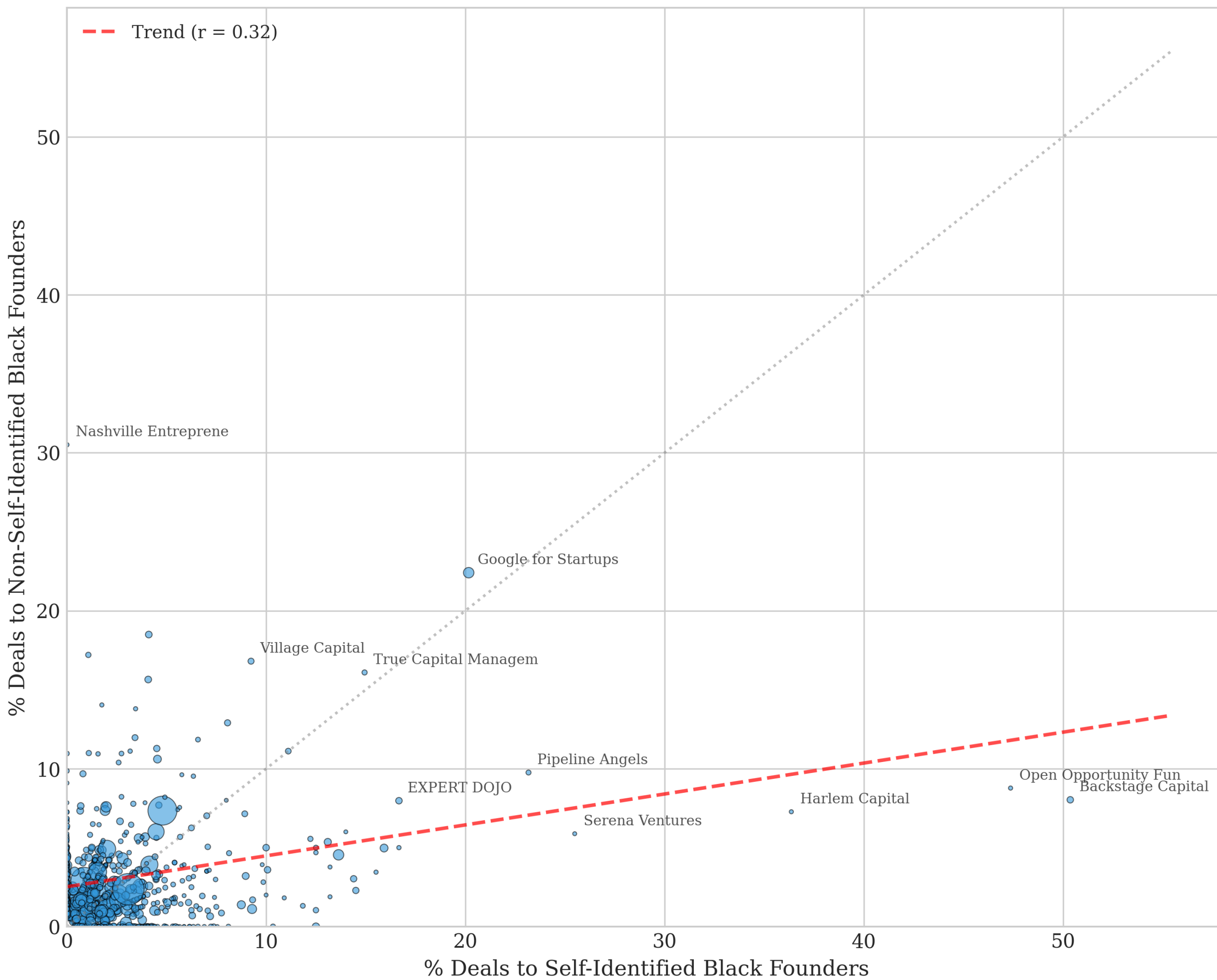

Figure 5: **Investor Backing of Self-Identified vs. Non-Self-Identified Black Founders**

This scatter plot shows the correlation between an investor's share of deals going to self-identified Black founders (x-axis) and share going to non-self-identified Black founders (y-axis). Each point represents an investor, with point size reflecting deal volume. The sample is restricted to investors with at least 10 deals.

Table 1: **Summary Statistics: Black Founders by Self-Identification Status**

This table compares characteristics and outcomes of Black founders identified via algorithmic classification who are at companies that self-reported to Crunchbase Diversity Spotlight (Self-Identified) versus those at companies that did not (Non-Self-Identified). The sample includes 8,564 Black founders identified through image classification using DeepFace combined with clerical review; approximately 75% were identified algorithmically and 25% through manual review for false negatives. Self-identified founders represent 12.6% of all Black founders, while 87.4% are "invisible" to self-reported diversity data. *Top School* indicates attendance at a university in the top 20 for producing VC-backed founders. *Good Exit* is defined as an IPO or an acquisition where the acquisition price is at least twice the total funding raised. *Any Exit* is defined as an IPO or any acquisition. *Post-GF Raised* is funding raised between June 2020 and December 2022. Dollar amounts in millions. All continuous variables are winsorized at the 1% and 99% levels. t-statistics test the null hypothesis of equal means; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

	Self-Identified (Diversity Spotlight)		**Non-Self-Identified ("Invisible")**		
	N = 1,077		**N = 7,487**		
	Mean	Std. Dev.	Mean	Std. Dev.	t-stat
Panel A. Founder Characteristics (LinkedIn)					
Years of Work Experience	15.17	8.70	15.04	8.39	0.24
I(Bachelor's Degree)	0.75	0.43	0.66	0.47	5.36***
I(Master's Degree)	0.18	0.38	0.18	0.38	0.12
I(MBA)	0.18	0.38	0.13	0.34	3.71***
I(PhD/JD/MD)	0.05	0.22	0.04	0.19	1.25
I(CS or Engineering Degree)	0.25	0.43	0.18	0.38	4.69***
I(Attended Top School)	0.33	0.47	0.19	0.40	8.66***
Senior Roles	1.65	1.90	1.39	1.61	4.06***
Prior Startups	0.53	0.89	0.32	0.68	7.05***
Age	39.74	10.69	42.30	11.53	−5.99***
Panel B. Startup Outcomes					
I(Raised Any Funding)	0.73	0.45	0.25	0.43	30.45***
Total Raised ($ Millions)	11.73	23.89	1.94	8.74	17.23***
Number of Funding Rounds	3.52	5.64	0.74	2.02	19.52***
I(Late Stage: Series B+)	0.11	0.31	0.02	0.12	13.67***
Post-GF Raised ($ M)	4.68	12.34	0.63	4.69	10.66***
I(Good Exit)	0.01	0.11	0.01	0.08	2.01**
I(Any Exit)	0.11	0.31	0.04	0.20	7.86***

Table 2: **Association between Black Ownership and Likelihood and Amount of VC Funding**

The unit of observation is the startup and standard errors are clustered at the startup level. Columns (1)–(2) use algorithmic classification to identify Black founders; Columns (3)–(4) use self-reported Crunchbase Diversity Spotlight data. The dependent variable in Columns (1) and (3) is an indicator for raising any VC funding; in Columns (2) and (4) it is ln(total VC funding) conditional on raising. *P(Black)* is the proportion of founders/CEOs who are Black (0 to 1). *Founder Controls* include education and experience measures. All regressions include state, founding year, and industry fixed effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Dependent Variable:	**I(Raised VC)**	**Ln(VC Funding)**	**I(Raised VC)**	**Ln(VC Funding)**
Sample:	*Algorithmic Classification*		*Self-ID Only*	
	(1)	(2)	(3)	(4)
P(Black)	-0.087***	-1.205***	0.193***	-0.266**
	(0.005)	(0.075)	(0.020)	(0.117)
Observations	174,347	46,431	174,347	46,431
Y_{mean}	0.261	15.21	0.261	15.21
Adjusted-R^2	0.10	0.12	0.10	0.12
State, Year, and Industry FE?	Yes	Yes	Yes	Yes
Founder Controls?	Yes	Yes	Yes	Yes

Table 3: **Association between Black Ownership and Funding by Stage**

The unit of observation is the startup and standard errors are clustered at the startup level. Panel A uses algorithmic classification to identify Black founders; Panel B uses self-reported Crunchbase Diversity Spotlight data. The dependent variable in odd columns is an indicator for raising funding at that stage; in even columns it is ln(funding) conditional on raising. Seed = pre-Series A; Early = Series A–B; Later = Series C+. *P(Black)* is the proportion of founders/CEOs who are Black (0 to 1). All regressions include founder controls, state, founding year, and industry fixed effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Dependent Variable:	**I(Raised Seed)**	**Ln(Seed Funding)**	**I(Raised Early)**	**Ln(Early Funding)**	**I(Raised Later)**	**Ln(Later Funding)**
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Algorithmic Classification			
P(Black)	-0.019***	-0.861***	-0.056***	-0.444***	-0.014***	0.037
	(0.005)	(0.052)	(0.002)	(0.102)	(0.001)	(0.219)
			Panel B: Self-ID Data Only			
P(Black)	0.290***	-0.198**	0.047***	-0.275**	0.001	-0.049
	(0.018)	(0.089)	(0.012)	(0.115)	(0.005)	(0.346)
Observations	194,193	37,903	194,193	20,104	194,193	5,477
Y_{mean}	0.195	13.64	0.104	16.41	0.028	17.66
State, Year, and Industry FE?	Yes	Yes	Yes	Yes	Yes	Yes

Table 4: **Exit Outcomes: Self-Identified vs. Non-Self-Identified Black Founders**

The unit of observation is a startup with at least one Black founder (identified via algorithmic classification) and standard errors are clustered at the startup level. *Good Exit* is an IPO or acquisition at ≥2x total funding raised. *Any Exit* includes all IPOs and acquisitions. *IPO* is an indicator for going public. *I(Self-Identified)* equals one if the startup appears in Crunchbase Diversity Spotlight. Columns (1) and (3) exclude funding controls; Columns (2), (4), and (5) control for ln(total raised). All regressions include state, founding year, and industry fixed effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Dependent Variable:	**Good Exit**	**Good Exit**	**Any Exit**	**Any Exit**	**IPO**
	(1)	(2)	(3)	(4)	(5)
I(Self-Identified)	0.002	-0.018***	0.001	-0.022***	-0.015***
	(0.003)	(0.007)	(0.004)	(0.007)	(0.005)
Ln(Total Raised)		0.012***		0.015***	0.015***
		(0.002)		(0.002)	(0.002)
Observations	8,723	2,697	8,723	2,697	2,697
Y_{mean}	0.0078	0.0078	0.0125	0.0125	0.0039
Adjusted-R^2	0.022	0.137	0.027	0.170	0.077
State, Year, and Industry FE?	Yes	Yes	Yes	Yes	Yes
Controls for Funding?	No	Yes	No	Yes	Yes

Table 5: **Investment Activity of Self-Identified Black-Led VC Firms**

The unit of observation is an investor-startup pair. This table reports investment activity by the 25 most active self-identified Black-led VC firms in Black-founded startups, ranked by number of Black-founded deals. A total of 113 investment firms self-reported Black leadership to Crunchbase Diversity Spotlight; the bottom row reports totals across all 113. *Total Deals* is the number of unique portfolio companies. *Black Founders* counts deals with startups that have at least one Black founder identified via algorithmic classification. *Self-ID* counts the subset that also appear in Diversity Spotlight. *Non-Self-ID* counts Black-founded startups not in Diversity Spotlight. *% Self-ID* is the share of Black founder deals going to self-identified startups.

		Black-Founded Deals			
Investor	**Total Deals**	**Total**	**Self-ID**	**Non-Self-ID**	**% Self-ID**
General Catalyst	751	658	6	652	0.9%
Precursor Ventures	352	285	35	250	12.3%
Ulu Ventures	208	187	10	177	5.3%
Pear VC	219	184	2	182	1.1%
Kapor Capital	188	172	25	147	14.5%
MaC Venture Capital	182	163	23	140	14.1%
Backstage Capital	144	133	63	70	47.4%
MAGIC Fund	157	127	4	123	3.1%
Rough Draft Ventures	140	127	7	120	5.5%
R/GA Ventures	119	108	11	97	10.2%
WndrCo	122	107	4	103	3.7%
GE Ventures	111	101	0	101	0.0%
Jumpstart Foundry	119	99	5	94	5.1%
Base Ventures	90	86	9	77	10.5%
Third Sphere	93	84	4	80	4.8%
Reach Capital	95	80	1	79	1.2%
Golden Seeds	85	79	2	77	2.5%
Sinai Capital Partners	80	77	7	70	9.1%
645 Ventures	81	72	3	69	4.2%
LoftyInc Capital	83	64	7	57	10.9%
Base10 Partners	82	64	3	61	4.7%
Riverwood Capital	68	62	0	62	0.0%
Cleveland Avenue	70	60	11	49	18.3%
Harlem Capital Partners	64	59	18	41	30.5%
Serena Ventures	60	54	12	42	22.2%
Total (All 113 Investors)	4,937	4,305	410	3,895	9.5%

Table 6: **Selection into Diversity Spotlight: Launch Partners vs. Diversity-Focused Investors**

The unit of observation is a Black-founded startup (identified via algorithmic classification) and standard errors are robust. The dependent variable is an indicator for appearing in Crunchbase Diversity Spotlight. *Launch Partner Backed* indicates investment from one of the five official Diversity Spotlight launch partners (Backstage Capital, Harlem Capital Partners, BLCK VC, All Raise, Precursor Ventures). *Diversity Investor Backed* indicates investment from any of the 113 investors that self-reported Black leadership to Diversity Spotlight. Column (5) restricts to funded startups and controls for log total funding raised. All regressions include state, founding year, and industry fixed effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Dependent Variable:	**I(Self-Identified in Diversity Spotlight)**				
Sample:	*All Black-Founded Startups*				*Funded Only*
	(1)	(2)	(3)	(4)	(5)
I(Launch Partner Backed)	0.675***	0.667***	0.246***	0.251***	0.255***
	(0.031)	(0.031)	(0.044)	(0.043)	(0.046)
I(Diversity Investor Backed)			0.456***	0.443***	0.273***
			(0.032)	(0.032)	(0.035)
Observations	6,607	6,607	6,607	6,607	2,574
State, Year, Industry FE	Yes	Yes	Yes	Yes	Yes
Founder Controls	No	Yes	No	Yes	Yes
Funding Control	No	No	No	No	Yes