Funding Black Innovators

Preliminary Draft – Please do not distribute without authors' permission

Lisa D. Cook[c,b], Steven Wu-Chaves[a], Emmanuel Yimfor[d]

[a] *Michigan State University*
[b] *NBER*
[c] *Federal Reserve Board*
[d] *University of Michigan*

July 17, 2022

Abstract

In the last decade, we have learned that high-growth, innovative startups are important to the economy because of their documented contributions to employment and economic growth. External finance can be critical to these firms. When the economics and finance literature has considered the impact of financing frictions on startup employment in the cross-section of entrepreneurs, it has typically focused on access to bank financing. However, recent work shows that venture capital firms also play an important role. Although venture capital firms back less than 1% of all startups, venture-backed companies account for 41% of total US market capitalization and 62% of US public companies' R&D spending. In this paper, we study access to venture capital for Black founders of innovation-driven startups. Earlier literature on this topic often uses survey data that combine high-growth innovation-driven startups with self-employed individuals and small and medium startups, many of which are not eligible for venture funding. Using data on startup founders who applied for at least one patent and raised at least some external funding, we find that Black founders raise 84 percent less venture capital funding in the five years following their first patent application, relative to non-Black founders. This is true even though Black founders' patents are just as likely to be approved and just as likely to be cited (conditional on being approved). Observable characteristics (e.g., alma mater prestige, being a serial founder, fundraising from non-venture capital sources) of the Black and non-Black startups are similar, which would seem to imply a similar demand for capital and growth ambitions. To further rule out a demand channel for our findings, we exploit the founders' implied race using their pictures as an instrument. Overall, the funding gap does not appear to be driven by a demand-side explanation. One factor alone seems to attenuate, and may even reverse, the funding gap: the presence of Black-owned venture capital investment firms.

⋆ We would like to thank Uday Rajan, Sugato Bhattacharyya, Andrew Toole, Shane Miller, Andrey Malenko, Nadya Malenko, Isaac Hacamo, Kristoph Kleiner, Erik Mayer and seminar participants at the University of Michigan, Rice University, and the USPTO for helpful comments.

I. Introduction

Broad participation in innovation has been shown to benefit the entire economy (Hunt et al., 2013; Cook and Yang, 2019). Similarly, participation in entrepreneurship is frequently discussed as an important factor in both wealth accumulation and the persistence of income inequality. Most of the wealth from entrepreneurship comes from exits of high-growth innovative startups via an initial public offering.[1] This paper is the first to investigate how and why access to venture capital funding for high-growth innovative startups varies by the race of their founder.

Fairlie et al. (2022) show some evidence of differences in business outcomes (employment, sales, and closures) for Black-owned startups and call attention to a lack of research explaining the differences in funding access among high-growth innovative startups. They surmise that one reason for this gap in the literature is the lack of detailed data on founders, startups, and investors. Even when such data are available, the data typically lack information on the founders' *and* investors' race. To test for and explain differences in access to funding between Black and non-Black founders, we processed over 20,000 profile images of founders that applied for a patent, and generated detailed characteristics of the founders and investors in high-growth innovative startups.

To construct our sample, we began with a set of first-time patent assignees that applied for, but did not necessarily obtain, a utility patent with the USPTO between 2001 and 2017. To identify high-growth startups, we further restricted the sample to firms that had raised at least a dollar of outside equity funding according to PitchBook. Specifically, we kept applicants that had raised at least a dollar of funding from venture capital groups, angel investors, grant funding, equity or product crowdfunding, or funding from the Small Business Innovation Research (SBIR) and the Small

[1] We define a high-growth innovative startup as a startup that filed a patent application with the USPTO and raised at least a dollar of external financing.

Business Technology Transfer (STTR) programs. We next collected information on the founders of these startups from PitchBook and various web searches. We then classified the founders as Black or non-Black using a combination of (i) their photo on LinkedIn or other online sources, (ii) their affiliations with Black groups on LinkedIn, and (iii) self-identification as a Black founder on various crowdsourced lists of Black founders (when their profile picture alone was not enough). Our final sample comprises about 5,425 startups that filed for a patent with the USPTO and raised at least a dollar of external funding. We define a startup as Black-owned if at least one founder is Black.

We first show that Black founders account for 2 percent of these high-growth innovative startups. Given that Blacks constitute 14.2 percent of the U.S. population in the 2020 census, Black founders' representation amongst high-growth startups is 86 percent lower than Blacks' representation in the general population.

Why is their share so small? To address this question, we first examine the characteristics of Black and non-Black startups. One stark difference is that Black founders raise about 84 percent less VC funding than non-Black founders in the five years following their first patent application.

This gap in venture funding could result from at least three different mechanisms. First, venture capital investors may provide less funding to Black startups because of bias, which could be driven by "taste-based" discrimination, biased beliefs, or even organizational practices, such as the common practice of sourcing deals from within the VC's network (which may not include Black founders) (Gompers et al. (2020), Small and Pager (2020)). Second, investors may use race as a proxy for an unobserved variable that is positively related to future success but negatively correlated with Black founders, at least unconditionally. In doing so, investors "statistically discriminate" against Black founders, reducing the amount of funding for Black-owned startups. Third, Black founders may raise less funding because investors observe variables (like Black founders' demand for equity funding or the quality of the founders' ideas) that we researchers do not. The key difference between the second explanation and the third is that the omitted variable is observed in the latter.

To address this identification problem, we first directly test whether certain omitted variables are likely to explain the funding gap. We use patent-examiner evaluations of the startups' patent applications to investigate whether the evidence supports the notion that ideas from Black founders are of lower quality. We also assess the importance of the demand-for-equity channel by testing whether similar funding gaps are present in other sources of funding. Finally, we use variation in the skin tone of founders (from their images) as an instrument to rule out demand-based explanations. We show that the instrument is positively related to being a Black founder, but not to exit outcomes such as acquisitions, IPOs, or firms' failures.

Next, we test for the relative influence of bias or statistical discrimination on the funding gap by considering follow-on funding rounds, startup outcomes, and the characteristics of venture capital investors in these startups.

We first show that the funding gap is absent when we look at other sources of equity funding for the *same* founders in the *same* year. Black startups raise as much equity funding as non-Black startups from non-venture-capital sources (defined as grants, equity and product crowdfunding, accelerators, and the SBIR and STTR programs). Unlike venture capital funding, these funding sources involve less personal interaction between founders and investors, partly due to established application and review processes. We also show that Black founders and non-Black founders are equally likely to be granted a patent by the USPTO, be a serial entrepreneur, and have attended a high-ranked college or university. These findings would seem to imply that the two groups are similar in terms of demand for capital, growth ambitions, and quality of ideas.

We also confirm the gap in funding when we use skin tone as an instrument for the proportion of a startup's founders that are Black. Because variation in the instrument occurs when our image-processing software predicts that dark-skin Asian founders are Black or light-skin Black founders are non-Black, we conduct a clerical review process of our ethnicity classifications. When an image is difficult to classify, we eliminate error using the founder's self identification as Black; the founder's appearing in various news and internet sources, including crowd-sourced lists of Black founders; and whether the founder attended a historically Black college or university (HBCU).

To investigate the type of prejudice that drives our results, we exploit the staggered nature of venture-capital funding rounds, the size of Black founder exits via an IPO or acquisition, and characteristics of investors. The evidence suggests that Black founders are *more* successful per dollar of VC funding and raise eight times more funding in a follow-on round, which is consistent with investors being biased against Black founders (Bohren et al., 2019). However, cross-sectional tests using investor characteristics do not support a pure taste-based discrimination story; rather, they suggest that organizational practices that introduce bias in the investment process are an important explanation of the funding gap. We also show that the funding gap is reversed when the venture capital fund is Black-owned, suggesting that bias-inducing organizational practices are not uniform across all venture-capital investors.

Our paper is related to the literature on the funding gap between Black and non-Black founders in other settings. Bates and Bradford (1992) report that minority-owned businesses and particularly Black-owned businesses have less access to venture capital financing. Park and Coleman (2009) find that Black-owned businesses tend to have a lower line of credit than White-owned firms and firms owned by other minorities, even though they have comparable demand for it. Bates et al. (2018) suggest that higher capital-search costs for minority-owned businesses explain the funding gap, and find empirical evidence that capital providers get higher returns from minority-owned projects—evidence, in their view, of taste-based discrimination. Fairlie et al. (2021) point out that Black-owned businesses are able to raise only a small fraction of outside equity, such as from VC and angel investors, compared to White-owned firms. Unlike our paper, however, these papers use survey data that combine high-growth, innovation-driven startups with self-employed individuals and small and medium startups, many of which are not eligible for venture funding. This makes the papers' estimates of the funding difficult to interpret. In contrast, our paper focuses on high-growth innovative startups, leverages founder and investor characteristics to explore reasons for the funding gap, and uses patent application data to test whether Black founders have lower-quality ideas.

The evidence on the funding gap for non-venture-capital or angel funding is mixed. On the one hand, Ginther et al. (2011) find that, after controlling for relevant applicant characteristics such as educational background and previous research awards, African Americans are less likely to receive NIH funding than White researchers. Ginther et al. (2018) note that Black applicants have fewer papers and citations and that this difference could explain a substantial amount of the funding gap. However, such patterns in publishing and citations have also been shown to be influenced, at least in part, by racial discrimination. Younkin and Kuppuswamy (2018) find that Black men are less likely to raise money from crowdfunding sources because the quality of their product is perceived to be low. On the other hand, Smith and Viceisza (2018) analyze the results of a TV pitch competition and find that there were no significant differences in the number of offers by race or ethnicity. Forscher et al. (2019) point out that there is little to no bias in the *initial* rounds of NIH R01 grant proposals, but they acknowledge that the possibility of bias at other stages of the funding process. We contribute to this research by showing that, amongst high-growth startups, Black founders raise as much non-venture-capital funding as non-Black founders and have similar characteristics.

More recently, Ewens and Townsend (2020) show evidence of gender biases amongst angel investors. Specifically, they find that male investors express interest in male-led startups even though male-led startups underperform female-led startups. Howell and Nanda (2019) show that networking frictions might explain the gender financing gap, as male participants in new venture competitions are more likely to proactively reach out to VC judges after the competition. Hebert (2020) shows that, while female-founded firms raise 27 percent less external equity, the gender-related funding gap reverses in female-dominated sectors (i.e., female entrepreneurs become more likely to raise funding than male entrepreneurs). Gornall and Strebulaev (2020) use a randomized field experiment to show that investors do not discriminate against female or Asian entrepreneurs when evaluating unsolicited pitch emails. Our paper extends this line of inquiry by studying the funding gap amongst female founders of innovation-driven startups.

II. Empirical Strategy

For most of our analysis, we will mainly rely on Poisson regressions to explain the amount of external funding a startup raises as a function of the proportion of its founders that are Black. Our choice is driven by the fact that some startups in our sample do not raise any external equity funding and thus have zero as their dependent variable. Cohn et al. (2021) show that Poisson regressions provide consistent estimates in this setting.

Our ultimate goal is to quantify the extent to which Black founders of high-growth startups face difficulties in raising external financing, and to examine possible explanations. We estimate the following reduced form equation using Poisson regression:

$$\begin{aligned}\text{Funding Raised}_i = \exp[&\alpha_1 + \beta_1 \text{P(Black)}_i + \beta_3 \text{P(Female)}_i \\ &+ X'\gamma + \lambda_j + \omega_s + \eta_t + \epsilon_{it}]\end{aligned}$$

Or equivalently

$$\begin{aligned}\ln(\text{Funding Raised}_i) = &\alpha_1 + \beta_1 \text{P(Black)}_i + \beta_3 \text{P(Female)}_i \\ &+ X'\gamma + \lambda_j + \omega_s + \eta_t + \epsilon_{it},\end{aligned} \tag{1}$$

The dependent variable, *Funding Raised*, is the cumulative amount of equity funding a startup (i) receives two, three, four, or five years following its first patent application. For startups that did not raise equity funding in the five years following their patent application, funding raised is zero. *Funding Raised* comprises funds from venture capital firms and other sources such as angel investors, accelerators, crowdfunding (proceeds from product or equity crowdfunding), grants, the small business innovation (SBIR) program, and the Small Business Technology Transfer (STTR) program. The unit of observation is a startup (i) that applied for a patent in year (t). *P(Black)* is the proportion of founders that are Black; *P(Female)* is the proportion of founders that are female. λ_j, ω_s, and η_t are indicators for the startup's industry and headquarter state and the year when it applied for its first patent, respectively. X is a vector of other control variables such as firm age, the log number of years between the application and the founding year, the proportion of founders that are inventors (as listed on the patent application), the number of founders, and the number of inventors.

Our hypothesis, informed by anecdotal evidence and various news reports, is that $\beta_1 < 0$. The identification challenge lies in quantifying β_1 and explaining *why* $\beta_1 < 0$. We consider three main reasons.

First, investors may provide less funding to Black startups because of bias. Such bias could have several possible sources. One possible source is taste-based. According to this view, investors *consciously* dislike working with Black founders and expect Black founders to "compensate" them for this distaste by accepting less funding (or a lower valuation) than non-Black founders—despite having the same probability of future success. Another possible source of bias is biased beliefs (Bohren et al., 2019). According to this view, investors simply underestimate black founders, at least initially. Yet another possible source of bias is organizational (Small and Pager, 2020). According to this view, certain organizational practices, such as habitually selecting investments based on referrals from people within an investor's network, have prevented Black founders from entering investor networks and made it harder for them to raise funding. Consistent with a heavy use of networking, Gompers et al. (2020) show that less than 28% of venture capital deals are sourced from outside the VCs' networks.

Second, investors may associate race with an unobserved variable, X_1, that is positively related to future success but negatively related to Black founders. According to this view, they "statistically discriminate" against Black founders by using race as a proxy for this variable. This discrimination, in turn, affects the amount of funding that investors provide to Black-owned startups. Implicit in the statistical discrimination theory is the notion that the investors' beliefs are correct—i.e., that Black-owned startups have worse outcomes, on average, than similar non-Black startups.

Third, the sign of β_1 might be driven by a variable that investors observe (like Black founders' demand for equity funding) but that we empirical researchers do not. To the extent that this "omitted variable," X_2, is positively associated with Black founders and negatively associated with fundraising, it might confound our estimates of equation 1. This explanation differs from the second explanation in that here the omitted variable is observed, whereas in the second explanation it is not.

We can formalize these competing explanations using the following system of equations:

$$\begin{aligned}\ln(\text{Funding Raised})_i = {} & \alpha_1 + \beta_1 \text{P(Black)}_i + \beta_3 \text{P(Female)}_i \\ & + X'\gamma + \lambda_j + \omega_s + \eta_t + \epsilon_i,\end{aligned} \tag{2}$$

$$\begin{aligned}\ln(\text{Funding Raised})_i = {} & \alpha_2 + \rho \text{P(Black)}_i + \delta_1 X_1 + \delta_2 X_2 \\ & + \beta_3 \text{P(Female)}_i \\ & + X'\gamma + \lambda_j + \omega_s + \eta_t + \epsilon_i,\end{aligned} \tag{3}$$

$$\text{P(Black)}_i = \alpha_3 + \gamma_1 X_1 + \gamma_2 X_2 + \epsilon_i, \tag{4}$$

$$\beta_1 = \rho + \delta_1 \gamma_1 + \delta_2 \gamma_2. \tag{5}$$

Equation (3) is the correct reduced-form Poisson regression we would like to run, but (2) is the regression we actually do run, because we do not know X_1 and X_2. We can derive (5), the omitted variable bias formula, by combining (2), (3), and (4) to show how β_1 is a combination of the three forces discussed above. Thus, if our hypothesis that $\beta_1 < 0$ is correct, it could be due to bias, ρ, "statistical discrimination," $\delta_1\gamma_1$, or an "omitted variable," $\delta_2\gamma_2$. Our identification strategy relies on direct tests of possible omitted variables and an instrument to mitigate concerns about omitted variables. We then conduct outcome and cross-sectional tests to gauge the relative importance of ρ relative to $\delta_1\gamma_1$.

A. Omitted variables versus statistical discrimination and/or bias

A.1. Demand for Funding

In our first test, we assess whether Black founders have a lower demand for external funding. To the extent that Black founders anticipate being discriminated against, we hypothesize that they should have a low ex ante demand for *all* forms of external financing. In fact, Fairlie et al. (2021) show that Black entrepreneurs often avoid applying for bank funding because they anticipate being denied a loan.

We implement this hypothesis by testing whether Black founders are less likely to raise VC and non-VC funding. A finding that Black founders are less likely to raise funding from both sources does not conclusively support the demand channel. However, a sharp difference between the two funding sources would cast doubt on demand-side factors as an explanation for the funding gap.

A.2. Quality of the idea

Our second test focuses on the quality of the idea. We test the extent to which Black and non-Black founders' respective ideas differ in quality. Specifically, we proxy for the quality of the idea by testing whether Black founders are less likely to have their patents granted, Black founders are less likely to have patent citations (conditional on the patent being granted), and Black founders' patents take longer to approve (assuming that hard-to-evaluate ideas take longer to be approved). To the extent that these signals are present and that investors rely on them when funding startups, Black founders are likely raise less funding.

A.3. Instrumental Variables Regression

In our final test of the influence of omitted variables on β_1, we use an instrument that influences how investors perceive race but is unrelated (an assumption we test) to a startup's probability of success (Rose (2022)).Our instrument is *P(Skin Tone)*, the average predicted probability that the founders of the startups are black, as computed by our image processing algorithm.[2] We will show that this instrument is positively related to the proportion of a startup's founders that are black, *P(Black)*, but unrelated to the startups' eventual outcomes, meeting the requirements for a valid instrument.

B. Statistical Discrimination versus Bias

After testing the relative importance of omitted variables versus statistical discrimination or bias, our next set of tests attempt to quantify the relative importance of each of the three sources of bias described above.

B.1. Follow-on funding

Our first test of the source of bias leverages the Bohren et al. (2019) prediction that biased beliefs might lead investors to fund Black founders at a lower rate, at least initially. However, following positive project updates from Black founders, investors will update their *biased* priors and provide larger amounts of funding to Black founders in follow-on funding rounds. Conceptually, the magnitude of the update should be greatest in the second round of funding relative to later rounds. We implement this prediction by running the following OLS regression:

[2] We use Google Image API and several other image-processing APIs and obtain similar results. Our results are not sensitive to the image-processing software that we use.

$$\left(\frac{\text{Deal Size Current Round}}{\text{Deal Size Previous Round}}\right)_{ivt} = \alpha + \beta_1 \text{P(Black)}_{iv}$$
$$+ \beta_2 \text{P(Black)}_{iv} \times \text{I(Round Two)}_{vt}$$
$$+ \beta_3 \text{I(Round Two)}_{vt}$$
$$+ \beta_4 \text{P(Female)}_{iv}$$
$$+ X'\gamma + \lambda_v + \epsilon_{ivt}$$

where a unit of observation is a startup-lead investor-year. We restrict our sample to the first three rounds of funding (when the update should be greatest). λ_v is an investor fixed-effect, and *I(Round Two)* is an indicator that equals one for startups raising their second funding round. Note that this test is only defined for startups that have raised at least one round of venture capital funding. Our hypothesis is that, within an investor, $\beta_1 > 0$ or $\beta_2 > \beta_1$ suggests that biased beliefs are an important component of the funding gap.

B.2. Post-Investment Outcomes

To evaluate the relative importance of statistical discrimination versus other bias sources, our next test focuses on outcomes. If the funding gap is caused by statistical discrimination, startups with a high proportion of Black founders should be less likely to be successful or more likely to fail. However, if the funding gap is caused by "taste-based" bias, the "outcome test" (Becker, 1993) suggests that firms with a high proportion of Black founders will be *more* successful due to the higher threshold they have to pass to obtain funding.

B.3. Investor Heterogeneity

Our final test to evaluate the relative importance of different sources of bias exploits heterogeneity across investors. We leverage cross-sectional characteristics of VC firms, including whether they have previously funded a Black-owned startup, have an abundance of experience, have achieved past success, and are themselves Black-owned. Our hypothesis is that taste-based bias is less likely among successful investors, experienced investors, investors that have previously funded Black founders, and investment firms that are Black-owned. This hypothesis builds on the labor economics literature. In particular, Arrow (1998) writes that "presumably the population of employers is not uniform in its discriminatory tastes. Then, under the usual assumption of constant (or increasing) returns to scale, competition would imply the elimination of all but the least discriminatory employers. If there are any non-discriminatory employers, they would drive out the others." Consistent with this, empirical evidence from Pager (2016) shows that firms that discriminate are more likely to go out of business.

III. Data

A. Sample Construction

Our goal is to study to what extent there are disparities in funding for Black founders of high-growth startups, and explore mechanisms driving the disparities. As such, our first task is carefully assembling a set of high-growth innovative startups that are Black-owned. To keep our search manageable, we focus of U.S.-based companies in the PitchBook database that have raised at least a dollar of external financing between 2001 and 2021, applied for a utility patent with the USPTO between 2001 and 2017, and were not formed more than five years before they applied for their first utility patent.

The next few sections detail how we arrive at our final sample by, first, matching PitchBook startups to USPTO assignee data, then classifying each of the matched startups as either Black or non-Black owned using pictures of startup founders from LinkedIn, and then merging in data on government grant funding.

B. Matching PitchBook to USPTO

For each startup in PitchBook, we get the complete history of their name changes, which PitchBook tracks. Next, we merge these startups, on name and state where they are headquartered, to data on all assignee's in the USPTO database as of 2017.[3] We focus on assignees that applied for a patent, but did not necessarily obtain one. When an assignee has multiple applications, we keep the first application ever filed. If multiple applications are filed on the same day, we keep the application that was eventually granted or generated the most forward citations. We further restricted the set of matches to assignees that were formed no more than five years before their first patent application. The restriction on founding year allows us to identify young startups, where financial constraints are more likely to be salient (Hadlock and Pierce (2010)). We matched about 19,000 startups in PitchBook to first-time applications in the USPTO database using this procedure. See section VI for a more in-depth discussion of the matching process.

C. Classifying Startups in the matched PitchBook-USPTO by founder race

Our next step is to calculate the proportion of Black founders for each startup in our PitchBook-USPTO sample of startups that applied for a patent for the first time between 2001 and 2017. To that end, we first get the list of all founders of these startups from PitchBook.[4] Next, we collect LinkedIn profile links for each of the founders and download their profile pictures, which we use to classify founders as Black or non-Black. We use an image processing algorithm to classify founders, and discuss the algorithm in more detail in section VI. We add clerical review to the

---

[3] We use the python package, fuzzywuzzy, to compare the standardized name in PitchBook to the standardized name in the USPTO assignee data. We hand verify all matches for accuracy.

[4] We search for the keywords "founder" or "owner" in job titles, and manually inspected the results to make sure that we identified the founders of the startup.

classification process to reduce errors. Specifically, we manually verified that each founder flagged as Black by our image-processing algorithm is indeed Black, and then calculate the proportion of a startup's founders that are Black. For our review process, we use a combination of the founder's picture, affiliation with Black groups on LinkedIn, and self identification as a Black founder from the founders LinkedIn profile, various news sources, and crowdsourced lists of Black founders.[5]

We were not able to classify all the founders in our sample by race. For some founders, we could not find a LinkedIn profile and others had a LikedIn profile but no picture. We dropped startups for whom we had no pictures for at least one founder. Our final sample comprises about 5,420 startups formed between 1996 and 2017, that applied for a patent between 2001 and 2017, where we could verify the race of at least one founder using profile pictures from LinkedIn. Ninety-four of these startups have at least one Black founder.

D. Other data on fundraising activity

PitchBook has fundraising data for startups that raised funding from venture capital investment firms, private equity firms, corporate venture capital firms, accelerators, incubators, grants, and equity and product crowdfunding. Nonetheless, we supplement the data by downloading SBIR (Small Business Innovation Research) and STTR (Small Business Technology Transfer) grant data from the SBIR's website[6]. We use the state where each award recipient is located and their name to match startups in our sample to this data.[7]

[5] Here is an example of a crowdsourced list of Black founders.
[6] `https://www.sbir.gov/sbirsearch/award/all`
[7] We also used the python package, fuzzywuzzy, to compare the standardized name in PitchBook to the standardized name in the SBIR data. We manually verified all matches for accuracy.

IV. Results

A. *Summary Statistics*

The first part of Table 1 presents the summary statistics of our sample for the different variables used in the analysis at the startup level. The first fact that stands out is that less than 2% of all startups have at least one black founder. Given that about 16% of the U.S. population of working adults is Black, Black founders are underrepresented amongst high-growth startups pursing innovation by a factor of 8. Some of the main variables are related to the amount of capital raised from different sources. In this respect, note that Black founders raise less venture capital (VC) funding over the five years following their first patent application, but raise as much capital as non-Black founders from from other sources (non-VC funding). These other sources comprise angel funding, crowdfunding, accelerators, incubators, the Small Business Innovation Research (SBIR) program, and the Small Business Technology Transfer (STTR) program. This table also shows that in our sample, Black-owned firms are not more likely to fail or be acquired, but we also see that no Black-owned startup went public over our sample period, a fact we return to when we look at startup exits.

A critical component of our analysis is the "quality" of the founder. We proxy this with indicator variables that is equal to one if founder graduated from a top school or is a serial founder. Our data, in Panel B of Table 1, shows that in this respect there are no significant differences between Black and non-Black founders.

B. *Sample Selection*

Given that our sample is the subset firms that applied for a patent for the first time and raised at least a dollar of external funding according to PitchBook, in the next two sub-sections we investigate how our sample compares to the population of first-time assignees and other startups in PitchBook.

B.1. Sample versus other startups in PitchBook

In this section, we investigate how our sample compares to the population of startups raising external financing. For this test, we compare the firms in our sample to all startups in PitchBook formed between 1996 and 2017 that raised at least a dollar of any external funding, startups that belong to the same cohort as those in our sample.

Panel C of Table 1 shows the descriptive statistics for startups in Pitchbook by whether they are in our sample. The results are very clear: the startups in our sample attract a large amount of funding, but do not necessarily have better outcomes. The results on funding differences between the firms in our sample and the firms in PitchBook suggest that we are focusing on the set of startups that likely attract the disproportionate share of funding dollars, where any funding gap is likely to be most salient.

B.2. Sample versus all Patent Assignees

In this section we compare our sample to the overall sample of first-time patent applicants. Given our focus on startups, we compare the firms in our sample to U.S.-based patent applicants that filed for a first patent between 2001 and 2017. We identify likely startups be restricting our sample to applicants that qualified for a reduced filing fee with the USTPO by satisfying criteria defining a "small business entity." Farre-Mensa et al. (2020) use a similar restriction.

Panel D of Table 1 shows the results. We see that applicants in our sample have a larger number of inventors, are less likely to be granted a patent, are equally likely to abandon their patents, applied for patents more recently, wait slightly longer to obtain their patents, and have fewer citations. Some of these differences could be a consequence of the startups in our sample being younger, since we have no way of verifying whether the other first-time assignees did not apply for the patent pre-2001, or that some of these assignees are not already public firms. Overall, applicants in our sample applied for patents more recently compared with the overall population of first-time assignees.

C. Black Innovation and Fundraising

This section presents our estimates of equation 1, which we show in Panels A and B of Table 2 and estimate using Poisson maximum likelihood. A unit of observation is a first-time assignee (startup) that applied for a patent between 2001 and 2017 and was formed no more than five years before its first patent application. The dependent variable in Panel A is the total amount of VC funding a startup raises 2, 3, 4, and 5 years following its first patent application. For startups that raise no venture capital funding, this variable is zero.[8] The dependent variable in Panel B is the total amount of non-VC funding the startups raises 2, 3, 4, and 5 years following its first patent application. All regressions include state (where the startup is headquartered), year (when the startup filed its patent application), and industry fixed-effects (where the industry can be one of seven sectors in PitchBook, as presented in the last seven rows in Panel A of Table 1). We fix the number of inventors listed on the startups patent application, the age of the startup when it filed the application (five years or fewer), the proportion of founders that are inventors, and the number of startup founders.

The estimates in Column (1) of Panel A imply that, all else equal, increasing the proportion of black founders from zero to one reduces the amount of venture capital funding a startup can expect to raise over the next two years by 84% ($100 \times (e^{-1.864} - 1)$). Similarly, increasing the proportion of female founders from zero to one reduces the amount of funding the startup can expect to raise over the next two years

[8] A round qualifies as a venture capital round if it is classified as “Early Stage,” “Late Stage,” of “Seed funding” by PitchBook *and* the primary investors are venture capital funds with a limited partnership structure. We further require that PitchBook lists the investor’s type as “Venture Capital,” “PE/Buyout,” “Growth/Expansion,” “Corporate Venture Capital,” “Other Private Equity,” or “Not-For-Profit Venture Capital.”

by 43%, or about half the estimated funding gap for black founders. At the five year horizon shown in Column (5), the funding gap is 77% for Black founders and 43% for female founders.

However, when we turn to Panel B, there does not appear to be a funding gap for Black startups, and there is now a positive funding gap for female startups. The estimates imply that Black startups that apply for patents raise as much funding from non-venture sources as non-Black startups, and that female startups that apply for patents raise about 82% more funding from non-VC sources, which comprises funding from angel investors, accelerators, equity crowdfunding, grants, the Small Business Innovation Research (SBIR) program, and the Small Business Technology Transfer (STTR) program.

We argue that this difference in VC and non-VC funding makes a demand-side explanation for the funding gap weaker. To the extent that the Black founders are not applying for external financing for fear of being denied funding, we should observe a similar pattern in VC *and* non-VC funding.

The signs on the other control variables are consistent with intuition: startups with more founders or inventors raise more funding (they can divide and conquer the fundraising process), and older startups raise more funding (are likely more established). We also find that founders that are also inventors are less likely to raise VC funding, suggesting that while these founders might have the technical skills to be inventors, they may lack the interpersonal skills required for fundraising.

C.1. Robustness

Given that we condition our analysis on a startup being in PitchBook, which is more likely to cover startups that have raised some external financing, we hypothesize that our estimates likely understate the magnitude of the funding gap for Black founders pursuing innovation, as we are likely missing more Black founders that did not raise *any* external funding. To test to what extent our estimates understate the true funding gap, we repeat our analysis using the same sample in Farre-Mensa et al. (2020), which also covers first-time assignees that applied for a patent.[9] We present our estimates using that sample in Table A.1. Consistent with our hypothesis, the funding gap is much larger in this sample. The estimates from Column (5) of Panel A imply that, all else equal, over a five year horizon, increasing the proportion of black founders in a startup from zero to one decreases the amount of venture funding the startup can expect to raise by 94%. However, consistent with our analysis using the PitchBook sample, there is no funding gap when we look at non-VC funding.[10]

C.2. Why is there a gap in VC funding for Black startups?

Across all columns, we see that there is a significant VC funding gap between Black and non-Black startups. As discussed in Section II, we must interpret this result with caution as it likely represents the combined effect of bias, statistical discrimination, or omitted variables bias. We continue our investigation of the extent to which omitted variables might drive our funding-gap estimates by testing whether patent applications by Black founders are "lower-quality".

C.3. Black Startups and Patent Grants

In this section, we test whether patent applications by Black startups are lower-quality, which might explain the difference in fundraising we document above. As a proxy for quality, we use an indicator that equals one if any of the patent claims on the patent application are granted by the second quarter of 2020, *I(Granted)* , patent citations (as of 2017), *Citations*, and number of years between the patent application and grant, *Years to Grant*. We also control for firm age (the number of years between

[9] We download their sample from the Journal of Finance's website

[10] We are less thorough in our classfication of Black founders in this sample, given the larger sample size. To predict inventor race in this sample, we first use NamePrism (also used in Egan et al. (2017)) to predict the probability that a founder is black. Then we downloaded and processed all images (using the same procedure that we used in the main analysis) where the probability that the inventor is black is 40% or greater. This procedure reduces the false positive rate (classifying a non-Black founder as Black), but likely has a high false negative rate (classifying a Black founder as non-Black). Thus, this coefficients also likely underestimate the true funding gap.

when the company was formed and when it applied for a patent), the number of inventors, and the proportion of founders that are also inventors. We further include USPC class by application year fixed-effect to control for variation in approval rates across different patent classes over time or selection of Black founders into certain patent classes. A unit of observation in this test is the first patent application a startup in our sample.[11]

Table 3 presents our results. Across all columns we do not find any evidence that patent applications by Black or female startups are less likely to be successful. This result suggests that the qualify of the idea is unlikely to explain the VC funding gap.

C.4. Robustness

We also repeat the test of patent outcomes using the sample in Farre-Mensa et al. (2020), which covers first-time assignees that applied for a patent between 2001 and 2009. Our estimates using this sample, see Table A.3, are virtually identical to what we find in the PitchBook-USPTO sample, suggesting that our results are not driven by how we constructed our sample or selection into PitchBook.

D. Instrumental Variables

To bolster our interpretation that the VC funding gap is not a result of omitted variables, such as Black-founder demand for capital, we turn to instrumental variables. We input images in our pre-trained model to predict the probability that a founder is Black and use that probability as an instrument. We call this probability *P(Skin Tone)*, because founders with dark skin tones are more likely to be predicted as Black.

We assume (and test) that skin tone is unrelated to a startup's probability of success, but positively to being a Black founder. From Panel B, row (1) of Table 1, we see that skin tone is positively correlated with being a Black founder. However,

[11] When a startup applied for multiple patents on the same date, we keep the most successful application — the application that was granted and generated the most citations.

not all founders with a dark skin tone are black. Variation in the instrument comes from clerical review, where we can eliminate Type I errors, such as dark-skin Asians that are predicted to be Black, and Type II errors, light skin Black founders with a light skin tone that are predicted to be non-Black. When an image is difficult to classify, we eliminate clerical error using founder's self identification as Black, various news and internet sources, and whether the founder attended an HBCU.

Table 4 tests the hypothesis that skin tone is unrelated to outcomes. Across all specifications, we see that skin tone is unrelated to a successful exit via and IPO or acquisition, and unrelated to the probability that a startup fails, supporting our assumption that the instrument meets the exclusion restriction.

Table 5 presents our two-stage least squares estimates of the relationship between having a high proportion of Black founders and funding raised.[12] The first two columns show that investors' perception of race affects the amount of VC funding raised in the five years following the patent application, while the last two columns look at non-VC funding. In line with our results in the previous sections, startups with a high proportion of Black founders struggle to raise VC funding, but there doesn't appear to be a gap in the amount of non-VC funding these startups raise.

Our evidence so far suggests that omitted variables are unlikely to be the primary explanation for the funding gap. In the next section, we test to what extent statistical discrimination or bias influences the VC funding gap.

[12] To facilitate reporting of our estimates, we run these regressions be taking the log of the dependent variables and adding one when it is zero. Given that the dependent variable is in millions, this is equivalent to adding a dollar to startups that raised no funding. We obtain similar estimates if we instead use a control function approach and run Poisson regressions.

E. Statistical Discrimination and/or Bias

E.1. Biased beliefs and follow-on funding rounds

Our first test to analyze the extent to which biased beliefs explain the Black funding gap leverages the staged nature of venture funding and the prediction from Bohren et al. (2019) discussed in Section II. Given that investments are typically staggered over time as a function of milestones, we hypothesize that, while the first funding round for Black founders is initially low, within an investor's portfolio, the amount of funding Black founders raise should be higher in later rounds, as investors positively update their initial mistaken beliefs about the prospects of Black startups.

Table 6 presents estimates of the relationship between the proportion of a startups founders that are Black and the size of the follow-on round relative to the first round. We restrict our attention to the first three rounds the startup raises, where we would expect the largest updates to investors' beliefs. A unit of observation in this test is a startup-lead investor-year, where year is the year when the startup raised its second or third round of funding. Note that all startups in this test raised at least some VC funding, hence the lower number of observations. All regressions have investor-fixed effects so that we can look at changes in investor beliefs, within an investor over time.

In Column (1) of Table 6, we see that, relative to non-Black founders, Black founder's second round of funding is eight times larger than their first round. However, when we go from the second to the third round, in Column (3), there is a smaller update in the amount of funding Black founders raise, consistent with investor's updating their priors on Black founders following the first round of funding. We also see a similar update for female founders, but only when they raise their third round of funding, suggesting perhaps a slower update to investors' beliefs. Given the lack of statistical significance in the second, third, and fourth columns, we view this results as weakly suggestive of biased beliefs.

Overall, the results in this section suggest that biased beliefs might be a possible driver of the Black funding gap amongst high-growth innovative startups.

E.2. Statistical discrimination and post-funding outcomes

Our next set of tests focus on outcomes. If statistical discrimination is the primary driver of the funding gap, we would expect Black-owned startups to be less likely to exit and more likely to fail. However, if bias explains the funding gap, we would expect Black-owned startups that exit to be more successful, given the higher threshold they had to pass to obtain funding.

From Panel A and B of Table 7, we see that Black startups are just as likely to exit via an acquisition and just as likely to fail, even though they raise less VC funding. Further, we see that female-owned startups are less likely to exit via an acquisition and are more likely to fail, even though the funding gap for Black startups is twice that of female startups. We do show separate results for IPOs because, as we show in Panel A of Table 1, we could not find a single IPO by a Black startup, which makes interpreting relative effects tricky. From previous work, we also know that part of the lack of IPOs for Black startups may be caused by the lack of VC funding we document above (Bernstein et al. (2016)).

Nonetheless, we combine IPOs and Acquisitions in Table 8, where we test whether Black-owned startups exit at higher valuations relative to non-Black startups. Again, the idea behind this test is that, under a mechanism based on bias perceptions, Black-owned startups would have higher exit proceeds per dollar of VC funding. From Column (4), we see that Black-owned startups are more likely to have higher exit proceeds per dollar of VC funding, favoring a bias argument over statistical discrimination.

Overall, the findings in this section are not consistent with statistical discrimination, but are weakly consistent with bias.

E.3. Using Investor Heterogeneity to investigate the source of bias

Our next test of the type of bias that drives the funding gap leverages cross-sectional characteristics of investors. In particular, we investigate whether the funding gap varies by whether the startup is funded by an investor that has been very successful in the past, has a lot of experience, has funded a Black-owned company in the past, or whether the venture capital firm is Black-owned. We define a venture group as Black-owned if 50 percent of the founders or senior partners of the group are Black.

Our hypothesis is that pure "taste-based" bias is less likely if we observe the same funding gap when a startup receives funding from an investor that has a lot of past success or experience, as these investors are less likely to have achieved success or lasted in the industry if they consistently practiced taste-based discriminated. Under a pure "taste-based" discrimination story, investors should eventually be punished for discriminatory behavior and should be more likely to exit the industry (Arrow (1998), Becker (1993), Becker (2010)). Under a non-taste based bias story, we also expect a smaller funding gap when the investor has funded another Black-owned startup or when the investor is also Black as these investors should be less likely to discriminate based on taste given their past exposure to Black founders or being Black founders themselves.

Table 9 presents our results. We see that the funding gap is not smaller for successful and experienced founders, but is smaller when a Black venture firm invests in the startup. Overall, these results are not consistent with a "taste-based" bias story and suggests that other frictions such as search costs, networks, or even unconscious bias might be the main driver of the funding gap. Note that these different explanations might be driven by systemic bias, which might have kept Black founders out of the networks of non-Black investors, increasing search costs for Black founders and unconscious bias from investors (Small and Pager (2020)).

V. Conclusion

We document a significant Black funding gap in VC funding amongst high-growth innovative startups. We leverage skin tone, patent applications, and type of funding to rule out demand-side explanations. Our evidence on the step-up in follow-on rounds of funding and exit proceeds per dollar of VC funding do not support a statistical discrimination story, but instead favor bias. We conclude by exploring the source of bias by leveraging characteristics of investors, and conclude that the evidence is not consistent with pure "taste-based" discrimination story. Other venture firm institutional practices such as referral-based investment recommendations might be an important driver of the funding gap.

Table 1: Characteristics of Startup Firms in Sample

This table reports summary statistics for startups that raised some external financing and applied for an utility patent with the USPTO. The unit of observation is a startup. *P(Black)* and *P(Female)* are the proportions of founders that are Black or Female, respectively. *Age (Yrs)* is the number of years between when the startup was formed and when it applied for a patent. *I(PE Hub)* is an indicator that equals one for companies located in California, Massachusetts, or New York. *5yrs VC funding* is the total amount of funding that the startup raised from venture capital firms (firms classified in PitchBook as "Venture Capital," "PE/Buyout," "Growth/Expansion," "Corporate Venture Capital," "Other Private Equity," or "Not-For-Profit Venture Capital") in the five years follwing its patent application. Similarly, *5 yrs Non-VC Funding* is the total amount of non-VC funding raised following the startup's first patent application. Non-VC funding is funding raised from angel investors, accelerators, equity crowdfunding, grants, the Small Business Innovation Research (SBIR) program, and the Small Business Technology Transfer (STTR) program. *I(IPO)* is an indicator variable that equals to unity if the firm went public by the second quarter of 2021. *I(M&A)* is an indicator that equals one if the startup was acquired by the second quarter of 2021. *I(Failed)* is an indicator that equals one if the startup has gone bankrupt or closed by the second quarter of 2021. *I(Business Products and Services)*, *I(Consumer Products and Services)*, *I(Energy)*, *I(Financial Services)*, *I(Healthcare)*, *I(Information Technology)*, and *I(Materials and Resources)* are indicators that equal one if the startup is in that industry. The final column reports the t-statistic for a test of differences in means of the different groups. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. All continuous variables are winsorized at the 1% and 99% levels to minimize the influence of outliers.

	Black Owned		**Non-Black Owned**		
	N = 94		**N = 5,331**		
	Mean	Std. Dev.	Mean	Std. Dev.	t-stat
A. Firm-level statistics					
P(Black)	0.68	0.31	0.00	0.00	21.46***
P(Female)	0.11	0.25	0.09	0.24	1.02
Age (Yrs)	2.16	1.42	2.20	1.43	−0.30
I(PE Hub)	0.54	0.50	0.56	0.50	−0.41
# Founders	1.99	1.07	1.80	0.97	1.69*
5yrs VC funding ($ Millions)	4.88	10.63	12.76	38.24	−6.52***
5yrs Non-VC Funding ($ Millions)	0.44	1.92	0.42	3.86	0.11
I(IPO)	0.00	0.00	0.02	0.14	−10.40***
I(M&A)	0.20	0.40	0.24	0.43	−0.91
I(Failed)	0.19	0.40	0.15	0.36	0.96
I(Business Products and Services)	0.18	0.39	0.15	0.36	0.80
I(Consumer Products and Services)	0.17	0.38	0.12	0.33	1.20
I(Energy)	0.01	0.10	0.03	0.18	−2.20**
I(Financial Services)	0.00	0.00	0.01	0.09	−6.27***
I(Healthcare)	0.22	0.42	0.24	0.42	−0.29
I(Information Technology)	0.38	0.49	0.43	0.50	−1.00
I(Materials and Resources)	0.03	0.18	0.02	0.13	0.86

Table 1 - *continued* (Founder Statistics)

This table reports summary statistics for founders of startups that raised some external financing and applied for a utility patent with the USPTO. A unit of observation is a founder of a startup in Table 1 above. *Black* is an indicator that equals one if any startup founder is Black. We classify a founder as Black using an image processing algorithm and clerical review. *P(Skin Tone)* is the probability that a founder is Black, as computed by the image processing algorithm. *P(African)* is the probability that a founder is African, as computed by NamePrism, a name-to-ethnicity algorithm used in other academic papers. *I(Founder-Inventor)* is an indicator that equals one if the founder is also the inventor listed on the USPTO patent application. *I(Serial Founder)* is an indicator that equals one if the founder's LinkedIn profile indicated that they started another company whose name differs from the focal assignee and whose formation predates the current startup that is the focus of our analysis. *(Top School Alumni)* is an indicator that equals one if the founder attended one of the top 20 universities, ranked by average SAT score of accepted freshmen in the year the startup raised funding. *I(Female)* is an indicator that equals one if the founder is female. The final column reports the t-statistic for a test of differences in means of the different groups. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. All continuous variables are winsorized at the 1% and 99% levels to minimize the influence of outliers.

	Black		**Non-Black**		
	N = 103		**N = 10,058**		
	Mean	Std. Dev.	Mean	Std. Dev.	t-stat
B. Founder-level statistics					
P(Skin Tone)	0.75	0.32	0.02	0.07	22.87***
P(African)	0.10	0.20	0.01	0.04	4.46***
I(Founder-Inventor)	0.65	0.48	0.51	0.50	2.93***
I(Serial Founder)	0.13	0.33	0.10	0.30	0.81
I(Top School)	0.43	0.50	0.36	0.48	1.42
I(Female)	0.14	0.34	0.09	0.29	1.37

Table 1 - *continued*

This table reports statistics on startups in our sample and other startups in PitchBook. A unit of observation is a startup formed between 1996 and 2017 that raised at least a dollar in external financing. *I(IPO)* and *I(M&A)* are indicator variables that equal one if the company has gone public or been acquired as of the second quarter of 2021. *I(Failed)* is an indicator that equals to unity if the company went bankrupt, failed, or closed as of the second quarter of 2021. *I(VC Funding)*, *I(Accelerator Funding)*, *I(Angel Funding)*, *I(Crowd Funding)*, *I(Grant Funding)*, *I(SBIR Funding)* are indicator variables that equal to one if the startup raised funding from the respective source. *Year Founded* is the year in which the company was founded. *Exit Value* is the value for which the company was acquired or valued at IPO respectively. The final column reports the t-statistic for a test of differences in means of the different groups. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. All continuous variables are winsorized at the 1% and 99% levels to minimize the influence of outliers.

	In Sample		**Others Startups**		
	N = 5,425		**138,994**		
	Mean	Std. Dev.	Mean	Std. Dev.	t-stat
C. Our Sample versus other PitchBook startups					
I(IPO)	0.02	0.14	0.03	0.17	-5.11^{***}
I(M&A)	0.24	0.43	0.32	0.47	-13.91^{***}
I(Failed)	0.15	0.36	0.11	0.31	9.47^{***}
I(VC Funding)	0.59	0.49	0.22	0.41	55.56^{***}
I(Accelarator Funding)	0.07	0.25	0.06	0.24	1.52
I(Angel Funding)	0.18	0.39	0.10	0.29	16.16^{***}
I(Crowd Funding)	0.05	0.21	0.02	0.15	8.23^{***}
I(Grant Funding)	0.16	0.37	0.05	0.21	22.65^{***}
I(SBIR/STTR Funding)	0.10	0.29	0.01	0.09	22.02^{***}
Year Founded	2008.72	4.63	2007.84	6.26	13.43^{***}
Exit Value ($ Millions)	223.12	573.49	270.41	1351.12	-1.74^{*}

Table 1 - *continued*

This table reports statistics for applications filed with the USPTO between 2001 and 2017 for a regular utility patent. We present statistics by whether we matched an application to LinkedIn and obtained a photograph of the founder of the first assignee assigned the application. We further require that our software could actually use the photograph to classify race. *# Inventors* refers to the numbers of inventors of a patent. *I(Patent Granted)* is an indicator variable that equals to one of the patent was granted. I(Patent Abandoned) is an indicator if an applicant fails to respond to an office action within six months. *Application Year* is the year in which the patent was filed. *Years between application and issuance* is the gap between when the patent was filed and when it was issued to the assignee. *# Citations* is the number of citations that a patent has received. *Years between application and issuance* and *# Citations* are only available only for the subset of applications that were eventually granted.

	Sample		**OTHERS**		**t-test**
# Observations	**5,425**		**103,484**		
	Mean	Std. Dev.	Mean	Std. Dev.	
D. Our Sample versus Other First-time Assignees					
# Inventors	2.63	1.71	2.09	1.41	22.78***
I(Patent Granted)	0.44	0.50	0.56	0.50	−17.05***
I(Patent Abandoned)	0.36	0.48	0.37	0.48	−0.52
Application Year	2010.92	4.42	2007.98	4.79	47.47***
Years between application and issuance	3.13	1.76	3.04	1.65	2.64***
# Citations	2.76	11.82	3.33	11.18	−3.52***

Table 2: Association Between Black Ownership and Funding Raised

This table presents coefficients from Poisson regressions run at the assignee (startup) level, with standard errors reported below in parentheses. We keep only the first patent application for each assignee. Where there are multiple applications for the same assignee, we keep the application that was granted or with the largest number of citations. The sample comprises assignees that applied for a patent between 2001 to 2017, where we could find at least one founder on Linkedin with a profile picture that allowed us to identify race. The dependent variable in Panel A, *VC Funding*, is the cumulative amount of VC funding the startup raised 2, 3, 4, or 5 years following the patent application. In Panel B, the dependent variable is *Non-VC Funding* the total amount of non-VC funding raised 2, 3, 4, or 5 years following the patent application. Non-VC funding is funding raised from angel investors, accelerators, equity crowdfunding, grants, the Small Business Innovation Research (SBIR) program, and the Small Business Technology Transfer (STTR) program. *P(Black)* is the proportion of founders (where we could identify race) that are Black, and *P(Female)* is the proportion of founders that are female. *Ln(Count Inventors)* is the log count of the number of inventors listed on the patent application. *Firm Age* is the number of years since the company was formed. *P(Founder-Inventor)* is the proportion of founders that are also inventors (listed on the patent application), and *Ln(Count Founders)* is the log of the number of founders of the startup. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by startup.

Panel A:	**VC Funding**			
Dependent Variable:	Next 2yrs?	Next 3yrs?	Next 4yrs?	Next 5yrs?
P(Black)	-1.864***	-1.819***	-1.844***	-1.498***
	(0.467)	(0.488)	(0.435)	(0.464)
P(Female)	-0.566***	-0.444***	-0.438**	-0.553***
	(0.196)	(0.171)	(0.184)	(0.181)
Ln(Count Inventors)	0.429***	0.373***	0.394***	0.373***
	(0.090)	(0.074)	(0.071)	(0.066)
Ln(Firm Age (yrs))	0.138***	0.099***	0.062**	0.054**
	(0.030)	(0.027)	(0.028)	(0.027)
P(Founder-Inventor)	-0.522***	-0.477***	-0.417***	-0.401***
	(0.126)	(0.109)	(0.112)	(0.106)
Ln(Count Founders)	0.433***	0.512***	0.540***	0.593***
	(0.131)	(0.109)	(0.101)	(0.095)
Log-likelihood	-52491.88	-66580.69	-82942.15	-96431.88
Panel B:	**Non-VC Funding**			
P(Black)	0.101	-0.038	-0.193	-0.154
	(0.641)	(0.635)	(0.553)	(0.520)
P(Female)	0.598***	0.457*	0.455**	0.461**
	(0.223)	(0.242)	(0.229)	(0.219)
Ln(Count Inventors)	-0.112	0.264	0.254	0.232
	(0.101)	(0.169)	(0.171)	(0.168)
Ln(Firm Age (yrs))	0.188***	0.091**	0.120***	0.097**
	(0.043)	(0.044)	(0.042)	(0.043)
P(Founder-Inventor)	0.259	0.109	0.154	0.154
	(0.157)	(0.153)	(0.160)	(0.157)
Ln(Count Founders)	0.640***	0.586***	0.620***	0.660***
	(0.167)	(0.135)	(0.162)	(0.152)
Observations	5410	5410	5410	5410
Log-likelihood	-2870.79	-4083.11	-5098.86	-5641.60
State, Year, and Industry FE?	Yes	Yes	Yes	Yes

Table 3: Association Between Black-Owned Assignees and Patent Outcomes

This table reports the estimated effect of the proportion of Black and female founders assigned a patent application on the likelihood the patent is granted, the number of citations the patent received, and the time between application and grant. The sample comprises assignees that applied for a patent between 2001 to 2017, where we could find at least one founder on Linkedin with a profile picture that allowed us to identify race. The dependent variable in Column (1) and (2), *I(Granted)*, is an indicator that equals one for startups whose patent application was ultimately successful. In Column (3), it is the number of citations the patent received as of 2017, (*Citations*), and in Column (4), it is the number of years between patent application and grant, *Years to Grant*. *P(Black)* is the proportion of founders (where we could identify race) that are Black, and *P(Female)* is the proportion of founders that are female. *Ln(Count Investors)* is the log count of the number of inventors listed on the patent application. *Firm Age* is the number of years between when the firm was formed and when it applied for a patent. *P(Founder-Inventor)* is the proportion of founders that are also listed as the inventor on the patent application. *Class X Year* denotes USPC class by application year fixed-effects. We estimate Columns (1) and (2) using linear probability models and Columns (3) and (4) using quasi-maximum likelihood Poisson models with robust standard errors. ^{g} denotes tests on the subset of applications that were granted. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by application.

Dependent Variable:	I(Granted)	I(Granted)	Citations	Years to Grant
	(1)	(2)	(3)	(4)
P(Black)	-0.006	0.078	0.510	-0.231**
	(0.067)	(0.078)	(0.617)	(0.110)
P(Female)	0.030	0.021	-0.148	0.011
	(0.028)	(0.029)	(0.289)	(0.053)
Ln(Count Inventors)	-0.022*	-0.004	0.019	-0.035*
	(0.011)	(0.012)	(0.095)	(0.018)
Ln(Firm Age (yrs))	-0.015***	-0.014***	-0.118***	0.013
	(0.005)	(0.005)	(0.040)	(0.008)
P(Founder-Inventor)	-0.019	0.002	0.099	0.066**
	(0.016)	(0.018)	(0.146)	(0.028)
Fixed Effects	Year	Class X Year	Class X Year	Class X Year
Observations	5425	4634	1363	1361
Log-likelihood	-3767.90	-2384.72	-6916.70	-2333.24

Table 4: Association Between Skin Tone and Outcomes (Instrument Exogeneity)

This table presents OLS regressions run at the startup level, with standard errors reported below in parentheses. Our sample comprises startups that applied for a patent between 2001 and 2017. We further require the startup's founders to have Linkedin profiles with pictures that allow us to identify race. The dependent variable in the first two columns, *I(Success)*, is an indicator variable that equal to unity if the startup is acquired, merged, or goes public. The response variable in the last two columns, *I(Failed)*, is an indicator variable that is equal to one if a firm failed or went bankrupt by the second quarter of 2021. *P(Skin Tone)* is the probability that the founder is Black based on their LinkedIn profile picture. *Ln(Count Inventors)* is the log of the number of inventors listed on the patent application. *Firm Age* is the number of years since the company was formed. *P(Founder-Inventor)* is the proportion of founders that are also inventors (listed on the patent application), and *Ln(Count Founders)* is the log of the number of founders of the startup. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by startup.

Dependent Variable:	**I(Success)**		**I(Failed)**	
	(1)	(2)	(3)	(4)
P(Skin tone)	0.028	0.031	-0.042	-0.049
	(0.062)	(0.061)	(0.042)	(0.042)
Ln(Count Inventors)		0.046***		0.011
		(0.010)		(0.008)
Ln(Firm Age (yrs))		0.022***		-0.017***
		(0.004)		(0.003)
P(Founder-Inventor)		-0.049***		0.016
		(0.015)		(0.011)
Ln(Count Founders)		-0.010		-0.040***
		(0.012)		(0.010)
Observations	5421	5421	5421	5421
Adjusted R^2	0.10	0.11	0.02	0.03
State, Year, and Industry FE?	YES	YES	YES	YES

Table 5: Association Between Black Ownership and VC Funding Raised (Instrumental Variables)

This table presents coefficients from 2SLS regressions run at the startup level, with standard errors reported below in parentheses. Our sample comprises companies founded after 2001 that raised external financing and filed a patent application with the USPTO, and whose founders have Linkedin profiles with pictures that allow us to identify race. The dependent variables, *Ln(VC Funding)* and *Ln(Non-VC Funding)*, are the cumulative amount of VC and non-VC funding raised in the five years following the patent application, respectively. Non-VC funding is funding raised from angel investors, accelerators, equity crowdfunding, grants, the Small Business Innovation Research (SBIR) program, and the Small Business Technology Transfer (STTR) program. The instrument is *P(Skin Tone)*, the probability that a founder is Black, as computed by an image-processing algorithm. *P(Black)* is the proportion of founders (where we could identify race) that are Black, and *P(Female)* is the proportion of founders that are female. *Ln(Count Inventors)* is the log of the number of inventors listed on the patent application. *Firm Age* is the number of years since the company was formed. *P(Founder-Inventor)* is the proportion of founders that are also inventors (listed on the patent application), and *Ln(Count Founders)* is the log of the number of founders of the startup. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by startup.

Dependent Variable:	**Ln(VC Funding)**		**Ln(Non-VC Funding)**	
	(1)	(2)	(3)	(4)
P(Black)	-0.807***	-0.693***	0.076	0.093
	(0.240)	(0.229)	(0.059)	(0.059)
P(Female)		-0.276***		0.055***
		(0.072)		(0.016)
Ln(Count Inventors)		0.299***		-0.015***
		(0.034)		(0.006)
Ln(Firm Age (yrs))		0.003		-0.005*
		(0.014)		(0.003)
P(Founder-Inventor)		-0.214***		0.011
		(0.047)		(0.009)
Ln(Count Founders)		0.568***		0.056***
		(0.045)		(0.008)
Observations	5421	5421	5421	5421
Adjusted R^2	-0.01	0.05	-0.01	-0.00
Cragg-Donald Wald F	6228.03	6231.44	6228.03	6231.44
State, Year, and Industry FE?	YES	YES	YES	YES

Table 6: Association Between Black Ownership and VC Funding Raised (Step Up Test)

This table presents coefficients from OLS regressions run at the investor-deal-year level, with standard errors reported below in parentheses. Our sample comprises startups that applied for a patent between 2001 and 2017, and whose founders have Linkedin profiles with pictures that allow us to identify race. We pair each startup with the lead investor in the deal. For this test, we restrict our sample to the first three deals for each startup. The dependent variable, $\frac{\textit{Deal Size Current Round}}{\textit{Deal Size Previous Round}}$, is the ratio of the amount of VC funding raised in the current deal to the amount of VC funding raised in the previous funding round. *P(Black)* is the proportion of founders (where we could identify race) that are Black, and *P(Female)* is the proportion of founders that are female. *I(Round Two)* is an indicator that equals one for a startup's second round of funding. *Age (yrs)* is the log number of years since the company was formed. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by investor-startup pair.

Dependent Variable:	$\frac{\textbf{Deal Size Current Round}}{\textbf{Deal Size Previous Round}}$			
	(1)	(2)	(3)	(4)
P(Black)	8.786*		8.537	8.403
	(5.218)		(5.273)	(5.241)
P(Female)		-1.498*	-1.395*	-1.951**
		(0.770)	(0.780)	(0.899)
P(Black) x I(Round Two)			1.524	
			(2.795)	
P(Female) x I(Round Two)				2.249*
				(1.356)
I(Round Two)			0.294	0.169
			(0.391)	(0.410)
Ln(Previous Round Size)	-2.842***	-2.894***	-2.875***	-2.890***
	(0.303)	(0.310)	(0.311)	(0.312)
Ln(Firm Age (yrs))	0.880**	0.994**	0.938**	0.928**
	(0.365)	(0.385)	(0.375)	(0.374)
Observations	715	715	715	715
Adjusted R^2	0.453	0.449	0.454	0.449
Investor FE?	YES	YES	YES	YES
Deal-Year FE?	YES	YES	YES	YES

Table 7: Association Between Black Ownership and Exits

This table presents coefficients from OLS regressions run at the startup level, with standard errors reported below in parentheses. Our sample comprises startups that applied for a patent between 2001 and 2017, and whose founders have Linkedin profiles with pictures that allow us to identify race. The dependent variable in Panel A, *I(M&A)*, is an indicator that equals one if the startup exited via an acquisition as of Q2 2021. In Panel B, *I(Failure)* is an indicator that equals one if the startup failed or went bankrupt as of Q2 2021. *P(Black)* is the proportion of founders (where we could identify race) that are Black, and *P(Female)* is the proportion of founders that are female. *Ln(Count Inventors)* is the log of the number of inventors listed on the patent application. *Firm Age* is the number of years since the company was formed. *P(Founder-Inventor)* is the proportion of founders that are also inventors (listed on the patent application), and *Ln(Count Founders)* is the log of the number of founders of the startup. *Ln(Other Funding)* is the total funding raised from non-VC investors in the five years following the patent application. *Ln(VC funding)* is the log of the total amount of funding raised from VCs in the five years following the patent application. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by startup.

Panel A:		**I(M&A)**		
	(1)	(2)	(3)	(4)
P(Black)	0.038	0.033	0.034	0.035
	(0.056)	(0.055)	(0.054)	(0.054)
P(Female)		-0.045**	-0.040*	-0.039*
		(0.020)	(0.020)	(0.020)
Ln(Count Inventors)		0.042***	0.042***	0.041***
		(0.009)	(0.009)	(0.009)
Ln(Firm Age (yrs))		0.005	0.006*	0.006*
		(0.004)	(0.004)	(0.004)
P(Founder-Inventor)		-0.039***	-0.038***	-0.037***
		(0.014)	(0.014)	(0.014)
Ln(Count Founders)		-0.026**	-0.021*	-0.022*
		(0.011)	(0.012)	(0.012)
Ln(Other Funding)			-0.064***	-0.064***
			(0.008)	(0.008)
Ln(VC Funding)				0.002
				(0.004)
Adjusted R^2	0.039	0.045	0.049	0.049
Panel B:		**I(Failure)**		
P(Black)	0.027	0.019	0.020	0.009
	(0.048)	(0.048)	(0.048)	(0.047)
P(Female)		0.041**	0.043**	0.037*
		(0.021)	(0.021)	(0.020)
Ln(Count Inventors)		0.012	0.012	0.019**
		(0.008)	(0.008)	(0.008)
Ln(Firm Age (yrs))		-0.016***	-0.016***	-0.016***
		(0.003)	(0.003)	(0.003)
P(Founder-Inventor)		0.016	0.017	0.011
		(0.011)	(0.011)	(0.011)
Ln(Count Founders)		-0.039***	-0.037***	-0.023**
		(0.010)	(0.010)	(0.010)
Ln(Other Funding)			-0.035***	-0.032***
			(0.007)	(0.008)
Ln(VC Funding)				-0.023***
				(0.003)
Observations	5410	5410	5410	5410
Adjusted R^2	0.022	0.030	0.032	0.041
State, Year, Industry FE?	YES	YES	YES	YES

Table 8: Association Between Black Ownership and Exits (Exit Value)

This table presents coefficients from Poisson regressions run at the startup level, with standard errors reported below in parentheses. Our sample comprises startups that applied for a patent between 2001 and 2017, and whose founders have Linkedin profiles with pictures that allow us to identify race. The dependent variable, *Exit Value*, is the value of the startup at exit. If the startup exits via an IPO, it is the market capitalization at IPO less the IPO proceeds, and if the startup exits via an acquisition, it is acquisition value. *P(Black)* is the proportion of founders (where we could identify race) that are Black, and *P(Female)* is the proportion of founders that are female. *Ln(Count Inventors)* is the log of the number of inventors listed on the patent application. *Firm Age* is the number of years between when the company was formed and when it applied for a patent. *P(Founder-Inventor)* is the proportion of founders that are also inventors (listed on the patent application), and *Ln(Count Founders)* is the log of the number of founders of the startup. *Ln(Other Funding)* is the total funding raised from non-VC investors in the five years following the patent application. *Ln(VC funding)* is the log of the total amount of funding raised from VCs in the five years following the patent application. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by startup.

Dependent Variable:	**Exit Value**			
	(1)	(2)	(3)	(4)
P(Black)	1.219	1.250	1.258	1.802*
	(1.012)	(0.803)	(0.804)	(0.952)
P(Female)		-1.178**	-1.142**	-1.043**
		(0.488)	(0.489)	(0.470)
Ln(Count Inventors)		0.473***	0.463***	0.254**
		(0.127)	(0.125)	(0.117)
Ln(Firm Age (yrs))		0.365***	0.368***	0.356***
		(0.085)	(0.086)	(0.085)
P(Founder-Inventor)		-1.046***	-1.024***	-0.825***
		(0.337)	(0.335)	(0.281)
Ln(Count Founders)		0.412*	0.433*	0.130
		(0.242)	(0.244)	(0.253)
Ln(Other Funding)			-0.495*	-0.532**
			(0.263)	(0.261)
Ln(VC Funding)				0.498***
				(0.101)
Observations	5229	5229	5229	5229
Log-likelihood	-274352.39	-246816.57	-245785.45	-218920.81
State, Year, Industry FE?	YES	YES	YES	YES

Table 9: Association Between Black Ownership and VC Funding Raised (Investor Heterogeneity)

This table presents coefficients from OLS regressions run at the investor-deal-year level, with standard errors reported below in parentheses. Our sample comprises startups that applied for a patent between 2001 and 2017, and whose founders have Linkedin profiles with pictures that allow us to identify race. We pair each startup with the lead investor in the deal. The number of observations is lower in this table because we only keep the set of startups for whom we can identify a lead investor. Note that *all* startups in these tests raised some VC funding. The dependent variable, *Ln(VC Funding)*, is the total amount of VC funding raised in each deal. *P(Black)* is the proportion of founders (where we could identify race) that are Black, and *P(Female)* is the proportion of founders that are female. *Ln(Count Inventors)* is the log of the number of inventors listed on the patent application. *Firm Age* is the number of years between when the company was formed and when it applied for a patent. *P(Founder-Inventor)* is the proportion of founders that are also inventors (listed on the patent application), and *Ln(Count Founders)* is the log of the number of founders of the startup. *I(High Success)* is an indicator that equals one if the lead investor had an above-median number of exits in the year before the deal. *I(High Experience)* is an indicator that equals one if the lead investor had an above-median number of investments in the year before the deal. *I(Funded Black Company)* is an indicator that equals one if the lead investor funded at least one other Black-owned company in the year before the deal. *I(Black VC)* is an indicator that equals one if the lead investor is a Black-owned venture capital group. We define a group as Black-owned if at least 50% of its founders and senior partners are Black. *Age (yrs)* is the log number of years since the company was formed. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by investor-startup pair.

Dependent Variable:	**Ln(VC Funding)**			
	(1)	(2)	(3)	(4)
P(Black)	0.388	0.350	-0.144	-0.403
	(0.426)	(0.426)	(0.261)	(0.287)
P(Black) x I(High Success)	-1.054**			
	(0.517)			
P(Black) x I(High Experience)		-1.013**		
		(0.514)		
P(Black) x I(Funded Black Company)			-2.122***	
			(0.627)	
P(Black) x I(Black VC)				0.873*
				(0.513)
I(High Success)	0.361***			
	(0.069)			
I(High Experience)		0.298***		
		(0.065)		
I(Black VC)				-0.083
				(0.195)
I(Funded Black Company)			0.110	
			(0.119)	
P(Female)	-0.296**	-0.295**	-0.271**	-0.272**
	(0.124)	(0.125)	(0.125)	(0.125)
Ln(Firm Age (yrs))	0.044**	0.046**	0.044**	0.044**
	(0.018)	(0.018)	(0.018)	(0.018)
Observations	3161	3161	3161	3161
Adjusted R^2	0.185	0.182	0.176	0.175
State, Industry, Deal Year FE?	YES	YES	YES	YES

REFERENCES

Arrow, K.J., 1998. What has economics to say about racial discrimination? Journal of economic perspectives 12, 91–100.

Bates, T., Bradford, W.D., 1992. Factors affecting new firm success and their use in venture capital financing. The Journal of Entrepreneurial Finance 2, 23–38.

Bates, T., Bradford, W.D., Jackson, W.E., 2018. Are minority-owned businesses underserved by financial markets? evidence from the private-equity industry. Small Business Economics 50, 445–461.

Becker, G.S., 1993. Nobel lecture: The economic way of looking at behavior. Journal of political economy 101, 385–409.

Becker, G.S., 2010. The economics of discrimination. University of Chicago press.

Bernstein, S., Giroud, X., Townsend, R.R., 2016. The impact of venture capital monitoring. The Journal of Finance 71, 1591–1622.

Bohren, J.A., Imas, A., Rosenberg, M., 2019. The dynamics of discrimination: Theory and evidence. American economic review 109, 3395–3436.

Cohn, J.B., Liu, Z., Wardlaw, M., 2021. Count data in finance. Available at SSRN .

Cook, L., Yang, Y., 2019. Missing Women and African Americans, Innovation, and Economic Growth. Mimeo. Michigan State University, East Lansing.

Egan, M.L., Matvos, G., Seru, A., 2017. When Harry fired Sally: The double standard in punishing misconduct. Technical Report. National Bureau of Economic Research.

Ewens, M., Townsend, R.R., 2020. Are early stage investors biased against women? Journal of Financial Economics 135, 653–677.

Fairlie, R., Robb, A., Robinson, D.T., 2021. Black and white: Access to capital among minority-owned start-ups. Management Science .

Fairlie, R., Robb, A., Robinson, D.T., 2022. Black and white: Access to capital among minority-owned start-ups. Management Science 68, 2377–2400.

Farre-Mensa, J., Hegde, D., Ljungqvist, A., 2020. What is a patent worth? evidence from the us patent "lottery". The Journal of Finance 75, 639–682.

Forscher, P.S., Cox, W.T., Brauer, M., Devine, P.G., 2019. Little race or gender bias in an experiment of initial review of nih r01 grant proposals. Nature human behaviour 3, 257–264.

Ginther, D.K., Basner, J., Jensen, U., Schnell, J., Kington, R., Schaffer, W.T., 2018. Publications as predictors of racial and ethnic differences in nih research awards. PLoS One 13, e0205929.

Ginther, D.K., Schaffer, W.T., Schnell, J., Masimore, B., Liu, F., Haak, L.L., Kington, R., 2011. Race, ethnicity, and nih research awards. Science 333, 1015–1019.

Gompers, P.A., Gornall, W., Kaplan, S.N., Strebulaev, I.A., 2020. How do venture capitalists make decisions? Journal of Financial Economics 135, 169–190.

Gornall, W., Strebulaev, I.A., 2020. Gender, race, and entrepreneurship: A randomized field experiment on venture capitalists and angels. Available at SSRN .

Hadlock, C.J., Pierce, J.R., 2010. New evidence on measuring financial constraints: Moving beyond the kz index. The review of financial studies 23, 1909–1940.

Hebert, C., 2020. Gender stereotypes and entrepreneur financing, in: 10th Miami Behavioral Finance Conference.

Howell, S.T., 2017. Financing innovation: Evidence from r&d grants. American economic review 107, 1136–64.

Howell, S.T., Nanda, R., 2019. Networking frictions in venture capital, and the gender gap in entrepreneurship. Technical Report. National Bureau of Economic Research.

Hunt, J., Garant, J.P., Herman, H., Munroe, D.J., 2013. Why are women underrepresented amongst patentees? Research Policy 42, 831–843.

Pager, D., 2016. Are firms that discriminate more likely to go out of business? Sociological Science 3, 849–859.

Park, Y., Coleman, S., 2009. Credit rationing and black-owned firms: is there evidence of discrimination? Journal of Developmental Entrepreneurship 14, 255–271.

Rose, E.K., 2022. A constructivist perspective on empirical discrimination research. Journal of Economic Literature .

Small, M.L., Pager, D., 2020. Sociological perspectives on racial discrimination. Journal of Economic Perspectives 34, 49–67.

Smith, B., Viceisza, A., 2018. Bite me! abc's shark tank as a path to entrepreneurship. Small Business Economics 50, 463–479.

Younkin, P., Kuppuswamy, V., 2018. The colorblind crowd? founder race and performance in crowdfunding. Management Science 64, 3269–3287.

Funding Black Innovators

Internet Appendix

Table of Contents

This Internet Appendix contains supplementary discussions and analyses, which are organized as follows:

1. VI discusses how we go from the original USPTO application data to our final sample
2. VII discusses how we merge data from the SBIR and STTR program to our data.
3. A.3 repeats the analysis in table 3 using the sample in Farre-Mensa et al. (2020).
4. A.1 repeats the analysis in table 2 using the sample in Farre-Mensa et al. (2020).
5. A.2 repeats the analysis in table 2 at a startup-year level, where the panel is a balanced panel that runs from the year the startup first applied for a patent to five years following its patent application.

VI. Public Pair Patent Application Data

A. Applications by Small Entities

We start our sample construction using the publicly available patent data from the United States Patent and Trademark Office. Our dataset uses applications from January 2001 to December 2017 to allow enough time for the patent to have been granted. In line with the patenting literature, we focus our attention on regular utility patent applications, which account for over 90% of all patents. Given our focus on small businesses, we further restrict our attention of applications by small entities by keeping applications by assignees that the USPTO refers to as SMALL or MICRO. The resulting data has about 1,340,069 patent-applications.

For these applications we only keep the assignee that was first to be assigned the patent, then we keep the first patent application by each assignee. In mapping applications to assignees, we keep only applications where the employer assignment indicator is one, which denotes the USPTO's best guess of an assignment of the patent by an employee to the employer. We also restrict the set of application-assignee pairs to assignments that were executed within 60 days of the patent application and incorporated assignees based in the United States. After this filter, we have 108,909 applications by 108,909 assignees filed between 2001 and 2017. To obtain an equal number of first-time applicants, when there are multiple applications for the same assignee, we keep the application that was granted or with the largest number of citations.

Note that the number of first-time assignees is larger than in Farre-Mensa et al. (2020) because, unlike that paper, we do have access to all applications prior to 2001, we do not screen out already public companies, and we do not screen out non-for-profit organizations.

For each application, we have the inventor(s), the date of the application, the final outcome of the application, the assignee, and the primary and secondary technology classification to which the application has been assigned.

B. *Merging Micro-Entity Application Assignees to PitchBook*

To match an assignee to PitchBook, we match assignees in the first step to PitchBook on name and state where the startup is located. [13] Given our focus on cross-sectional variation in access to venture capital for our innovative firms, we further require that each company be formed no more than five years before the first patent application. We matched 18,595 applications by 18,595 assignees to LinkedIn following this procedure.

C. *Isolating founder-Investors*

Of the 18,595 applications, PitchBook has data on the LinkedIn links of the founders of the assignees for 11,620 applications representing 11,620 assignees.[14]

D. *Identifying founder-race*

To identify founder race, we use the founder's profile picture on LinkedIn. We have links for the public LinkedIn profiles for 20,916 founders of the 11,620 assignees. Of these founders, we have pictures for 12,965 founders, representing 6,290 assignees, and 6,290 applications filed between 2001 and 2017.

E. *Final Sample of Applications by Race*

We use Google's vision API to train a model to identify race given a picture. We create the initial training data by manually classifying 1,000 pictures using four labels, black, female, male, and other. We use a combination of pictures, names, and self-reported

[13] We use the fuzzywuzzy python package for name matching and require a 99 percent match rate.

[14] We identified founders by keeping only those employees of the assignees we matched to LinkedIn with the keyword "founder," or "owner" in their title.

race (from Bios on the company's website and various news sources) for our manual classification. We then deployed the model on all the 12,965 images we have in our sample. The algorithm used 88 percent of the images for training and 12 percent for testing. The precision and recall in the test sample is 96.88%. Type II errors mainly came from classifying light-skin blacks as other, and Type I errors come from classifying darker skin people of other races as black. For each image, we have a predicted probability that the person is black or female. We are able to classify race for 10,161 applications in our sample.[15]

Table 1 compares the characteristics of applications for which we could classify race of at least one assignee founder to all small entity applications from 2001 to 2017. We can see that the subset of applications we use for our analysis differs little from the population of applications by small entities.

VII. Fundraising Data

In this section we discuss how we matched applicants in our sample to fundraising data from the SBIR/STTR program.

A. SBIR Program

According to Howell (2017), the SBIR program was established in 1982 by congress. The program requires 12 federal agencies to set-aside 3.2 percent of their R& D budgets to award to small businesses.[16]

We download all award data as of 2020 from the SBIR's website. We then use the

[15] We lost founders representing about 1,000 applications because the algorithm was not able to process some low-quality images that we downloaded from LinkedIn. We used the Python Rest API to deploy the model on our sample of images.

[16] Participating agencies comprise the Department of Defense, the Department of Health and Human Services, the Department of Energy, the National Science Foundation, the National Aeronautics and Space Administration, the U.S. Department of Agriculture, the Department of Homeland Security, the Department of Commerce, the Department of Education, the Department of Transportation, the Department of Commerce, and the Environmental Protection Agency.

name of each recipient and the state where they are located to match the recipent to our list of assignees in our matched USPTO-PitcBook sample.[17]

[17] We use the fuzzywuzzy python package for name matching and require a 99 percent match rate.

Table A.1: Association Between Black Ownership and Funding Raised (Same Sample as Farre-Mensa et al. (2020))

This table presents coefficients from Poisson regressions run at the assignee level, with standard errors reported below in parentheses. We use the same sample that is used in Farre-Mensa et al. (2020). The dependent variable in Panel A, *VC Funding*, is the cumulative amount of VC funding the startup raised 2, 3, 4, or 5 years following the patent application. In Panel B, the dependent variable is *Non-VC Funding* the total amount of non-VC funding raised 2, 3, 4, or 5 years following the patent application. Non-VC funding is funding raised from angel investors, accelerators, equity crowdfunding, grants, the Small Business Innovation Research (SBIR) program, and the Small Business Technology Transfer (STTR) program. *P(Black)* is the proportion of founders (where we could identify race) that are black, and *P(Female)* is the proportion of founders that are female. To predict inventor race in this sample, we first use Name Prism to predict the probability that a founder is black. Then we downloaded and processed all images (using the same procedure that we used in the main analysis) where the probability that the inventor is black is 40% or greater. We use the Genderize.IO algorithm to predict gender. *Ln(Count Investors)* is the log count of the number of inventors listed on the patent application. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by startup.

Panel A:	**VC Funding**			
Dependent Variable:	Next 2yrs	Next 3yrs	Next 4yrs	Next 5yrs
P(Black)	-2.261	-2.776*	-2.642*	-2.856**
	(1.463)	(1.623)	(1.419)	(1.441)
P(Female)	-0.977*	-0.674	-0.675*	-0.376
	(0.541)	(0.435)	(0.404)	(0.484)
Ln(Count Inventors)	0.735***	0.699***	0.725***	0.663***
	(0.095)	(0.084)	(0.076)	(0.091)
Log-likelihood	-87619.36	-105571.02	-123939.25	-149849.93
Panel B:	**Non-VC Funding**			
Dependent Variable:	Next 2yrs	Next 3yrs	Next 4yrs	Next 5yrs
P(Black)	0.050	0.080	0.156	0.148
	(0.070)	(0.101)	(0.178)	(0.182)
P(Female)	0.006	-0.002	0.001	-0.000
	(0.013)	(0.030)	(0.031)	(0.033)
Ln(Count Inventors)	0.027***	0.035***	0.041***	0.047***
	(0.006)	(0.008)	(0.009)	(0.010)
Observations	33964	33964	33964	33964
Log-likelihood	-25200.65	-61214.70	-62724.37	-64639.86
Fixed Effects?	Class X Year	Class X Year	Class X Year	Class X Year

Table A.2: Association Between Black Ownership and Funding Raised (Panel regression)

This table presents coefficients from Poisson regressions run at the assignee-year level, with standard errors reported below in parentheses. We keep only the first patent application for each assignee. Where there are multiple applications for the same assignee, we keep the application that was grated or with the largest number of inventors listed. Our sample comprises assignees that applied for a patent between 2001 and 2017, and whose founders have Linkedin profiles with pictures that allow us to identify race. The dependent variable in Panel A, *Total Funding*, is the total amount of funding the startup raised each year following its founding and the minimum of Q2 2021 or five years following the patent application. In Panel B, the dependent variable is *VC Funding* the total amount of VC funding raised each year. In Panel C, the dependent variable is *Non-VC Funding* the total amount of non-VC funding raised each year. Non-VC funding is funding raised from accelerators, equity crowdfunding, grants, and from the Small Business Innovation Research (SBIR) program. *P(Black)* is the proportion of founders (where we could identify race) that are black, and *P(Female)* is the proportion of founders that are female. *P(Founder-Inventor)* is the proportion of founders that are also inventors (listed on the patent application). *Ln(Count Investors)* is the log count of the number of inventors listed on the patent application, and *Ln(Age)* is the log number of years since the company was formed. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by startup.

Panel A:		**VC Funding**		
P(Black)	-1.606***	-1.585***	-1.478***	-1.484***
	(0.404)	(0.396)	(0.402)	(0.463)
P(Female)		-0.585***	-0.526***	-0.544***
		(0.160)	(0.162)	(0.179)
Ln(Count Inventors)				0.365***
				(0.066)
Firm Age (yrs)				0.057**
				(0.029)
P(Founder-Inventor)				-0.377***
				(0.106)
Ln(Count Founders)				0.608***
				(0.093)
Log-likelihood	-187353.88	-186829.13	-177169.45	-171212.94

Panel B:		**Non-VC Funding**		
P(Black)	-0.296	-0.305	-0.258	-0.154
	(0.434)	(0.434)	(0.469)	(0.520)
P(Female)		0.450*	0.395*	0.461**
		(0.263)	(0.204)	(0.219)
Ln(Count Inventors)				0.232
				(0.168)
Firm Age (yrs)				0.097**
				(0.043)
P(Founder-Inventor)				0.154
				(0.157)
Ln(Count Founders)				0.660***
				(0.152)
Observations	32328	32328	32220	32220
Log-likelihood	-10531.39	-10511.87	-9107.18	-8946.23
State FE?	NO	NO	NO	YES
Industry FE?	NO	NO	YES	YES
Year FE?	YES	YES	YES	YES

Table A.3: Association Between Black Assignees and Patent Outcomes (Same Sample as Farre-Mensa et al. (2020))

This table reports the estimated effect of the proportion of black and female founders assigned a patent application on the likelihood the patent is granted, abandoned, the time between application and grant, and the number of citations the patent received. We use the same sample that is used in Farre-Mensa et al. (2020). The dependent variable in Column (1), *I(Granted)*, is an indicator that equals one for startups whose patent application was ultimately successful. In Column (2), it is an indicator for whether the patent application was abandoned, *I(Abandoned)*. In Column (3), it is the number of citations the patent received as of 2017, (*Citations*), and in Column (4), it is the number of years between patent application and grant, *Years to Grant*. *P(Black)* is the proportion of founders (where we could identify race) that are black, and *P(Female)* is the proportion of founders that are female. To predict inventor race in this sample, we first use Name Prism to predict the probability that a founder is black. Then we downloaded and processed all images (using the same procedure that we used in the main analysis) where the probability that the inventor is black is 40% or greater. We use the Genderize.IO algorithm to predict gender. *Ln(Count Investors)* is the log count of the number of inventors listed on the patent application. *Class X Year* denotes USPC class by application year fixed-effects. We estimate Columns (1) and (2) using linear probability models, and Columns (3) and (4) using quasi-maximum likelihood Poisson models with robust standard errors. ^{g} denotes tests on the subset of applications that were granted. $^{***}p < 0.01$ denotes significance at the 1% level, $^{**}p < 0.05$ denotes significance at the 5% level, and $^{*}p < 0.10$ denotes significance at the 10% level. We cluster standard errors by application.

Dependent Variable:	I(Granted)	I(Abandoned)	Citations^{g}	Years to Grant^{g}
P(Black)	0.086	-0.086	0.186	0.061
	(0.076)	(0.076)	(0.228)	(0.081)
P(Female)	-0.023*	0.023*	-0.084	0.031*
	(0.014)	(0.014)	(0.067)	(0.018)
Ln(Count Inventors)	0.012***	-0.012***	0.237***	0.029***
	(0.005)	(0.005)	(0.026)	(0.005)
Fixed Effects	Class X Year	Class X Year	Class X Year	Class X Year
Observations	33964	33964	21672	21672
Log-likelihood	-19754.71	-19754.71	-182000.75	-37106.48