Discrimination in Hiring: Evidence from Retail Sales*

Alan Benson†, Simon Board‡, and Moritz Meyer-ter-Vehn§

June 30, 2022

Abstract

We propose a simple model of racial bias in hiring that encompasses three major theories: taste-based discrimination, screening discrimination, and complementary production. We derive a test that can distinguish these theories using the mean and variance of workers' productivity under managers of different pairs of races. We apply this test to study discrimination at a major U.S. retailer using data from around 50,000 newly-hired commission-based salespeople. White, black and Hispanic managers within the same store are significantly more likely to hire workers of their own race, consistent with all three theories. For white-Hispanic and black-Hispanic pairs, mean productivity is greater and productivity variance is lower when manager and worker races match, driven by greater productivity in lower-tail of the productivity distribution. For black-Hispanic pairs the evidence indicates screening discrimination; for white-Hispanic pairs it indicates some combination of screening discrimination and complementary production. Our tests for white-black pairs are less clear.

1 Introduction

More than fifty years after the U.S. prohibited employers from using race as a factor in employment, the persistence of racialized labor market outcomes—including segregation and pay gaps—remains among the most striking features of the labor market. Controlling for education and experience, black and Hispanic workers have significantly lower employment rates and wages than white workers, especially for less-educated workers (Altonji and Blank, 1999; Ritter and Taylor, 2011; Lang and

---

*Keywords: discrimination, screening, hiring. JEL codes: J71, M51. We thank Aislinn Bohren, Adriana Lleras-Muney, Mandy Pallais, Andres Santos and seminar participants at the AEAs, Minnesota, MIT, NBER, Penn, and UCLA. We also thank the W.E. Upjohn Institute for Employment Research for support.
†University of Minnesota, https://carlsonschool.umn.edu/faculty/alan-benson
‡UCLA, http://www.econ.ucla.edu/sboard/
Lehmann, 2012). Understanding the causes of racial disparities is important for policymakers and managers who wish to promote equality and enhance productivity.

In this paper, we study how workers' outcomes depend on the race of their hiring manager, known as relative employment discrimination. We then seek to understand why such discrimination arises by studying three major theories. Taste-based discrimination (Becker, 1957) is the simplest, and perhaps the presumptive, theory of discrimination. If employers have an intrinsic taste for hiring same-race workers then they lower the hiring bar for their own race, leading to labor market segregation and wage disparities (e.g. Black, 1995). Screening discrimination (Cornell and Welch, 1996), though less prominent in the literature, poses an alternative theory of racialized labor market outcomes. This model adapts Phelps's (1972) classic model of statistical discrimination to allow employers to be better at screening same-race applicants.¹ The third theory of discrimination is complementary production (e.g. Lang, 1986) in which workers are more productive when working with managers of the same race, leading to higher rates of same-race hiring.² All three theories are consistent with a wide variety of empirical facts (e.g. labor market segregation, pay gaps, and differences in employment rates), but call for different remedies. For example, racial quotas raise productivity under taste-discrimination, as biased managers hire more qualified minorities, but lower it under screening discrimination, as uninformed managers hire mismatched minorities. Conversely, algorithmic hiring tools that provide managers information about candidates improve diversity and productivity under screening discrimination, but not under taste-based discrimination.

This paper proposes a common framework to distinguish and test for these three theories of discrimination. In the model, managers observe a noisy signal of applicants' productivity and hire those whose expected productivity lies above a threshold. With taste-based discrimination, managers adopt lower thresholds when evaluating same-race applicants. With screening discrimination, managers observe a more precise signal when evaluating same-race applicants. And with complementary production, managers obtain higher output when employing a same-race worker.

All three theories predict that managers hire more same-race applicants, although for different reasons. With taste-based discrimination, managers lower their standards for same-race candidates; with screening discrimination, managers receive stronger signals of ability from same-race candidates; with complementary production, managers obtain higher output for a given worker if they share the manager's race. All three theories also predict same-race workers have lower turnover. With taste-based discrimination, managers treat same-race workers more leniently; with screening discrimination, managers make fewer mistakes in hiring same-race workers; with complementary production, same-race workers are less likely to be marginal hires and at risk of being fired. Thus

¹Statistical discrimination captures the idea that firms have more accurate signals of majority applicants. Screening discrimination captures the idea that managers obtain more accurate signals from same-race applicants.
²By “complementarity production” we mean that workers are more productive when working with others of the same race, so the most productive assignment is to match people with the same race. In contrast, acts of “favoritism”, like assigning better shifts to same-race workers, are zero sum.
one cannot separate between the three theories using hiring or turnover data.

In contrast, the three theories make different predictions regarding the relative productivity of same-race and cross-race hires. Under taste-based discrimination, managers lower the hiring threshold for same-race workers, meaning that the productivity of same-race hires has a lower mean and higher variance. Under screening discrimination, managers can screen same-race workers with greater accuracy, so the productivity distribution of same-race hires has a higher mean and lower variance. And under complementary production, productivity of same-race workers shifts up, so the productivity distribution of same-race hires has a higher mean and higher variance. These theoretical results are summarized in Table 1.

Using these predictions, we derive a test for discrimination in a model that nests the three theories and allows for a small amount of manager and worker heterogeneity (e.g. different races of workers may have different means, and different races of managers may have different hiring thresholds). Our test is based on the sub/supermodularity of the productivity mean and variance. If mean productivity is submodular in manager-worker race, then there exists taste-based discrimination. If log productivity variance is submodular, then there exists screening discrimination. And if mean productivity and log productivity variance are both supermodular, then there exists complementary production. The first two results follow immediately from Table 1 but the third is more subtle: No combination of the other forms of discrimination can produce both types of supermodularity simultaneously.

Our test is conservative, in that it avoids false positives, but can lead to false negatives. For example, it may be that all three forces are present to some extent, but we identify none because they cancel one another out. Indeed, we can identify at most two forces from the same data set.

Our test is also based on comparing managers of different races. Our test would fail to pick up, say, taste-based discrimination if all managers were biased against a particular race of worker.

We apply our model to longitudinal administrative data from a large national retailer from 2009.

---

3 If \( \bar{y}_{ij} \) is the mean productivity of race-\( i \) workers hired by race-\( j \) managers, then \( \bar{y} \) is supermodular if \( \bar{y}_{ii} + \bar{y}_{jj} \geq \bar{y}_{ij} + \bar{y}_{ji} \), and submodular if the inequality is reversed. Similarly, if \( v_{ij} \) is the productivity variance of race-\( i \) workers hired by race-\( j \) managers, then \( v \) is log-supermodular if \( \log v_{ii} + \log v_{jj} \geq \log v_{ij} + \log v_{ji} \), and log-submodular if the inequality is reversed.
to 2015. Each store typically employs several hiring managers of different races, allowing us to examine the difference in their behavior after controlling for store, department and month fixed effects. The data include the races of 48,755 white, black, and Hispanic newly-hired salespeople and the races of their 7,892 hiring managers at 997 store locations over 74 months. The data further include sales versus targets, the performance measure used to determine commissions.

First, we show that the probability that a hire is a given race is 3.6pp higher when the manager shares the same race, after controlling for fixed effects. This gap is positive and statistically significant for all same-race pairs, and represents a particularly large increase for minority workers since they account for a much smaller percentage of hires. For example, a Hispanic worker is 25.5% more likely to be hired if the hiring manager is also Hispanic. These results are consistent with all three theories.

Next, we use the productivity data to shed light on which of the three theories causes this discrimination in hiring. The cleanest evidence comes from black-Hispanic pairs, where the mean gap between same-race and cross-race productivity is 1.9pp while the same-race variance is 84% of the cross-race variance. We reject the null that there is no screening discrimination at the 1% level. Indeed, at the bottom of the distribution, the gap between same-race and cross-race productivity is 10pp, which suggest managers eliminate low-productivity same-race workers. For white-Hispanic pairs the mean gap between same-race and cross-race productivity is 1.3pp while the same-race variance is 96% of the cross-race variance. We reject the null that there is neither screening discrimination nor complementary production. For white-black pairs the mean gap between same-race and cross-race productivity is 0.6pp while the same-race variance is 99% of the cross-race variance. This suggests a mixture of all three forces is present.

Our results should not be interpreted as saying that taste-based discrimination is absent or unimportant. Instead, we provide a methodology for distinguishing the three types of discrimination and show that screening discrimination seems to play an important role, especially when it comes to Hispanics. Thus, managers and policy makers should take screening discrimination into account when designing reforms (e.g. racial quotas, centralized hiring, and diverse hiring committees) that seek to identify the best applicants and reduce racial outcome gaps. Our results also suggest that the effectiveness and appropriateness of interventions also vary by minority group.

The paper ends with a number of extensions. Under a symmetry assumption, we show that we can use the three pairwise relation to identify the bias of each manager race. We briefly examine internal transfer data and discuss other possible theories (in particular, referral networks, favoritism, and worker preferences). And we study the long-term implications of our results for racial composition of the firm’s employees.

\footnote{For linguistic simplicity, we refer to white, black, and Hispanic all as “races.”}
1.1 Literature

There is widespread evidence for discrimination in recruiting. For example, in an audit study, Bertrand and Mullainathan (2004) find that applicants with white-sounding names receive a 50% higher call-back rate than black-sounding names. Closer to our study, Giuliano, Leonard and Levine (2009, 2011) study hiring practices of managers of different races at a large chain of retail stores and find evidence of same-race bias with black and Hispanic employees (the latter in highly Hispanic locations). Our objective productivity measure allows us to go beyond these findings by partially identifying the cause of segregation.\(^5\)

Other studies have sought to test the three types of discrimination contained in our model. There is widespread evidence of taste-based discrimination. For instance, Hedegaard and Tyran (2018) find that Danish students discriminate in choosing co-workers based on students’ names, even after productivity information is disclosed. Looking at a fruit picking farm, Bandiera, Barankay, and Rasul (2009) show that managers assist socially connected workers at the expense of unconnected workers when paid a fixed wage, but not when paid a performance bonus.

A few recent papers have documented how the productivity of a workers depends on who they work with (“complementary production”). Within a Kenyan flower packing firm, Hjort (2014) shows that workers try to lower the performance of team members from different tribes, even at a cost to themselves. Closer to our paper, Glover, Pallais, and Pariente (2017) study a French grocery chain and show that minority cashiers are as productive as non-minority cashiers on average, but become less productive when working under managers who were deemed biased by an implicit association test. They write that “our evidence is most consistent with a theory in which biased managers interact less with minority workers.” This is less likely to be relevant in our setting since our managers hire the workers they supervise and develop long-term relationships.

There is also a literature that examines screening discrimination.\(^6\) Dustmann, Glitz, Schönberg, and Brücker (2016) using matched employee-employer data to show that firms in Germany with more minority employees hire more minority applicants and pay them slightly higher wages. They argue that these results come from referrals that provide information about match quality, which they

---

\(^5\)Comparing our Table 2 with Giuliano et al. (2009, Table 9), the black/white gap looks similar, but our white/Hispanic gap is larger. This may be because of different locations or demographics; e.g. in Giuliano et al. (2009, 2011), front-line workers are 70% female with an average age of 22, whereas our salespeople are 41% female with a median age of 32. Giuliano et al. have data on turnover and promotions, so one might hope one can use it as a proxy for productivity to identify the cause of discrimination. Our paper urges caution since these are decisions are made by potentially biased managers; indeed, in our model, all three theories share the same prediction for turnover.

\(^6\)There is also a large literature on statistical discrimination (whereby firms have more accurate signals of majority applicants). Altonji and Pierret (2001) study the correlation between wages and intelligence scores as workers age, and find mixed evidence. Ağan and Starr (2017) use audit methods to show that that “ban-the-box” laws that prohibit employers from asking about job applicants’ criminal histories can lead employers to reduce the call-back rate for black applicants. Our model allows for workers to be drawn from different distributions, as in a traditional model of statistical discrimination, but our test concerns relative discrimination and thus only identifies screening discrimination.
interpret through a model of screening discrimination. Relatedly, Åslund, Hensvik, and Skans (2014) show that immigrant managers in Sweden are more than twice as likely to hire immigrants compared to native managers. Immigrants are also paid higher wages when managers are also immigrants; this is mainly explained by individual worker effects, meaning that immigrant managers hire higher quality immigrant recruits, consistent with screening discrimination. Relative to these papers, our data concerns relatively homogeneous jobs at a single US employer; we then show how the mean and variance of productivity can be used to tease apart different theories of discrimination.  

Methodologically, we are closest to Anwar and Fang (2006) who show how relative comparisons can be used to overcome the infra-marginality problem associated with Becker’s (1957) “outcome test”. They study police officers’ decisions to stop and search a driver and propose a simple test of relative taste-based discrimination. Specifically, police officers are prejudiced if the search rate exhibits rank reversal (e.g. white officers search more black drivers, while black officers search more white drivers). The test makes few distributional assumptions (e.g. about the underlying populations or police officers’ private signals) but it is a rather weak test (e.g. if white police pull over many more people than black police then it is hard to detect bias). Our theory has a continuous outcome variable, which allows us to study both the mean and variance of productivity. This allows us to separate taste-based discrimination, screening discrimination and complementary production, whereas Anwar-Fang focus on taste-based discrimination. To study all these forces together we impose more parametric structure, in terms of normal distributions of priors and signals, and assume there is “small heterogeneity,” enabling us to use Taylor approximations. Given these assumptions, we say that managers are prejudiced if the mean output is submodular which is less demanding than Anwar-Fang’s rank reversal test.

The use of relative comparisons to evaluate discrimination has been used in other areas outside the employment domain. Ayres and Siegelman (1995) observe that black customers in car dealerships receive worse offers, no matter the race of the salesperson, which they interpret as evidence of statistical discrimination. Similarly, List (2004) finds that baseball card traders believe minority customers have a wider variance in willingness-to-pay for baseball cards, leading them to offer minorities less favorable quotes. Fisman, Paravisini, and Vig (2017) looks at Indian loan decisions and shows that, consistent with screening discrimination, same-group borrowers receive more loans, more variable loan terms, and have lower default rates (where a “group” corresponds to religion or caste). In comparison to our formulation, Fisman et al. have a continuous decision variable (loan terms) and a binary outcome variable (default), whereas we study the mean and variance of a continuous outcome variable (productivity).

Screening discrimination has also been studied in the psychology and communication. For example, in an ethnographic study of 185 Chicago-area firms, Neckerman and Kirschenman (1991) find evidence that “interviewing well goes beyond interpersonal skills to common understandings of appropriate interaction and conversational style—in short, shared culture.”
There are papers that seek to overcome the "inframarginality problem" in different ways. Knowles, Persico, and Todd (2001) develop a model of the decision to carry contraband, in which marginal and average success rates of police searches will coincide. Anwar and Fang (2015) use the timing of release decisions of parole boards. Arnold, Dobbie, and Yang (2018) and Arnold, Dobbie, and Hull (2022) use variation between judges in bail decisions to infer the marginal treatment effect via two different methods; the latter paper seeks to identify taste-based and statistical discrimination. The bail data has the advantage that one observes each defendant (whereas we only see hired workers) and each judge makes a large number of decisions about randomly assigned cases. However, it uses a binary outcome variable (pretrial misconduct), whereas we study the mean and variance of a continuous outcome variable (productivity).

Looking across this expansive literature, most prior studies of labor market discrimination aim to provide evidence for individual theories. However, the effect of different policies depends on which theory is dominant. Our contribution is to provide an intuitive method to identify three different types of discrimination in one common framework.

2 Theory

We first describe a baseline model of recruiting based on Phelps (1972) and use it to examine the predictions that result from taste-based discrimination, screening discrimination, and complementary production.\(^8\) We then enrich the model to allow for manager and worker heterogeneity and propose our empirical tests.

Managers screen applicants for sales positions. Motivated by our data (see Figure C1 in Online Appendix C), we assume the log of workers’ sales is normally distributed, \(y \sim N(\mu, \sigma_0^2)\) and refer to \(y\) as the worker’s productivity. A manager’s signal is given by \(\tilde{y} = y + \epsilon\), where the noise \(\epsilon \sim N(0, \sigma_\epsilon^2)\) is independent of \(y\). Using Bayes’ rule, the expected productivity of the applicant given the signal, the estimate \(\hat{y}\), is a weighted sum of the prior and the signal, where the weights reflect the signal variance \(\sigma_\epsilon^2\) and the cross-sectional productivity variance \(\sigma_0^2\),

\[
\hat{y} = E[y|\tilde{y}] = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2} \mu + \frac{\sigma_\epsilon^2}{\sigma_0^2 + \sigma_\epsilon^2} \tilde{y}.
\]

The estimate \(\hat{y}\) is a sufficient statistic of the signal \(\tilde{y}\) with a meaningful economic interpretation; so we work with \(\hat{y}\) instead of \(\tilde{y}\). As usual in Bayesian updating, it is important to distinguish between two variances. Ex-ante, the estimate \(\hat{y}\) is distributed normally with mean \(\mu\) and estimator variance \(\eta^2 := \frac{\sigma_\epsilon^2}{\sigma_0^2 + \sigma_\epsilon^2}\). Notably, the estimator variance \(\eta^2\) falls in the signal variance \(\sigma_\epsilon^2\). For example, with

---

\(^8\)With taste-based discrimination, a version of Proposition 1(a) was shown by Cornell and Welch (1996). With screening discrimination, Propositions 2(b),(d) are closely related to the analysis of Simon and Warner (1992) and Dustmann, Glitz, Schönberg, and Brücker (2016). We formalize these results and unify them in a common analytical framework.
a completely noisy signal, $\sigma^2 = \infty$, all workers have the same expected ability $y\hat{\;}$, so the estimator variance $\eta^2$ is zero. Ex-post, conditional on $y\hat{\;}$, the realized ability $y$ is distributed normally with mean $y\hat{\;}$ and residual variance $\gamma^2 := \frac{\sigma^2 \sigma^2}{\sigma^2 + \sigma^2 \epsilon}$. For example, with a completely noisy signal, $\sigma^2 = \infty$, the residual variance equals the cross-sectional variance of productivity $\sigma^2_0$.

For ease of exposition, we assume that managers have log-utility so wish to hire a worker if expected log-sales $y\hat{\;}$ exceeds a cutoff $y^\ast$. The hiring probability is

$$p := Pr[y\hat{\;} \geq y^\ast];$$

we also assume $y^\ast \geq \mu$, so at most half of all applicants are hired. Average productivity and productivity variance of accepted applicants are given by

$$\bar{y} := E[y|y\hat{\;} \geq y^\ast] \quad \text{and} \quad v := Var[y|y\hat{\;} \geq y^\ast].$$

Finally, we suppose the firm eventually learns the actual productivity $y$, and fires a worker if it is more than $\tau$ below the cutoff $y^\ast$. Turnover then equals

$$x := Pr[y \leq y^\ast - \tau|y\hat{\;} \geq y^\ast].$$

We want to understand the effects of taste-based discrimination, screening discrimination, and complementary production on these four outcome variables ($p, \bar{y}, v, x$). The proofs from this section are in Appendix A.

### 2.1 Taste-based Discrimination

Under taste-based discrimination, managers are biased towards their own race. Specifically, they apply a lower hiring threshold to same-race hires than cross-race hires, $y^\ast_s < y^\ast_c$. They then fire workers whose productivity drops $\tau$ below the respective hiring threshold.

**Proposition 1.** In the taste-based model:

(a) The hiring probability is higher for same-race applicants than cross-race applicants, $p_s > p_c$.

(b) Average productivity is lower for same-race hires than cross-race hires, $\bar{y}_s < \bar{y}_c$.

(c) Productivity variance is higher for same-race hires than cross-race hires, $v_s > v_c$.

(d) Turnover is lower for same-race hires than cross-race hires, $x_s < x_c$.

The intuition is illustrated in Figure 1a (left). Part (a) says that managers are biased toward workers of their own race, so lower the required standard, and have a higher hiring probability, $p_s > p_c$. Part (b) shows that since managers apply a lower standard to same-race workers, their average productivity is lower. Part (c) says that the lower same-race standard $y^\ast_s < y^\ast_c$ raises the productivity variance of same-race recruits as additional low-productivity workers are hired.
Specifically, by the law of total variance, we can decompose the productivity variance of recruits into the sum of estimator variance $\text{Var}(\hat{y}|\hat{y} \geq y^*_i)$ and residual variance $\text{Var}(y|\hat{y}) = \gamma^2$. The estimator variance is higher for same-race recruits, $\text{Var}(\hat{y}|\hat{y} \geq y^*_s) > \text{Var}(\hat{y}|\hat{y} \geq y^*_c)$, while the residual variance is equal for both races. Finally, part (d) says that turnover is lower for same-race hires. This result is somewhat counter-intuitive since same-race recruits have lower average productivity and higher variance. The result follows since managers apply the same bias when firing as when hiring. An agent is thus fired if the measurement error $\hat{y} - y$ exceeds the “safety margin” $\hat{y} - y^* + \tau$. Both same- and cross-race recruits have the same residual variance, but cross-race recruits have lower safety margins (in the likelihood ratio order). Indeed, in Figure 1a (left), one can see that cross-race recruits are closer to the hiring threshold than same-race recruits.

To illustrate the productivity results, Figure 1a (right) plots the difference between same-race and cross-race productivity by quantile. In the figure, quantile 0 corresponds to the difference between the worst same-race and cross-race recruits, while quantile 1 corresponds to the best recruits. One can see that the same-race recruits are less productive, as in part (b). This gap shrinks as the recruits get better since the best workers are hired no matter the race of the manager; this raises the variance for same-race hires, as in part (c).

### 2.2 Screening Discrimination

We now turn to screening discrimination, assuming signals are more accurate for same-race applicants than cross-race, so with signal variance $\sigma^2_s < \sigma^2_c$. Managers are unbiased and hire applicants with expected productivity above threshold $y^*$, and fire workers whose realized productivity falls below $y^* - \tau$.

**Proposition 2.** In the screening model:

(a) The hiring probability is higher for same-race applicants than cross-race applicants, $p_s > p_c$.
(b) Average productivity is higher for same-race hires than cross-race hires, $\bar{y}_s > \bar{y}_c$.
(c) Productivity variance is lower for same-race hires than cross-race hires, $v_s < v_c$.
(d) Turnover is lower for same-race hires than cross-race hires, $x_s < x_c$.

The intuition is illustrated in Figure 1b (left). For part (a), managers have better signals about same-race applicants, so the estimator variance is higher and candidates have a higher probability of passing the hiring threshold. For part (b), note that since managers have better signals about same-race applicants, the estimator variance is higher. That is, there are many same-race candidates

---

9Given the different signal variances, one might argue that the ability to fire workers induces different cutoffs $y^*$ for different races. This can go in either direction. If the managers take into account the option value of employment, the ability to fire low-productivity workers means managers would lower their standards for high-variance cross-race applicants. Conversely, if reversing one’s hiring decision is viewed as a bad signal by the store manager, managers would raise their standards for cross-race applicants. For simplicity, we assume $y^*$ is the same for each race.
(a) Taste-Based Discrimination

(b) Screening Discrimination

(c) Complementary Production

Figure 1: Distribution of Applicants’ Estimators and Recruits’ Productivity

Notes: The left figure illustrates the distribution of estimators \( \hat{y} \) under same-race and cross-race managers. The right figure shows the difference between same-race and cross-race productivity by quantile (the y-axis shows the de-logged scale for ease of interpretation). These figures assume ability has mean \( \mu = -0.5 \) and variance \( \sigma^2 = 0.4 \), the signal variance is \( \sigma^2_s = 0.2 \) and cutoff is \( y^* = -0.3 \); this generates a productivity distribution with mean 0.046, median 0.028 and variance 0.21, similar to our data (see Figure C1 in the Online Appendix C). Under taste-based discrimination, the same- and cross-race cutoffs are \( y^*_s = -0.3 \) and \( y^*_c = -0.2 \). Under screening discrimination, the same- and cross-race signals have variance \( \sigma^2_s = 0.2 \) and \( \sigma^2_c = 0.3 \). Under complementary production, the same- and cross-race means are \( \mu_s = -0.4 \) and \( \mu_c = -0.5 \).
whose expected productivity $\hat{y}$ far exceeds the threshold $y^*$. In comparison, the estimated productivity of most cross-race hires is relatively close to $y^*$, and so same-race hires have higher expected productivity than cross-race hires. Part (c) states that the higher accuracy of the same-race signals means that the productivity variance is lower for same-race hires. Intuitively, if managers screen on a very noisy signal then the variance of realized productivity equals that of the prior; as the accuracy rises, more unproductive workers are excluded, and the variance of realized productivity falls. Finally, part (d) follows since same-race recruits have higher mean productivity and lower residual productivity variance. Both of these forces lower the chance of outliers, and thus turnover.

The predictions for hiring and turnover are the same as with taste-based discrimination (Proposition 1). Thus, hiring data alone is not enough to differentiate between the two models. To do this one needs productivity data, for which the models generate opposite predictions. To gain some further intuition for the productivity difference, Figure 1b (right) plots the difference between same-race and cross-race productivity by quantile. One can see that the gap is initially positive, shrinks as the workers get better, and eventually becomes negative. Intuitively, cross-race managers make some very bad hires since their signal is imprecise, leading to a large gap at the lowest quantiles, while everyone hires the best workers.\footnote{The cross-race manager also makes fewer hires overall, so the very best are a larger percentage of their hires. This explains why the productivity gap in Figure 1b (right) is negative at the highest quantiles.}

2.3 Complementary Production

We now turn to complementary production, assuming workers are more productive under same-race managers, $\mu_s > \mu_c$. For example, same-race managers may provide superior supervision, feedback, and training. Managers take into account the higher productivity of same-race workers, hiring any worker with expected productivity above threshold $y^*$; they fire any worker with realized productivity below threshold $y^* - \tau$.

Proposition 3. In the complementary production model:

(a) The hiring probability is higher for same-race applicants than cross-race applicants, $p_s > p_c$.
(b) Average productivity is higher for same-race hires than cross-race hires, $\bar{y}_s > \bar{y}_c$.
(c) Productivity variance is higher for same-race hires than cross-race hires, $v_s > v_c$.
(d) Turnover is lower for same-race hires than cross-race hires, $x_s < x_c$.

The intuition is illustrated in Figure 1c (left). For part (a), same-race workers are more productive, so managers hire more of them. For parts (b)-(d), the higher productivity increases the proportion of infra-marginal same-race workers with productivity far above the threshold. Thus, the productivity of same-race workers has higher mean and variance, and turnover is lower. To illustrate the productivity differences, Figure 1c (right) plots the difference between same-race and
cross-race productivity by quantile. One can see that the gap is positive and grows as workers get better. Intuitively, complementarity is at work at the top of the distribution, whereas it is offset by selection at the bottom.

2.4 Discrimination Tests

In Sections 2.1-2.3, we consider one theory at a time. To bring our model to the data we nest the three theories in one common model that additionally allows for worker and manager heterogeneity by race. This analysis directly leads to our supermodularity tests.

**General Model.** Consider a population containing mass \( n_\Theta \) race-\( \Theta \) workers where \( \Theta \in \{W, B, H, \ldots\} \), that are screened by race-\( \theta \) managers where \( \theta \in \{w, b, h, \ldots\} \). We write \( \Theta = \theta \) when, say, \( \Theta = W \) and \( \theta = w \). The model is indexed by parameters \( (\chi_\Theta, \xi_\theta; d_{\Theta\theta}, z_{\Theta\theta}, k_{\Theta\theta}) \) that come from two components. First, we model worker heterogeneity by parameters \( \chi_\Theta \) and manager heterogeneity by parameters \( \xi_\theta \). For workers, \( \chi_\Theta \) could reflect differences in the population mean \( \mu \) or variance \( \sigma^2_\theta \). For managers, \( \xi_\theta \) could reflect differences in hiring standards \( y^* \) or signal quality \( \sigma^2_\epsilon \). Second, we incorporate three forms of discrimination. Taste-based discrimination is indexed by parameters \( d_{\Theta\theta} \geq 0 \), which additively enter the hiring threshold \( y^*_{\Theta\theta} = y^* + d_{\Theta\theta} \). Screening discrimination is indexed by parameters \( z_{\Theta\theta} \geq 0 \), which multiplicatively enter the signal variance \( \sigma^2_{\Theta\theta} = (\sigma_\epsilon(1 + z_{\Theta\theta}))^2 \). Complementary production is indexed by parameters \( k_{\Theta\theta} \geq 0 \), which additively enters the population productivity average \( \mu_{\Theta\theta} = \mu - k_{\Theta\theta} \).\(^{11}\) We normalize \((d_{\Theta\theta}, z_{\Theta\theta}, k_{\Theta\theta}) = (0, 0, 0)\) for worker-manager pairs of the same race. When \((d_{\Theta\theta}, z_{\Theta\theta}, k_{\Theta\theta}) = (0, 0, 0)\) then race-\( \theta' \) managers do not discriminate against race-\( \Theta \) workers.

We wish to understand how heterogeneity and discrimination \( (\chi_\Theta, \xi_\theta; d_{\Theta\theta}, z_{\Theta\theta}, k_{\Theta\theta}) \) affect our output variables \((p, \bar{y}, \nu, x)\). We trace out the impact of the parameters via Taylor expansions. For example, the productivity mean of race-\( \Theta \) workers equals \( \bar{y}_\Theta = \bar{y}(\chi_\Theta, \xi_\theta, 0, 0, 0) \) when hired by same-race managers \( \theta = \Theta \), and \( \bar{y}_{\Theta\theta} = \bar{y}(\chi_\Theta, \xi_\theta, d_{\Theta\theta}, z_{\Theta\theta}, k_{\Theta\theta}) \) when hired by cross-race managers \( \theta' \neq \Theta \). Taking a Taylor approximation around \((\chi_\Theta, \xi_\theta; d_{\Theta\theta}, z_{\Theta\theta}, k_{\Theta\theta}) = 0\)

\[
\bar{y}_{\Theta\theta} = \bar{y}(0) + \chi_\Theta \frac{\partial \bar{y}}{\partial \chi} + \xi_\theta \frac{\partial \bar{y}}{\partial \xi} + d_{\Theta\theta} \frac{\partial \bar{y}}{\partial d} + z_{\Theta\theta} \frac{\partial \bar{y}}{\partial z} + k_{\Theta\theta} \frac{\partial \bar{y}}{\partial k}
\]

(1)

We can similarly take Taylor approximations for the other outcome variables. Echoing Table 1,

\(^{11}\)Note that \( k \) enters as a negative term in the productivity of cross-race workers, so \( k \) more literally measures “mismatch” rather than “complementarity”. We do this so we can allow for heterogeneity across race-pairs and because taste-based discrimination \( d \) and screening discrimination \( z \) are also formulated for of cross-race hires.
Propositions 1-3 then tell us about their partial derivatives:

**Taste-Based Discrimination:**  
\[
\frac{\partial p}{\partial d} < 0, \quad \frac{\partial \bar{y}}{\partial d} > 0, \quad \frac{\partial v}{\partial d} < 0, \quad \frac{\partial x}{\partial d} > 0
\]

**Screening Discrimination:**  
\[
\frac{\partial p}{\partial z} < 0, \quad \frac{\partial \bar{y}}{\partial z} < 0, \quad \frac{\partial v}{\partial z} > 0, \quad \frac{\partial x}{\partial z} > 0
\]

**Complementary Productivity:**  
\[
\frac{\partial p}{\partial k} < 0, \quad \frac{\partial \bar{y}}{\partial k} < 0, \quad \frac{\partial v}{\partial k} < 0, \quad \frac{\partial x}{\partial k} > 0
\]

The underlying assumption behind the use of Taylor expansions is that the heterogeneity and bias are small. This assumption is consistent with our empirical results: The mean productivity of different race pairs varies by at most 4% (see Table 5); these numbers capture both worker and manager heterogeneity \((\chi, \xi)\) and the discrimination parameters \((d, z, k)\).

**Tests of Discrimination.** Using our general model, we seek to measure the form and extent of discrimination. One might first think about comparing one type of worker under different types of managers (e.g., asking whether Hispanic workers are more productive under white or Hispanic managers). However, this “within worker test” is undermined by the managerial heterogeneity \(\xi\). One can see this by using our Taylor approximation (1) to compare the productivity of worker \(\Theta\) under same-race manager \(\theta\) and cross-race manager \(\theta'\),

\[
\bar{y}_{\Theta\theta} - \bar{y}_{\Theta\theta'} = \bar{y}(\chi_{\Theta}, \xi_{\Theta}, 0, 0, 0) - \bar{y}(\chi_{\Theta}, \xi_{\Theta'}, d_{\Theta\theta'}, z_{\Theta\theta'}, k_{\Theta\theta'})
\]

\[
= (\xi_{\Theta} - \xi_{\Theta'}) \frac{\partial \bar{y}}{\partial \xi} - d_{\Theta\theta'} \frac{\partial \bar{y}}{\partial d} - z_{\Theta\theta'} \frac{\partial \bar{y}}{\partial z} - k_{\Theta\theta'} \frac{\partial \bar{y}}{\partial k}.
\]

Manager Heterogeneity

Discrimination Effects

For example, Hispanic workers may be more productive under Hispanic managers than under white managers because Hispanic managers have a large positive fixed effect, \(\xi_{\Theta}\).  

To overcome the problem of worker and manager heterogeneity, we take a diff-in-diff approach and propose a **supermodularity test** of mean productivity. Define

\[
\Delta^y_{\Theta\Theta'} := \frac{(\bar{y}_{\Theta\theta} - \bar{y}_{\Theta\theta'})}{\text{Productivity gap for } \Theta} - \frac{(\bar{y}_{\Theta'\theta'} - \bar{y}_{\Theta'\theta'})}{\text{Productivity gap for } \Theta'}
\]

where \(\Theta = \theta\) and \(\Theta' = \theta'\). For example, this compares the productivity boost Hispanic workers get from having a Hispanic manager to the productivity boost that white workers get from having a Hispanic manager. Using the Taylor expansion (1),

\[
\Delta^y_{\Theta\Theta'} = -(d_{\Theta\theta'} + d_{\Theta\theta}) \frac{\partial \bar{y}}{\partial d} - (z_{\Theta\theta'} + z_{\Theta\theta}) \frac{\partial \bar{y}}{\partial z} - (k_{\Theta\theta'} + k_{\Theta\theta}) \frac{\partial \bar{y}}{\partial k}.
\]

12One can similarly compare the productivity of different races of workers under the same manager. But the sign of this “within manager effect” may simply reflect differences in worker populations, \((\chi_{\Theta} - \chi_{\Theta'}) \frac{\partial \bar{y}}{\partial \chi}\).

13We use the double-capitalized notation \(\Theta\Theta'\) to indicate a general pair of races, rather than a worker-manager race pair \(\Theta\theta'\).
Recall from (2) that $\partial \bar{y}/\partial d$ is positive, whereas $\partial \bar{y}/\partial z$ and $\partial \bar{y}/\partial k$ are negative. Therefore, if mean output is submodular, $\Delta^y_{\Theta^\prime} < 0$, then $d_{\Theta^\prime} + d_{\Theta^\prime} > 0$ which means at least one race of manager uses taste-based discrimination.

We next consider productivity variance. Analogous to the classic “ratio of variances” test, we define the log-supermodularity test of productivity variance,

$$\Delta^v_{\Theta^\prime} := \frac{v_{\Theta^\prime}/v_{\Theta^\prime}}{v_{\Theta^\prime}/v_{\Theta^\prime}}$$

(5)

Taking logs and applying the Taylor approximation

$$\log \Delta^v_{\Theta^\prime} = (\log v_{\Theta^\prime} - \log v_{\Theta^\prime}) - (\log v_{\Theta^\prime} - \log v_{\Theta^\prime})$$

$$= \frac{1}{v} \left(- (d_{\Theta^\prime} + d_{\Theta^\prime}) \frac{\partial v}{\partial d} - (z_{\Theta^\prime} + z_{\Theta^\prime}) \frac{\partial v}{\partial z} - (k_{\Theta^\prime} + k_{\Theta^\prime}) \frac{\partial v}{\partial k}\right)$$

By (2), $\partial v/\partial z$ is positive, whereas $\partial v/\partial d$ and $\partial v/\partial k$ are negative. Therefore, if the variance is log-submodular, in that $\log \Delta^v_{\Theta^\prime} < 0$ or $\Delta^v_{\Theta^\prime} < 1$, then $z_{\Theta^\prime} + z_{\Theta^\prime} < 0$ which means at least one race of manager uses screening discrimination.

Using these insights, we can use productivity data to identify all three discrimination theories:

**Proposition 4 (Tests of Discrimination).** Consider the general model with worker and manager heterogeneity, and assume the Taylor approximations are valid.

(a) If $\Delta^y_{\Theta^\prime} < 0$, then there exists taste-based discrimination: $d_{\Theta^\prime} + d_{\Theta^\prime} > 0$.

(b) If $\Delta^v_{\Theta^\prime} < 1$, then there exists screening discrimination: $z_{\Theta^\prime} + z_{\Theta^\prime} > 0$.

(c) If $\Delta^y_{\Theta^\prime} > 0$ and $\Delta^v_{\Theta^\prime} > 1$, then there exists complementary production: $k_{\Theta^\prime} + k_{\Theta^\prime} > 0$.

Part (c) says that there is no combination of taste-based and screening discrimination that could explain $\Delta^y_{\Theta^\prime} > 0$ and $\Delta^v_{\Theta^\prime} > 1$. Intuitively, taste-based discrimination most directly affects expected productivity, while screening discrimination most directly affects productivity variance. Formally, we show in Appendix A.2

$$-\frac{\partial v/\partial d}{\partial \bar{y}/\partial d} < -\frac{\partial v/\partial z}{\partial \bar{y}/\partial z}.$$  

(6)

Figure 2 (left) illustrates the impact of the three types of discrimination ($d, z, k$) on ($\Delta^y, \Delta^v$). In particular, it shows that an increase in an one variable ($d, z, k$) cannot be replicated by a increase in the other two variables. The matrix in Figure 2 (right) then summarizes our tests.

One may wonder whether Proposition 4 captures all empirical implications of our model: Are there additional combinations of ($\Delta^y, \Delta^v$) that allow us to conclude that one of our discrimination forces must be at play? For screening discrimination this is not the case. We show in Appendix A.2 that the slope of any of the three vectors in Figure 2 is linear in the estimator variance $\eta$, which is unobservable, and hence potentially arbitrarily small. For small $\eta$, the cone spanned by $\partial_d$ and $\partial_k$
Figure 2: **Identifying Discrimination from Productivity Data**

*Notes:* The figure on the left illustrates how taste-based discrimination, screening discrimination and complementary production affect the supermodularity of mean productivity $\Delta y$ and productivity variance $\Delta v$. By (6), $\partial_k$ is steeper than $\partial_y$. The matrix on the right shows the table we will use for our empirical tests.

covers the entire upper half-plane, so Proposition 4(b) is tight. However, the complementarity test, Proposition 4(c), can be tightened: The slope of $\partial_d$ is bounded above, so the triangle area above $\partial_d$ in the upper-left quadrant can only be reached with $\partial_k$, and is hence proof of complementary production. We do not pursue this angle further in this paper since such considerations would not change our empirical analysis.\(^\text{14}\)

**Tests of the Model.** Figure 2 implies that linear combinations of $\partial_d$, $\partial_z$, and $\partial_k$ with positive coefficients span the entire plane; in other words productivity data cannot reject the joint model. However, all three discrimination theories have the same predictions for hiring and turnover. This gives us a way of rejecting our model, even in the presence of worker and manager heterogeneity.

Hiring probabilities $p_{\Theta \theta'}$ are not empirically observable. We address this problem by measuring the share of race-$\Theta$ workers among the $N_{\theta'} = \sum_{\theta'} n_{\Theta \theta'} p_{\Theta \theta'}$ workers hired by race-$\theta'$ managers, $r_{\Theta \theta'} = \frac{n_{\Theta \theta'} p_{\Theta \theta'}}{N_{\theta'}}$. As with productivity variance we define the log-supermodularity test of hiring,

$$\Delta^{r}_{\Theta \theta'} := \frac{r_{\Theta \theta'} / r_{\Theta' \theta'}}{r_{\Theta' \theta} / r_{\Theta \theta'}} \quad (7)$$

\(^{14}\)The above discussion concern restrictions on individual vectors $(\partial_d, \partial_z, \partial_k)$. We also know that there exists joint restrictions on these vectors. In particular, Figure 2 (left) shows that there always exists a half-space to the upper right of the origin that is proof of complementary production, although we do not know which half-space this is. If 95% of our bootstraps lie in all half-planes to the upper left of the origin, we could conclude there is complementary production; this is a more powerful test than the one proposed in Proposition 4(c), where we just look at the upper right quadrant. For our data, this does not affect our conclusions.
Our theory predicts that $\Delta_{\Theta'} > 1$. Using a Taylor approximation and the inequalities in (2),

$$\log \Delta_{\Theta'} = \log r_{\Theta} - \log r_{\Theta'} - (\log r_{\Theta'} - \log r_{\Theta'})$$

$$= \frac{1}{p} \left( -(d_{\Theta'} + d_{\Theta'}) \frac{\partial p}{\partial d} - (z_{\Theta'} + z_{\Theta'}) \frac{\partial p}{\partial z} - (k_{\Theta'} + k_{\Theta'}) \frac{\partial p}{\partial k} \right) > 0.$$  

Finally, we define the supermodularity test of turnover

$$\Delta_{\Theta'}^x := (x_{\Theta} - x_{\Theta'}) - (x_{\Theta'} - x_{\Theta'}).$$ (8)

Our theory predicts that $\Delta_{\Theta'}^x < 0$. Using a Taylor approximation and the inequalities in (2),

$$\Delta_{\Theta'}^x = -(d_{\Theta'} + d_{\Theta'}) \frac{\partial x}{\partial d} - (z_{\Theta'} + z_{\Theta'}) \frac{\partial x}{\partial z} - (k_{\Theta'} + k_{\Theta'}) \frac{\partial x}{\partial k} < 0.$$

We have thus shown

**Proposition 5 (Tests of Model).** Consider the general model with worker and manager heterogeneity, and assume the Taylor approximations are valid.

(a) Hiring is log-supermodular, $\Delta_{\Theta'}^r > 1$.

(b) Turnover is submodular, $\Delta_{\Theta'}^x < 0$.

Part (a) states that the ratio of workers hired by manager $\theta$, $r_{\Theta}/r_{\Theta'}$, exceeds the ratio hired by manager $\theta'$, $r_{\Theta'}/r_{\Theta'}$. One might wonder whether, more strongly, each manager will hire a greater proportion of same-race workers, $r_{\Theta}/r_{\Theta'}$ and $r_{\Theta'/\theta}$ and $r_{\Theta'/\theta'}$. This need not be the case: If, say, white managers strongly discriminate against Hispanic workers then they may end up hiring more black workers than a black manager. But the ratio of white/black workers will always be higher for the white manager.

### 2.5 Discussion

Here we discuss our key assumptions and compare our supermodularity test with the literature.

**Power of the Test.** Our test will not always detect discrimination, even as the sample size grows to infinity. If all three forms of discrimination are present, then we can detect at most two (see Figure 2). Moreover, if the different forms of discrimination offset one another, our test may be unable to identify any of them.

**Random Assignment.** Our key identifying assumption is that workers are randomly assigned to managers. Formally, this means that the number of applicants $n_{\Theta}$ and the distribution of worker skills ($\mu_{\Theta}, \sigma^2_{\Theta}$) are independent of manager race, $\theta$. Several forces may conflict with this assumption:
Regional variation. Our data contains stores across the US, and applicants at a given store may vary in their racial mix, $n_\Theta$, and their skill distribution ($\mu_\Theta, \sigma^2_\Theta$). Since manager racial composition also varies across stores, we control for store fixed effects.

Job differentiation. Jobs in some departments may be harder and higher paid than others, attracting different numbers of applicants $n_\Theta$ and skill distributions ($\mu_\Theta, \sigma^2_\Theta$). Departments may differ in their managerial race composition, so we control for department fixed effects.

Intertemporal changes. As the retail sector changes over time (because of seasonality or sectoral decline), the composition of applicants may change in their racial mix, $n_\Theta$, and their skill distribution ($\mu_\Theta, \sigma^2_\Theta$). Managerial race composition may also change over time, so we control for month fixed effects.

Ultimately, the hiring analysis (Proposition 5(a)) assumes each manager faces an identical distribution of workers in terms of the number of applicants $n_\Theta$ and distribution of skills ($\mu_\Theta, \sigma^2_\Theta$) after controlling for fixed effects. The test of discrimination (Proposition 4) just requires that each manager faces an identical distribution of skills ($\mu_\Theta, \sigma^2_\Theta$) after controlling for fixed effects; differences in the number of workers $n_\Theta$ do not change the modularity of productivity.\(^{15}\)

**Log Utility.** The model assumes that managers have log utility, which means that they hire a worker if expected log-sales $\hat{y}$ exceeds a cutoff, $y^*$. Instead, suppose managers are risk neutral and wish to maximize expected sales; they are then risk-loving with respect to log-sales. Under taste-discrimination and complementary production the analysis is unchanged. Under screening discrimination, things become a little more complicated since managers now apply a lower hiring threshold to cross-race hires than same-race hires because of the higher residual variance of cross-race applicants. This new force works against the bias for hiring same-race candidates, Proposition 2(a), but the result still holds if the hiring threshold is sufficiently high. On the other hand, the lower threshold means managers hire more low-productivity cross-race workers, and thus reinforces Propositions 2(b,c,d). See Online Appendix B.1 for details.

**Absolute vs. Relative Discrimination.** The classic “outcome test” (Becker, 1957) examines whether a firm is biased by comparing the performance of marginal applicants. We do not observe the managers’ private signals, and thus do not observe the marginal worker. A naive comparison of mean productivity $E_\theta[\bar{y}_{\Theta}]$ across worker $\Theta$ would then suffer from both false positives and false negatives. Specifically, if worker $\Theta$ were more productive than $\Theta'$ because of higher fixed effects $\chi_\Theta$ (e.g. a higher prior mean), we would see $E_\theta[\bar{y}_\Theta] > E_\theta[\bar{y}_{\Theta'}]$ even if there were no discrimination. Conversely, if two equally sized populations exhibit symmetric discrimination (e.g. $d_{\Theta\theta} = d_{\Theta'\theta} > 0$)\(^{15}\)

\(^{15}\)Similarly, the test of turnover (Proposition 5(b)) just requires that each manager faces an identical distribution of skills ($\mu_\Theta, \sigma^2_\Theta$) after controlling for fixed effects.
then we could see $E_{\theta}[\bar{y}_{e\theta}] = E_{\theta}[\bar{y}_{e\theta}]$ even in the presence of discrimination. Rather than try to identify the marginal recruit, we follow a second approach pioneered by Anwar and Fang (2006) of looking for relative bias between managers of different races.

3 Empirical Setting and Approach

Our data come from the U.S. operations of a large national retailer from February 2009 to March 2015. Each establishment is led by a store manager and a team of department managers (henceforth “managers”) who hire for their respective departments, among other duties.

**Hiring and Employment.** When a department has a vacancy, the manager requests a shortlist of qualified applicants from the regional HR representative, who assembles a list from online advertisements specific for that posting or from qualified “evergreen” applicants who can apply for entry-level positions at any time. Applicants to sales roles take a screening test online or at a store location, which generally requires less than an hour to complete. The test asks candidates about their background, technical qualifications, and experience, how they would respond under hypothetical scenarios with customers or colleagues, and a personality test. These tests are scored by an algorithm that provides a three-tier recommendation for whether to proceed with an interview. The department managers observe the algorithmic recommendations and select whom to interview; they are not obligated to follow the recommendation.

Job interviews can vary substantially, but managers receive a guide for conducting behavioral interviews that further ask applicants how they would respond to hypothetical scenarios. Consistent with this training, most employees report that managers use behavior-based questions. For example, the applicant may be asked to describe a life or work experience in which they overcame an obstacle, helped a stranger, or witnessed dishonesty. After being hired, a worker typically undergoes about one week of formal online training and a week of job shadowing before moving to regular status. Salespeople are highly incentivized to maximize their productivity because commissions and other forms of incentive pay account for about 40% of their income over this period, with the bulk of this pay tied directly or indirectly to their individually-credited sales performance.

Department managers have many duties outside hiring. They supervise salespeople, schedule shifts, give feedback to employees, handle customer complaints, and so on. They also play a significant role running the whole store, e.g. opening and closing the store, helping the manager prepare for sales events. These managers are also highly incentivized to maximize the productivity of their team since they are assessed and rewarded on the sales performance of their departments, which depends on the cumulative sales performance of their team. In particular, the company uses annual bonuses (averaging 5-10% of salary, but potentially much higher), retention offers, and promotion to encourage managers to hire productive salespeople and ultimately build a profitable department.
**Data Features.** The model assumes that the number of applicants $n_{\Theta}$ and their productivity $(\mu_{\Theta}, \sigma^2_{\Theta})$ is independent of the manager’s race $\theta$. Given that regional HR compiles the list of applicants, we believe that this is a reasonable assumption, but it is ultimately untestable. One employee within HR told us that it is rare that department managers know applicants personally; we discuss such referrals in Section 5.4.

The data have a number of features that make it well suited to study discrimination in hiring. First, the nature of the job means that productivity largely reflects individual performance. In comparison, in many other industries (e.g. manufacturing) and even other sales jobs (e.g. complex business-to-business sales), performance reflects the work of an entire team. Second, as discussed above, salespeople and managers are highly incentivized to maximize their productivity. Third, managers have a large degree of autonomy in assessing applicants, allowing us to identify the impact of managerial race. Fourth, the firm is demographically and educationally similar to the greater retail trade industry, which employs over 16 million workers (more than 10% of the US labor force). These jobs are also similar in skill and labor force attachment to those held by many working class Americans where racial disparities seem to be largest (Lang and Lehmann, 2012). Finally, this firm employs about 1.2 million workers over the seven year period, affording a large amount of data on demographics, hiring, and turnover, as well as perhaps the largest such data on commissioned workers. Similar sized data, like the American Community Survey’s sample of employed workers (which covers about 1% of all U.S. workers each year), are missing key variables like job performance and manager race.

**Descriptive Statistics.** The data we use for our main analyses include longitudinal administrative records on 63,842 commissioned salespeople and their managers, including productivity, demographic identity, department, and store location. Of these, 56,071 have an observable hiring manager because they are either directly observed at their hire date (47,025) or have a known hiring manager due to available position tenure data (9,046). To focus on racial combinations for whom we have the greatest statistical power, we restrict the sample to white, black, and Hispanic (WBH) workers hired by WBH managers. The resulting sample consists of 48,755 salespeople hired by 7,892 managers at 997 store locations over 74 months, or 335,867 worker-months. Salespeople are 64% white, 19% black, and 17% Hispanic, whereas managers are 77% white, 11% black, and 12% Hispanic. The age quartiles are (21, 25, 42), the mean age is 32, and 41% identify as female. For department categorization, we use the 44 departments with at least 1,000 person-months in the full sample; the remaining 3.78% of salespeople are in an “other” category.

Our first variable of interest is the **new hire race.** As one would expect, there is a fair degree of regional segregation: white managers’ WBH hires are 71.5% white, black managers WBH hires are 43.9% black, and Hispanic managers’ WBH hires are 48.7% Hispanic. Fortunately, there is still a lot of variation of race within stores. Of the 997 stores, 423 have white and black managers, 412
have white and Hispanic managers, 217 have black and Hispanic managers, and 214 stores have all three. Nearly all (98.9%) new commissioned salespeople were hired into locations with variation in the race of new commissioned salespeople.

Our second variable of interest is sales productivity. We begin with a worker’s sales per hour as a percent of their sales targets, which serves as the basis for their commission pay. Sales targets depend on several factors including shift, location, department, period, and tenure (but not race). Targets are set centrally by corporate headquarters and not by the managers. As shown in Figure C1 in Online Appendix C, the sales divided by the target distribution is roughly log-normal, so we use the performance measure \( \ln \left( \frac{\text{sales per hour}}{\text{target}} \right) \), and winsorize this variable at 1% within race pairs. To be consistent with the theory, we restrict the data to workers who work under their hiring manager. The resulting sales productivity measure has a mean of 0.008 reflecting that the median worker hits their target almost exactly. It has a standard deviation 0.5178, reflecting substantial variation in sales performance. This performance measure features small differences in the mean productivity of salespeople/managers of different races (see Table 5). These could come from heterogeneity in the underlying population or from compositional effects in the model (e.g. under screening discrimination or complementary production, the minority race has lower mean productivity simply because they are in the minority). Our supermodularity tests are robust to such heterogeneity.

The third variable of interest is turnover. We first remove 2,703 salespeople who were hired before the start of our data and 6,713 salespeople who were hired within six months of the end of the data, leaving us with 39,339 people. After six months, 32.6% leave the firm directly from sales, 3.5% switch to a non-sales job, and 59.7% stay within sales. Our theory posits that a worker is fired if the output drops below a cutoff; consistent with this, the average \( SPH \) is \(-6.1\)pp for leavers, \(-7.8\)pp for switchers and \(+0.5\)pp for stayers. Given the switchers perform worse than the leavers, we pool the two groups and say someone has turned over if they leave sales. Turnover is then 26.4% after 3 months, 36.1% after 6 months, 44.5% after 9 months and 51.7% after 12 months.\(^{16}\)

Data Limitations and Extensions. While we have productivity for 48,755 salespeople, the data are limited in some dimensions. First, we do not observe the number or race of applicants, which would help us test the random assignment of worker race, and provide data about the probability of acceptance. Second, we do not observe the company’s three-tier recommendation, which would help us test for the random assignment of worker productivity. Third, we do not observe the shift assignments; however, the target does control for shift. This is important because “sales per hour/target” is the basis for salespeople’s commission.

---

\(^{16}\)Our theory also assumes that the hiring manager makes the decision to fire the worker. Managerial changes will not affect screening discrimination, but could generate supermodular turnover under taste-based discrimination if a biased manager is replaced with an unbiased one. We don’t think this is likely to be a large problem as most workers remain under their hiring manager.
Our baseline results examine hiring and productivity for newly hired salespeople. In Section 4.1, we assess the hiring and turnover hypotheses with the full set of employment records, rather than just salespeople, greatly enhancing our sample size. Such workers include cashiers, backroom operators, and service technicians. In all, there are 812,399 WBH workers at 4,646 store locations, which may be the largest such employment data. Of these workers, 57% are white, 25% are black, and 18% are Hispanic. The turnover is 26.5% at 3 months, 49.2% at 6 months, 62.2% at 9 months, and 70.6% at 12 months.

3.1 Empirical Implementation

We first describe our regressions:

**Hiring.** Our model predicts that managers hire more same-race workers (see Proposition 5). To test this we run the following regression for each worker race $\Theta$,\(^\dagger\)

$$100 \times \text{Worker}_{\Theta} = \beta_{\Theta \theta} \text{Manager}_{\theta} + \text{controls}_{i,t} + \epsilon_i^r \tag{9}$$

On the LHS, we have an indicator that tells us whether the worker’s race is $\Theta$ multiplied by 100. On the RHS, we have indicators that tell us whether that worker’s manager is $\theta$. The controls consist of 997 store fixed effects, 45 department fixed effects, and 74 month fixed effects. Thus, $\beta_{\Theta \theta}$ equals the adjusted share of a $\theta$ managers’ employees who are of race $\Theta$. Based on (7), define

$$\hat{\Delta}_{\Theta'=\Theta}^r = \frac{\hat{\beta}_{\Theta \theta}^r / \hat{\beta}_{\Theta' \theta}^r}{\frac{\hat{\beta}_{\Theta \theta}}{\hat{\beta}_{\Theta' \theta}}} \tag{10}$$

where $\Theta = \theta$ and $\Theta' = \theta'$. Under the assumption that workers are randomly assigned to managers after controlling for fixed effects, Proposition 5(b) predicts that $\hat{\Delta}_{\Theta'=\Theta}^r > 1$.

**Productivity.** We use the mean and variance of productivity to tease apart the three types of discrimination (see Proposition 4). As a first step, we filter out the effect of tenure; we separate this from the other controls since we are concerned that workers with longer tenures are positively selected.\(^\ddagger\) Specifically, we restrict the set of workers to those who are still employed after 12 months and work under the hiring manager and, using the first 12 months of data for each worker, run the regression

$$\ln(\text{sales per hour}/\text{target})_{i,t} = \beta_T^0 + \beta_T^1 \ln(\text{tenure})_{i,t} + \text{controls}_{i,t} + \epsilon_i^T \tag{11}$$

For all salespeople, we then define the tenure-adjusted sales per hour by

$$SPH_{i,t} := \ln(\text{sales per hour}/\text{target})_{i,t} - \hat{\beta}_0^T - \hat{\beta}_1^T \ln(\text{tenure})_{i,t}$$

\(^\dagger\)Our findings are the same for logit or probit regressions. The interpretation is easier with a linear regression.

\(^\ddagger\)The firm also accounts for tenure when setting the salesperson’s target, but we have the benefit of hindsight to filter out the realized effects.
$SPH_{i,t}$ has mean zero with a standard deviation 0.531. The average $SPH_{i,t}$ for a given worker over time has standard deviation 0.427, reflecting substantial variation in performance across individuals.

To perform our productivity test, we consider the random effects model

$$SPH_{i,t} = \beta_{\Theta \theta} y + \text{controls}_{i,t} + u_i + \epsilon_{i,t}$$

for $i = 1 \ldots N_{\Theta \theta}$ and $t = 1 \ldots T_i$, where $N_{\Theta \theta}$ are the number of $\Theta$ workers under $\theta$ managers and where $T_i$ are the number of observations of worker $i$. We allow for the same controls as in (9). Agent $i$’s productivity $u_i \sim N(0, \sigma_{\Theta \theta}^2)$ and the monthly noise $\epsilon_{i,t}^y \sim N(0, \sigma_{\epsilon, \Theta \theta}^2)$ are IID.\(^\text{19}\) Thus, $\beta_{\Theta \theta}^y$ is the mean productivity of worker $\Theta$ under manager $\theta$, and $\sigma_{\Theta \theta}^2$ is the variance of worker $\Theta$ under manager $\theta$.

We estimate mean group productivity $\beta_{\Theta \theta}^y$ via weighted least squares, where each worker is weighted by the inverse of their tenure. Thus, $\beta_{\Theta \theta}^y$ reflects the person-weighted average productivity of group $\Theta \theta$ rather than the person-month-weighted average. This mitigates attrition bias and can raise efficiency, since agent $i$’s productivity shock $u_i$ is common across all of their observations. Based on (3), define

$$\hat{\Delta}_y^{\Theta \theta'} = (\hat{\beta}_{\Theta \theta}^y - \hat{\beta}_{\Theta \theta'}^y) - (\hat{\beta}_{\Theta \theta}^{y'} - \hat{\beta}_{\Theta \theta'}^{y'})$$

Productivity gap for $\Theta$ Productivity gap for $\Theta'$

(12)

We calculate standard errors by 10,000 bootstraps stratified by race pairs.\(^\text{20}\)

To estimate the variance of group productivity $\sigma_{\Theta \theta}^2$, we use a simplified version of the Kline, Saggio, and Sølvsten (2020) leave-out estimator.\(^\text{21}\) Letting $e_{it}$ be $i$’s residual at time $t$, we first estimate the monthly noise:

$$\sigma_{\epsilon, \Theta \theta}^2 = \frac{1}{N_{\Theta \theta}} \sum_{i=1}^{N_{\Theta \theta}} \frac{1}{T_i - 1} \sum_{t=1}^{T_i} (e_{it} - \bar{e}_i)^2$$

where $\bar{e}_i := \left( \sum_{t=1}^{T_i} e_{it} \right) / T_i$ is worker $i$’s average residual. We allow this to differ at the group level to capture, say, heterogeneous volatility at different stores. We then estimate productivity variance

\(^\text{19}\)In the model, the prior distribution of productivity is normal, however the posterior is truncated based on a signal and thus technically not normal. Empirically productivity looks normal (see Figure C1) in Online Appendix C. This is consistent with the model if the signal is fairly weak.

\(^\text{20}\)Bootstrapping this allows us perform joint tests with productivity mean and variance. Table 5 shows that clustering by worker leads to similar results.

\(^\text{21}\)Equation (13) coincides with KSS equation (4), but we estimate monthly variance at the group level (rather than the individual level) and weight by person (rather than person-month). We do the former for efficiency, and the latter to mitigate attenuation bias. The true KSS estimator also re-estimates the regression (11) after leaving out each observation; this is computationally infeasible for our bootstrap.
by subtracting the monthly variance from the across-worker variance,\textsuperscript{22}

$$\hat{\sigma}^2_{\Theta \theta} = \frac{1}{N_{\Theta \theta}} \sum_{i=1}^{N_{\Theta \theta}} \left( \bar{e}_i - \bar{e}_{\Theta \theta} \right)^2 - \frac{1}{T_i} \hat{\sigma}^2_{\epsilon, \Theta \theta}$$  \hfill (13)

where \( \bar{e}_{\Theta \theta} := \frac{1}{N_{\Theta \theta}} \sum_{i=1}^{N_{\Theta \theta}} \bar{e}_i \) is the average residual across workers. Based on (5), define

$$\hat{\Delta}^\circ_{\Theta \theta'} = \frac{\hat{\sigma}^2_{\Theta \theta} / \hat{\sigma}^2_{\Theta \theta'}}{\hat{\sigma}^2_{\Theta \theta'} / \hat{\sigma}^2_{\Theta \theta'}}$$  \hfill (14)

Proposition 4 shows that whether \( \hat{\Delta}^y_{\Theta \theta'} \leq 0 \) and \( \hat{\Delta}^v_{\Theta \theta'} \leq 1 \) allow us to identify the nature of discrimination. We calculate standard errors by 10,000 bootstraps stratified by race pairs.

**Turnover.** Our model predicts that same-race workers have a lower turnover rate (see Proposition 5). To implement the test, we run the following regression for each race-\( \Theta \) worker,

$$TURN_i = \beta x_{\Theta \theta} + \text{controls}_{i,t} + \epsilon_i$$  \hfill (15)

where the indicator \( TURN_i \) equals 100 if worker \( i \) has left the job in the first six months, and \( \omega_{\Theta \theta} \) is the turnover of worker \( \Theta \) under manager \( \theta \). In Section 5.1 we examine how the results change for cutoffs other than 6 months. We allow for the same controls as in (9). Based on (8), define

$$\hat{\Delta}^x_{\Theta \theta'} = (\hat{\beta}_{x \Theta \theta} - \hat{\beta}_{x \Theta \theta'}) - (\hat{\beta}_{x \Theta \theta'} - \hat{\beta}_{x \Theta \theta'})$$  \hfill (16)

Proposition 5(b) predicts that \( \hat{\Delta}^x_{\Theta \theta'} \leq 0 \).

### 4 Results

In this section we apply our tests of discrimination. In Section 4.1 we present our hiring results, showing that all races are significantly more likely to hire same-race recruits. In Section 4.2 we examine the results for productivity, where we test for different types of discrimination. In Section 4.3 we consider turnover.

#### 4.1 Hiring Tests

We first present our test for discrimination in hiring. This is a test about the existence of discrimination rather than the cause (Proposition 5(a)). Table 2 shows that, for all three races, same-race hiring is higher than cross-race hiring, \( \hat{\beta}_{r \Theta \theta} > \hat{\beta}_{r \Theta \theta'} \), and this difference is significant at the 1% level for all race pairs. This inequality implies that hiring is log-supermodular \( \hat{\Delta}^r_{\Theta \theta'} = \hat{\beta}_{r \Theta \theta} \hat{\beta}_{r \Theta \theta'} / \hat{\beta}_{r \Theta \theta'} \hat{\beta}_{r \Theta \theta} > 1 \),

\textsuperscript{22}Intuitively, this formula comes from \( E \left[ (\bar{e}_i - \bar{e}_{\Theta \theta})^2 \right] = \text{Var} \left[ \frac{1}{T_i} \sum_{t=1}^{T_i} (u_i + \epsilon_i, t) \right] = \hat{\sigma}^2_{\Theta \theta} + \frac{1}{T_i} \hat{\sigma}^2_{\epsilon, \Theta \theta} \).
### Table 2: Hiring Tests

<table>
<thead>
<tr>
<th></th>
<th>White worker</th>
<th>Black worker</th>
<th>Hispanic worker</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>65.0</td>
<td>18.9</td>
<td>16.1</td>
</tr>
<tr>
<td>manager</td>
<td>[64.6, 65.4]</td>
<td>[18.5, 19.3]</td>
<td>[15.8, 16.5]</td>
</tr>
<tr>
<td>Black</td>
<td>61.7</td>
<td>22.0</td>
<td>16.3</td>
</tr>
<tr>
<td>manager</td>
<td>[60.4, 63.0]</td>
<td>[20.9, 23.1]</td>
<td>[15.3, 17.4]</td>
</tr>
<tr>
<td>Hispanic</td>
<td>60.5</td>
<td>19.3</td>
<td>20.2</td>
</tr>
<tr>
<td>manager</td>
<td>[59.2, 61.9]</td>
<td>[18.1, 20.4]</td>
<td>[19.2, 21.2]</td>
</tr>
</tbody>
</table>

Notes: Reports results for 48,755 commissioned salespeople, as described in the text. Below each coefficient is the 95% confidence interval.

consistent with the model.\textsuperscript{23}

Quantitatively, the gap between same-race and cross-race hiring is 3.9pp, 2.9pp and 4.0pp for white, black and Hispanic salespeople. The average gap between same-race and cross-race hiring is 3.6pp. This is economically significant. A Hispanic worker is \(\frac{20.2}{16.1} - 1 = 25.5\%\) more likely to be hired if the manager is Hispanic than if the manager is white. Given the 50,000 salespeople employed over the course of our study, the firm would employ over 2,000 extra Hispanic employees if managers were all Hispanic as compared to being all white.

#### 4.2 Productivity Tests

We now turn to the productivity tests that allow us to separate our three theories (see Proposition 4). The first column of Table 3 shows that mean productivity is supermodular for all three race combinations, but especially for white-Hispanic and black-Hispanic pairs. The standard errors are fairly large but the former is significantly different from zero at the 10% level. Quantitatively, the mean productivity gap between same-race and cross-race managers is \(\hat{\Delta}^{\bar{y}}_{WH}/2 = 1.3\)pp and \(\hat{\Delta}^{\bar{y}}_{BH}/2 = 1.9\)pp for white-Hispanic and black-Hispanic pairs. To put this in context, a 1% increase in a store’s sales increases its profits by about 6%.\textsuperscript{24}

The second column of Table 3 shows that productivity variance is log-submodular for all three race combinations, but especially for white-Hispanic and black-Hispanic pairs. The standard errors are fairly large but the latter is significantly different from one at the 5% level. Quantitatively, the same-race variance is \(\sqrt{\hat{\Delta}^{\bar{y}}_{WH}} = 0.96\) and \(\sqrt{\hat{\Delta}^{\bar{y}}_{BH}} = 0.84\) of the size of cross-race variance for

\textsuperscript{23}Technically, if both \(\hat{\beta}_{\text{ee}r} > \hat{\beta}_{\text{ee}r'}\) and \(\hat{\beta}_{\text{ee}r} > \hat{\beta}_{\text{ee}r'}\) at the 1% level then \(\hat{\beta}_{\text{ee}r} \hat{\beta}_{\text{ee}r'} > \hat{\beta}_{\text{ee}r'} \hat{\beta}_{\text{ee}r'}\) at the 2% level by Boole’s inequality.

\textsuperscript{24}Suppose the firm’s profit function is \(\pi = (p - c)q - k\). Of the stores in our sample, the average margin is 31.7%, so \(pq = 1.317cq\). Moreover, the average profit margin is 5.11%, so \(pq = 1.0511(cq + k)\). Normalizing total costs to 1, we have \(pq = 1.0511, cq = 0.7470, k = 0.2530\), and \(\pi = 0.0511\). If sales rise by 1% then we have \(pq = 1.0616, cq = 0.7545, k = 0.2530\), and \(\pi = 0.0541\), which is a 5.9% rise in profitability.
Table 3: Supermodularity Tests

<table>
<thead>
<tr>
<th></th>
<th>Mean Prod. $\hat{\Delta}^y$</th>
<th>Var Prod., $\hat{\Delta}^v$</th>
<th>Turnover, $\hat{\Delta}^x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>White-Black</td>
<td>0.0061</td>
<td>0.981</td>
<td>$-0.0154$</td>
</tr>
<tr>
<td>relations</td>
<td>$[-0.0249, 0.0370]$</td>
<td>$[0.857, 1.123]$</td>
<td>$[-0.0490, 0.0181]$</td>
</tr>
<tr>
<td>White-Hispanic</td>
<td>0.0268</td>
<td>0.925</td>
<td>$-0.0354$</td>
</tr>
<tr>
<td>relations</td>
<td>$[-0.0036, 0.0561]$</td>
<td>$[0.801, 1.069]$</td>
<td>$[-0.0698, -0.0011]$</td>
</tr>
<tr>
<td>Black-Hispanic</td>
<td>0.0387</td>
<td>0.701</td>
<td>0.0009</td>
</tr>
<tr>
<td>relations</td>
<td>$[-0.0162, 0.0949]$</td>
<td>$[0.551, 0.919]$</td>
<td>$[-0.0589, 0.0607]$</td>
</tr>
</tbody>
</table>

Notes: Below each coefficient is the 95% confidence interval. For productivity mean and variance, the results come from 48,755 commissioned salespeople. The coefficients are given by the median of 10,000 bootstrap estimates; confidence interval also comes from the bootstrap. For turnover, the results comes from 39,339 commissioned salespeople.

white-Hispanic and black-Hispanic pairs.

Discrimination Tests. Table 4 shows our tests for the source of discrimination, which can be compared with the theoretical predictions in Figure 2 (right). The most clear-cut evidence comes from black-Hispanic pairs. Suppose our null hypothesis is “there is no screening discrimination”. This translates into

$$H_0 : \Delta^v_{BH} \leq 1$$

We can reject this hypothesis since 99.4% of our bootstraps have $\hat{\Delta}^v_{BH} > 1$.

For white-Hispanic pairs, only 85.6% exhibit $\hat{\Delta}^v_{BH} > 1$, so we cannot similarly conclude there must be screening discrimination. However, the fact that mean productivity is supermodular indicates there is either complementarity or screening. With this in mind, suppose our null hypothesis is “there is neither screening discrimination nor complementary production.” Using Figure 2, this translates into

$$H_0 : \Delta^y_{WH} \leq 0 \text{ and } \Delta^v_{WH} \leq 1$$

We can reject this hypothesis since 99.2% of our bootstraps have either $\hat{\Delta}^y_{WH} > 0$ or $\hat{\Delta}^v_{WH} > 1$.

For white-black pairs, there is little evidence in favor of any one hypothesis.

Overall, the hiring results in Table 2 show us that there is discrimination with all three pairs of races. Table 4 tells us there is screening discrimination for black-Hispanic pairs, and some combination of screening discrimination and complementary production for white-Hispanic pairs. We note two caveats when interpreting the results. First, we cannot rule out taste-discrimination for any pairs. It may be that different types of discrimination coexist and offset one another. Second, we do not know which manager races are driving the supermodularity. If productivity is supermodular, Proposition 4(b) tells us that $z_{\Theta \Psi'} + z_{\Theta \Psi} > 0$, but does not tell us whether the
screening discrimination is coming from $\theta$ managers, $\theta'$ managers or both.

**Quantile Analysis.** We can get further insight by looking at the supermodularity of output quantiles. Define $y^q$ to be the $q^{th}$ quantile of output and define $\Delta^q_{\Theta\Theta'} := y^q_{\Theta\theta} + y^q_{\Theta'\theta'} - y^q_{\Theta\theta'} - y^q_{\Theta'\theta}$ as the supermodularity for the $q^{th}$ quantile of output. As in Section 2.4, a Taylor approximation implies

$$\Delta^q_{\Theta\Theta'} = -(d_{\Theta\theta'} + d_{\Theta'\theta}) \frac{\partial y^q(0)}{\partial d} - (z_{\Theta\theta'} + z_{\Theta'\theta}) \frac{\partial y^q(0)}{\partial z} - (k_{\Theta\theta'} + k_{\Theta'\theta}) \frac{\partial y^q(0)}{\partial k}$$

As one can see from the right-hand side of Figure 1, $\partial y^q / \partial d > 0$, $\partial y^q / \partial k < 0$ and $\partial y^q / \partial z < 0$ for most quantiles. To find the empirical equivalent we run an aggregate version of (11)

$$SPH_{i,t} = \beta^q + \text{controls}_{i,t} + \epsilon^q_{i,t}$$

(17)

with the same controls as in (9). Letting $\varepsilon_{i,t}$ be $i$'s residual at time $t$, worker $i$'s average residual is $
abla_i = \left( \sum_{t=1}^{T_i} \varepsilon_{i,t} \right) / T_i$. Given a quantile $q \in [0,1]$, let $\varepsilon^q_{\Theta\theta'}$ be the residual of the $q^{th}$ highest $\Theta$ worker under a $\theta$ manager. Then define

$$\hat{\Delta}^q_{\Theta\Theta'} = \varepsilon^q_{\Theta\theta} + \varepsilon^q_{\Theta'\theta'} - \varepsilon^q_{\Theta\theta'} - \varepsilon^q_{\Theta'\theta}$$

(18)

where $\Theta = \theta$ and $\Theta' = \theta'$\footnote{One concern is that $\varepsilon_i$ measures $u_i$ with noise. If the noise only depends on the quantile (e.g. better workers stay longer and have less noise) then this should cancel out in the supermodularity calculation and should not bias $\Delta^q_{\Theta\Theta'}(q)$. However, if same-race workers have substantially longer tenure (and thus less noise) than cross-race workers, $\Delta^q_{\Theta\Theta'}(q)$ would tend pivot clockwise around its mid-point. The differences in tenure is small, so we do not expect this effect to be important in practice.} One thus broadly expects $\hat{\Delta}^q_{\Theta\Theta'}$ to be (i) negative and increasing under taste-discrimination, (ii) positive and decreasing under screening discrimination, and (iii) positive and increasing under complementary production. Figure 3 show the results. The three pictures are remarkably similar, showing a “ski jump” style curve. At the bottom of the distribution, $\hat{\Delta}^q_{\Theta\Theta'}$ is initially positive and decreasing, consistent with screening discrimination. This is particularly stark with black-Hispanic pairs: For example, at the 5th percentile the difference between same-race and

<table>
<thead>
<tr>
<th>Table 4: Tests of Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\Delta}^y_{WB} &lt; 0$ &amp; $\hat{\Delta}^y_{WB} &gt; 0$</td>
</tr>
<tr>
<td>$\hat{\Delta}^y_{WB} &gt; 1$</td>
</tr>
<tr>
<td>$\hat{\Delta}^y_{WB} &lt; 1$</td>
</tr>
<tr>
<td>$\hat{\Delta}^y_{WH} &lt; 0$ &amp; $\hat{\Delta}^y_{WH} &gt; 0$</td>
</tr>
<tr>
<td>$\hat{\Delta}^y_{WH} &gt; 1$</td>
</tr>
<tr>
<td>$\hat{\Delta}^y_{WH} &lt; 1$</td>
</tr>
<tr>
<td>$\hat{\Delta}^y_{BH} &lt; 0$ &amp; $\hat{\Delta}^y_{BH} &gt; 0$</td>
</tr>
<tr>
<td>$\hat{\Delta}^y_{BH} &gt; 1$</td>
</tr>
<tr>
<td>$\hat{\Delta}^y_{BH} &lt; 1$</td>
</tr>
</tbody>
</table>

(a) White-Black Relations (b) White-Hispanic Relations (c) Black-Hispanic Relations

Notes: Reports bootstrap results for 48,755 commissioned salespeople, as described in the text.
cross race productivity is about $\frac{\hat{\Delta}^q_{BH}}{2} = 0.1$ log points (or 10%). Around the median worker, there is little difference between same-race and cross-race workers. At the top, the gap becomes positive again, perhaps indicative of some complementarity. (In the model, we assume that complementarity is uniform across all workers, but this need not be the case.)

### 4.3 Turnover Results

Finally, we come to turnover. All three models predict that turnover is submodular. The third column of Table 3 shows that $\hat{\Delta}^q_{q_0}$ is negative for two pairs and essentially zero for the third; only the white-Hispanic combination is significantly negative. Thus we do not reject the model. Quantitatively, the average difference between same-race and cross-race turnover rates is $(\hat{\Delta}^q_{WB} + \hat{\Delta}^q_{WH} + \hat{\Delta}^q_{BH})/6 = -0.8\%$ compared to a base six-month turnover rate of 36.1%.

### 5 Discussion

#### 5.1 Sensitivity

In this section we discuss the sensitivity of our results in several dimensions.

**Productivity.** In Section 4 we estimate the productivity regression (11) via weighted-least squares for 10,000 bootstraps and present the median supermodularity coefficient (12) and corresponding standard errors. As a comparison, Table 5 estimates the productivity via weighted-least squares.
and clusters the standard errors at the worker level. One can see that the estimates and confidence intervals are very similar to those in Table 3.

We wish to understand whether the productivity estimates are driven by workers early or late in their tenure. Figure C2 in Online Appendix C recalculates Figure 3 for (a) the first six months of workers’ tenure, and (b) later months. One can see that the “ski-jump” pattern is remarkably consistent across all pairs and periods.

**Turnover.** Section 4.3 considered turnover after 6 months, showing that $\Delta x_{\Theta} < 0$ for all three pairs, with white-Hispanic being significant. The underlying regression is shown in Table 6. Table C1(a) in Online Appendix C shows that the results are robust to measuring turnover at 3, 9 and 12 months. All the coefficients are negative except black-Hispanic at 6 months; only the white-Hispanic coefficient are significantly negative.

Table C1(b) then includes the worker’s average productivity, $\frac{1}{T} \sum_{t=1}^{T_i} SPH_{i,t}$, as a control. As expected, this has a negative sign; e.g. with the “6 month” regression, the coefficient is $-0.0835$, meaning a one standard deviation increase in $SPH$ of 0.427 lowers turnover by 3.6% from a base of 36.1%. In the model, turnover is based on comparing the realized productivity with a threshold. If average SPH perfectly captures this realized productivity, then turnover should be modular under screening discrimination and complementary production, once we include it as a control. Under taste-based discrimination, the biased manager still discriminates against cross-race workers, and submodularity should increase.\(^{26}\) Table C1(b) shows that the submodularity tends to decrease.

**Full Sample of Workers.** Our results in Section 4 focus on 48,755 WBH newly hired salespeople for whom we have productivity data. We can also run the hiring regression (9) and turnover regression (15) on the entire sample of 812,399 WBH newly hired workers that includes cashiers, backroom operators, and service technicians. Table C2 in Online Appendix C shows our hiring results. As before, managers are significantly more likely to hire same-race workers. Specifically, the gap between same-race and cross-race hiring is 2.5pp, 3.2pp and 2.4pp for white, black and Hispanic workers, which is similar to the corresponding gaps for salespeople (3.9pp, 2.9pp and 4.0pp). Table C3 in Online Appendix C shows our turnover results. As before, turnover is submodular; since we have more data, these results are all significant at the 1% level. The size of these coefficients is ($\hat{\Delta}_{WB} + \hat{\Delta}_{WH} + \hat{\Delta}_{BH}$)/6 = $-2.9\%$ compared with $-0.8\%$ for salespeople. Table C1(c) shows these results are robust measuring at turnover at 3, 9 and 12 months.

**Peer Race.** Our hiring result in Table 2 shows how the race of recruits varies with manager race. Table C2 in Online Appendix C runs the same regression and also includes the race of managers’ other employees. Fixing the team composition, the gap between same-race and cross-race hiring

\(^{26}\)In Proposition 3(d) we saw that same-race candidate have lower turnover despite having a worse distribution of productivity. Once one controls for productivity, the amount of submodularity should increase.
Table 5: Productivity Regression

<table>
<thead>
<tr>
<th>Sales performance</th>
<th>Point estimate</th>
<th>95% confidence interval</th>
<th>Number of workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>White-White ((\hat{\beta}_{Ww}^{y}))</td>
<td>reference</td>
<td></td>
<td>26,847</td>
</tr>
<tr>
<td>White-Black ((\hat{\beta}_{Wb}^{y}))</td>
<td>0.0067</td>
<td>[-0.0140, 0.0273]</td>
<td>2,285</td>
</tr>
<tr>
<td>White-Hispanic ((\hat{\beta}_{Wh}^{y}))</td>
<td>-0.0133</td>
<td>[-0.0342, 0.0077]</td>
<td>2,122</td>
</tr>
<tr>
<td>Black-White ((\hat{\beta}_{Bw}^{y}))</td>
<td>-0.0200</td>
<td>[-0.0338, -0.0063]</td>
<td>6,172</td>
</tr>
<tr>
<td>Black-Black ((\hat{\beta}_{Bb}^{y}))</td>
<td>-0.0088</td>
<td>[-0.0337, 0.0161]</td>
<td>2,344</td>
</tr>
<tr>
<td>Black-Hispanic ((\hat{\beta}_{Bh}^{y}))</td>
<td>-0.0192</td>
<td>[-0.0518, 0.0134]</td>
<td>877</td>
</tr>
<tr>
<td>Hispanic-White ((\hat{\beta}_{Wh}^{y}))</td>
<td>-0.0150</td>
<td>[-0.0299, -0.0002]</td>
<td>4,557</td>
</tr>
<tr>
<td>Hispanic-Black ((\hat{\beta}_{Bh}^{y}))</td>
<td>-0.0382</td>
<td>[-0.0769, 0.0005]</td>
<td>710</td>
</tr>
<tr>
<td>Hispanic-Hispanic ((\hat{\beta}_{Hh}^{y}))</td>
<td>-0.0030</td>
<td>[-0.0266, 0.0205]</td>
<td>2,841</td>
</tr>
</tbody>
</table>

Location FEs Yes
Month FEs Yes
Department FEs Yes
Observations 335867 48,755

Pairwise relations
White-Black (\(\hat{\Delta}_{WB}^{y}\)) | 0.0045 | [-0.0263 , 0.0354] | 37,648 |
White-Hispanic (\(\hat{\Delta}_{WH}^{y}\)) | 0.0253 | [-0.0041 , 0.0547] | 36,367 |
Black-Hispanic (\(\hat{\Delta}_{BH}^{y}\)) | 0.0456 | [-0.0103 , 0.1016] | 6,772  |

Notes: This table shows the productivity regression (11) with standard errors clustered at the worker level. These parameters are estimated via WLS, inversely weighted to tenure. The supermodularity coefficient \(\hat{\Delta}_{WB}^{y}\) is defined by (12). The coefficient \(\hat{\Delta}_{WB}^{y}\) and corresponding confidence intervals differ a little from Table 3, which reports the median of the bootstrap estimates.

is 3.3pp, 2.5pp and 3.8pp for white, black and Hispanic salespeople, which is very similar to the benchmark regression (3.9pp, 2.9pp and 4.0pp). At the same time, peer race is predictive of worker race. For example, a manager of an all-white team has 7.7pp more chance of hiring a white worker than a manager of an all-black team (although the standard errors are large). This could reflect heterogeneous treatment effects, with some managers discriminating more than others. Or, it could be causal, with managers with more exposure to minorities being more likely to hire minorities.

Table C5 in Online Appendix C runs the benchmark productivity regression in Table 5 and also includes the share of same-race employees. All pairwise relations are supermodular, with white-Hispanic and black-Hispanic significant at the 5% level. The mean gap between same-race and cross-race productivity \((\hat{\Delta}_{WB}^{y} + \hat{\Delta}_{WH}^{y} + \hat{\Delta}_{BH}^{y})/6 = 0.9\) pp is compared to 1.3pp in the benchmark regression. Interestingly, the coefficient on peer race is negative, so a white worker in an all-white team is 1pp less productive than a white worker in a non-white team. Since our model does not speak to peer relations, we leave this for future research. For example, Benson and LePage (2022) use the data set to look at how a manager’s hiring decision depends on their past hires.
### Table 6: Turnover Regression

<table>
<thead>
<tr>
<th>Pairwise relations</th>
<th>Point estimate</th>
<th>95% confidence interval</th>
<th>Number of workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>White-White ($\hat{\beta}_{WW}$)</td>
<td>0.0033</td>
<td>[-0.0191, 0.0257]</td>
<td>21,613</td>
</tr>
<tr>
<td>White-Black ($\hat{\beta}_{WB}$)</td>
<td>0.0033</td>
<td>[-0.0191, 0.0257]</td>
<td>1,885</td>
</tr>
<tr>
<td>White-Hispanic ($\hat{\beta}_{WH}$)</td>
<td>-0.0128</td>
<td>[-0.0364, 0.0107]</td>
<td>1,694</td>
</tr>
<tr>
<td>Black-White ($\hat{\beta}_{BW}$)</td>
<td>0.0425</td>
<td>[0.0278, 0.0572]</td>
<td>5,007</td>
</tr>
<tr>
<td>Black-Black ($\hat{\beta}_{BB}$)</td>
<td>0.0304</td>
<td>[0.0081, 0.0527]</td>
<td>1,930</td>
</tr>
<tr>
<td>Black-Hispanic ($\hat{\beta}_{BH}$)</td>
<td>-0.0157</td>
<td>[-0.0512, 0.0198]</td>
<td>717</td>
</tr>
<tr>
<td>Hispanic-White ($\hat{\beta}_{HW}$)</td>
<td>-0.0073</td>
<td>[-0.0239, 0.0094]</td>
<td>3,699</td>
</tr>
<tr>
<td>Hispanic-Black ($\hat{\beta}_{HB}$)</td>
<td>-0.0104</td>
<td>[-0.0499, 0.0292]</td>
<td>572</td>
</tr>
<tr>
<td>Hispanic-Hispanic ($\hat{\beta}_{HH}$)</td>
<td>-0.0555</td>
<td>[-0.0767, -0.0344]</td>
<td>2,222</td>
</tr>
</tbody>
</table>

**Notes:** This table shows the turnover regression (15) for all newly hired WBH salespeople hired by WBH managers. The supermodularity coefficient $\hat{\Delta}_{WB}$ is defined by (16).

#### 5.2 Triangulating Individual Bias

Our supermodularity test examines relative bias. This has the advantage that it allows for worker and manager heterogeneity, and overcomes the problem of infra-marginality in the outcomes test. The downside is that it does not identify the bias of any one type of manager. For example, the fact that $\Delta_{BH}^v < 1$ could mean that black managers exhibit screening discrimination, Hispanic managers exhibit screening discrimination, or both. However, looking at Table 3, the three pairwise variance are ($\Delta_{WB}^v, \Delta_{WH}^v, \Delta_{BH}^v) = (0.981, 0.925, 0.701)$, which suggests that Hispanics are driving these results. We now propose a formalization of this intuition.

Say that managers **discriminate symmetrically** if

$$d_{\Theta \theta} = d_{\Theta \hat{\theta}} =: d_{\theta}, \ z_{\Theta \theta} = z_{\Theta \hat{\theta}} =: z_{\theta}, \text{ and } k_{\Theta \theta} = k_{\Theta \hat{\theta}} =: k_{\theta},$$  

so that manager $\theta$ treats all cross-race workers symmetrically. Assumption (19) makes sense if managers discriminate in favor of same-race candidates rather than against people of other races (e.g., because of shared culture). Empirically, this is consistent with Table 2, since we cannot reject
that cross-race hiring rates are statistically different from one another.\footnote{Formally, the hypothesis that \( r_{WB} = r_{WA} \) has a p-value of 0.23, \( r_{BW} = r_{BH} \) has a p-value of 0.55, and \( r_{HW} = r_{HB} \) has a p-value of 0.69. Assumption (19) is actually weaker than this, e.g. if white managers hired a lot of whites then \( r_{BW} < r_{BH} \), even if white managers treat black and Hispanic workers symmetrically.}

Under the symmetry assumption (19), the supermodularity test (3) becomes

\[
\Delta^y_{\Theta \Theta'} = -(d_\theta + d_{\Theta'}) \frac{\partial \tilde{y}(0)}{\partial d} - (z_\theta + z_{\Theta'}) \frac{\partial \tilde{y}(0)}{\partial z} - (k_\theta + k_{\Theta'}) \frac{\partial \tilde{y}(0)}{\partial k}.
\]

Using the three supermodularity estimates, we can then triangulate the impact of manager \( \theta \) on the mean productivity of cross-race workers,

\[
\Delta^y_{\theta} := \frac{1}{2} (\Delta^y_{\Theta \Theta'} + \Delta^y_{\Theta' \Theta'} - \Delta^y_{\Theta' \Theta'}) = -d_\theta \frac{\partial \tilde{y}(0)}{\partial d} - z_\theta \frac{\partial \tilde{y}(0)}{\partial d} - k_\theta \frac{\partial \tilde{y}(0)}{\partial d}
\]

Similarly, we can derive the impact of manager \( \theta \) on the variance,

\[
\log \Delta^v_{\theta} := \frac{1}{2} (\log \Delta^v_{\Theta \Theta'} + \log \Delta^v_{\Theta' \Theta'} - \log \Delta^v_{\Theta' \Theta'}) = -d_\theta \frac{\partial \nu(0)}{\partial d} - z_\theta \frac{\partial \nu(0)}{\partial d} - k_\theta \frac{\partial \nu(0)}{\partial d}
\]

We can then apply Proposition 4 to test for bias of individual managers. In particular, if \( \Delta^y_{\theta} < 0 \) then the manager exhibits taste-based bias, if \( \Delta^v_{\theta} < 1 \) then the manager exhibits screening discrimination, and if \( \Delta^y_{\theta} > 0 \) and \( \Delta^v_{\theta} > 1 \) then the manager exhibits complementary production.

We can illustrate the ideas with the numbers in Table 3. For productivity mean, \((\hat{\Delta}^y_{w}, \hat{\Delta}^y_{b}, \hat{\Delta}^y_{h}) = (-0.0029, 0.0090, 0.0297)\), and for the variance, \((\hat{\Delta}^v_{w}, \hat{\Delta}^v_{b}, \hat{\Delta}^v_{h}) = (1.138, 0.862, 0.813)\). We thus see that black and (especially) Hispanic managers have supermodular productivity and submodular variance, consistent with screening discrimination, whereas white managers have (slightly) submodular productivity and supermodular variance, consistent with taste-based discrimination or complementary production. These results rely on the symmetry assumption (19), so one should interpret them with caution. But this approach illustrates the benefit of having a study with multiple racial groups.

### 5.3 Internal Transfers

In Section 2-4 we provide and execute a test of discrimination using new recruits. Further evidence can be obtained from studying internal transfers. The hope is that one can compare separate different theories of discrimination by comparing a worker’s performance under their hiring manager and their current manager. The problem is that internal transfers are often not random, meaning we may see the same selection issues with internal transfers as seen with new hires, and that we see relatively few workers transition between managers.

A proper analysis of transfers warrants a paper unto itself, but a brief theoretical discussion is useful. Assume transfers are random (e.g. when a manager is promoted, the replacement is independent of the ethnic composition of the team). Under taste-based discrimination and screening
discrimination the agents’ productivity is fixed and there should be no relation between the worker’s productivity and the new manager’s race; hence productivity should be modular, $\Delta_{\Theta^\prime} = 0$. Under complementarity, workers are still more productive under same-race managers, so productivity should be supermodular, $\Delta_{\Theta^\prime} > 0$. Conversely, if transfers are selective, one would broadly expect the results from Section 2 to still hold.\textsuperscript{28}

To test the assumption of random transfers, Table C6 in Online Appendix C recalculates the hiring test for internal transfers. This shows that managers hire about 4pp more same-race candidates (a similar number to Table 2). The standard errors are rather large, but this suggests that not all transfers are random.

As a second test, we can study what happens to a worker’s productivity when they change manager. If we restrict the sample of salespeople to those who change managers and run the productivity regression (11) including pairwise dummies for the race of the hiring manager, the controls in (9), and a dummy for when the worker moves manager. The coefficient on moving is $-0.78$pp with standard error of $0.34$pp, meaning the average worker’s productivity falls slightly when they move. This suggests that not all transfers are selective.

Finally, Table C7 in Online Appendix C runs the supermodularity regression (11) for salespeople working under non-hiring managers. Unfortunately there are relatively few data points, and the standard errors on the supermodularity coefficients $\hat{\Delta}^y$ are too large to make any conclusions.\textsuperscript{29}

5.4 Alternative Theories

This paper has considered three theories (taste-based discrimination, screening discrimination, complementary production). Here we briefly discuss other factors that might be at play.

Referral Networks. Our model assumes that each manager selects from an equal distribution of applicants (controlling for store, department and month fixed effects). This excludes the possibility that the manager may obtain referrals through their social networks. If referrals have the productivity distribution as other applicants of that race, this would violate our random assignment assumption, raise the number of same-race hires, but would not affect productivity. However, if referrals are positively selected (e.g. Beaman and Magruder (2012), Pallais and Sands (2016)) then they would generate supermodular mean productivity. Referrals might be better because the manager receives valuable information from their social contacts; this would look like our screening

\textsuperscript{28}To formally model selective transfers is complicated. Presumably such transfers only occur if there if both the worker and the manager benefit. This means having to understand managers’ preferences (e.g. taste-based discrimination), managers’ information (e.g. screening discrimination), workers’ productivity (e.g. complementary production), workers’ comparative advantage at different jobs, and the workers’ preferences.

\textsuperscript{29}One might also consider looking at workers who experience a change in the race of their manager. Unfortunately, there is not a lot of such variation. For example, if one looks at white workers hired by a white manager who transfer internally, less than 7% work for a cross-race manager.
model. Alternatively, referrals might be have a higher mean distribution \( \mu_\Theta \); this would look like our complementary production model.

It is hard to evaluate the importance of referrals since we do not observe applications. At an anecdotal level, one HR representative told us that, in her view, it is rare for managers to know applicants. We can also look for indirect evidence of referrals. Brown, Setren, and Topa (2016) find that “most referrals take place between a provider and a recipient with similar characteristics in terms of age, gender, ethnicity, education, and division and staff level within the corporations.” Thus, if more of the same-race workers are referrals, one might expect the same-race workers share other characteristics with the manager. Table C8 in Online Appendix C looks at the average age and gender differences between the worker and manager by each racial pair and show a precise zero estimate.

**Favoritism.** Our theory allows for complementarity, whereby a worker is more productive with a same-race manager (e.g. Becker (1973), Lang (1986)). This is different from favoritism, whereby a manager gives preferential treatment to same race workers. The latter is intrinsically zero-sum, e.g. assigning same-race workers to the best shifts. Our measure of productivity controls for the worker’s shift via their target sales, eliminating the most obvious source of favoritism. In terms of theory, favoritism doesn’t directly affect hiring. But, when paired with taste-based discrimination, a manager could hire low-productivity same-race workers, and then artificially boost their productivity by giving them the best shifts. If the favoritism was sufficient to overcome the low initial productivity, this would look like complementarity in our tests. Ultimately we cannot differentiate between these hypotheses, and one should view “complementarity” as a catch-all for same-race favoring actions.\(^{30}\)

**Worker Preferences.** The model of taste-discrimination assumes that managers are biased towards same-race workers. An alternative model is that managers are neutral, but workers are biased towards same-race managers, accepting the offer with a higher probability when the manager shares the same race. Such a model predicts the same managers hire more same-race candidates. The impact on productivity depends on the details of the model. If a random selection of workers are biased and turn down offers, productivity is modular. Alternatively, if workers have a fixed outside option that depends on the race of the manager, then low-productivity workers will be more likely to reject cross-race offers, meaning productivity is submodular. Ultimately, we cannot rule out such forces, but they seem less important with hiring decisions than retention or transfer decisions.

\(^{30}\)One might be able use data on the entire composition of the team (e.g. favoritism is irrelevant when the entire team is one race). But identification is unclear if managers are heterogeneous in their biases (e.g. the manager with the single-race team may be particularly discriminatory). These are interesting directions that we leave for future work.
5.5 Long-Run Implications of Discrimination

The results presented in Section 4 provide evidence of screening discrimination, but not of taste-based discrimination. This conclusion may be somewhat of a relief. It is not a surprise that people with similar backgrounds see past stereotypes and are better at assessing individual characteristics. And this feels less pernicious than if managers have extensive taste-based biases. However, we now argue that screening discrimination systematically disadvantages minority candidates.\footnote{This logic is similar to Müller-Itten and Öry (2022), which focuses on mentoring.}

Consider a simple model of employment dynamics based on Section 2. There is a continuum of agents who belong to one of two races. Fraction $n > \frac{1}{2}$ of the population is of the “majority” race and fraction $1 - n$ is of the “minority” race. The races are otherwise symmetric. A firm employs initially fraction $M_0$ managers and $N_0$ workers of the majority race. These fractions $\{M_t, N_t\}$ evolve over continuous time. At each point in time, workers and managers quit at an exogenous rate and are replaced. When a worker leaves, the manager sequentially looks at workers until one passes the bar, in that the manager’s signal exceeds the hiring threshold. As in Propositions 1(a)-3(a), a manager hires each same-race candidate with probability $p_s$ and each cross-race candidate with probability $p_c < p_s$; assume these numbers are the same for majority and minority managers. When a manager leaves, a random worker is promoted into the vacancy.

**Proposition 6.** Starting from any initial condition $\{M_0, N_0\}$, the system converges to a unique steady state $\{M^*, N^*\}$. Workers and managers have the same composition of races, and the majority population is over-represented, $M^* = N^* > n$.

**Proof.** See Online Appendix B.2.

Minorities are at a disadvantage simply because they are minorities. Intuitively, the problem is that most minorities are interviewed by a manager who discriminates against them; the exact source of the discrimination does not matter. We can use the hiring numbers in Table 2 to understand the quantitative significance of the result. Given these hiring numbers, the firm converges to a steady state of $(N^*_W, N^*_B, N^*_H) = (64.2, 19.4, 16.4)$. We don’t know what “neutral hiring” looks like, since we do not know the bias of the different managers, but if we pick a manager race $\theta \in \{w, g, h\}$ with probability $\{1/3, 1/3, 1/3\}$, we obtain $(N_W, N_B, N_H) = (63.1, 19.8, 17.2)$ by taking the average of the columns in Table 2. Thus white workers are over-represented by 1.1pp while Hispanics are under-represented by 0.8pp. This is quantitatively significant: There are 4.8% fewer Hispanics employed than in a “neutral” workplace simply because the Hispanics are in the minority. Moreover this calculation may underestimate the true effect: If productivity is supermodular in race (as in Table 3), minority workers and managers have lower average productivity than their majority counterparts; one would thus expect a minority workers to be promoted to managerial positions less frequently than majority workers.
This “institutional bias” can be mitigated if the firm promotes more minority managers. This is especially useful as a way to get to steady state, but it is harder to use such a policy to shift an inequitable steady state. Indeed, Appendix B.2 shows that as the number of minority workers vanish, the number of minority managers would have to equal $p_c/(p_c + p_s)$ to achieve equitable recruitment, meaning the ratio of minority managers to workers grows without bound. A different strategy is to centralize recruitment and take away the job of hiring from the day-to-day manager. For example, one scheme would match the candidate with the interviewer best equipped to interview them on a variety of characteristics (e.g. race, age, school, gender). One can then evaluate the interviewer based on their acceptance rate and the subsequent productivity of their hires.

6 Conclusion

The purpose of this paper is to examine three theories of relative discrimination. Methodologically, we derive the theoretical predictions for each theory, show how to combine them in one common framework, and derive a novel test for discrimination. The data shows statistically and economically significant discrimination in hiring for all three pairs of races. For black-Hispanic pairs, screening discrimination plays a role. For white-Hispanic pairs, a combination of screening discrimination and complementary production play a role. Other forms of discrimination may also be present.

These findings may be relevant for policy. As discussed in Section 5.5, screening discrimination may be just as detrimental to the employment opportunities of minorities as taste-based discrimination, though the solutions are quite different. In particular, it suggests policies that improve the ability of the firm to identify talent (e.g. diverse interviewers, algorithmic hiring, improved incentives and training for managers).

Our tests of discrimination can be easily applied by firms to measure discrimination along the lines of race, gender, age, disability, educational background, national origin, and so on. For example, one might be interested in testing whether young and old managers treat LGBTQ applicants differently. The test is based on recruiting numbers and realized output, which allows firms to monitor their human resource strategies in real time. The method is also applicable in other contexts, wherever there is a continuous output variable. For example, one can use GPA to evaluate college admissions or profitability to study venture capital funding.

One limitation of the paper is that our tests are qualitative. We thus ignore quantitative relations imposed by a model between, say, the bias in hiring and the supermodularity of productivity. Such relationships might pin down the extent of discrimination, rather than identifying the presence of one or two types. However, this does not seem possible with our data since one must identify the prior distributions and the biases. Having a quantitative understanding of the various biases is important for policy prescriptions, and is an important topic for future research.
Appendix

A Proofs for Section 2

A.1 Useful Properties of the Normal Distribution

**Normal distribution:** Let \( \phi(\alpha) = \exp(-\alpha^2/2)/\sqrt{2\pi} \) be the density of the normal distribution, and \( \Phi \) be the corresponding cdf. We note the derivative is

\[
\phi'(\alpha) = -\alpha \phi(\alpha). \tag{20}
\]

Moreover, the density is log-concave since \((\log \phi)' = -1\).

**Truncated normal:** Given a normal random variable \( \hat{y} \sim N(\mu - k, \eta^2) \) and cutoff \( y^* + d \), the expectation and variance of the truncated normal are given by:

\[
E[\hat{y}|\hat{y} \geq y^* + d] = \mu - k + \lambda(\alpha) \eta \tag{21}
\]

\[
Var[\hat{y}|\hat{y} \geq y^* + d] = \eta^2 (1 + \alpha \lambda(\alpha) - \lambda(\alpha)^2) \tag{22}
\]

where \( \alpha := (y^* + d - (\mu - k))/\eta \) is the normalized cutoff and \( \lambda(\alpha) = \phi(\alpha)/(1 - \Phi(\alpha)) \) is the hazard rate of the standard normal. We recall the estimator variance \( \eta = \sqrt{\sigma_0^2/(\sigma_0^2 + \sigma^2(1 + z)^2)} \) and write \( \eta' = \partial \eta/\partial z < 0 \); greater signal noise reduces the estimator variance, as discussed in Section 2. The normalized cutoff then depends on our parameters as follows

\[
\frac{\partial \alpha}{\partial d} = \frac{1}{\eta} > 0 \quad \frac{\partial \alpha}{\partial z} = -\frac{\alpha \eta'}{\eta} > 0 \quad \frac{\partial \alpha}{\partial k} = \frac{1}{\eta} > 0. \tag{23}
\]

Write \( \delta = \hat{y} - (y^* + d) \geq 0 \) for the **excess productivity**, i.e. the random variable that measures the amount by which expected productivity \( \hat{y} \) exceeds the hiring threshold \( y^* + d \).

**Lemma 1.** Excess productivity \( \delta \) falls in \( d, z, \) and \( k \) in the likelihood ratio order (MLRP).

**Proof.** The density of \( \delta \) is given by

\[
g(\delta) = g(\delta; d, z, k) := \frac{1}{\eta(z)} \phi \left( \frac{y^* + d + \delta - (\mu - k)}{\eta(z)} \right) / \left( 1 - \Phi \left( \frac{y^* + d + \delta - (\mu - k)}{\eta(z)} \right) \right). \]

Thus, \( \delta \) MLRP-decreases in \( d \) if

\[
\frac{g(\delta; \tilde{d}, z, k)}{g(\delta; d, z, k)} < \frac{g(\delta; \tilde{d}, z, k)}{g(\delta; d, z, k)}
\]

for all \( \delta < \tilde{\delta} \) and \( d \leq \tilde{d} \). Rearranging, we obtain

\[
\log g(\delta; \tilde{d}, z, k) + \log g(\delta; d, z, k) \leq \log g(\delta; \tilde{d}, z, k) + \log g(\delta; d, z, k),
\]

---

Easily derived with integration by parts, or found on, e.g., https://en.wikipedia.org/wiki/Truncated_normal_distribution
which is to say that the density is log-submodular in $\delta$ and $d$. Differentiating,

$$\frac{\partial^2(\log g)}{\partial y^* \partial \delta} = \frac{\partial^2}{\partial d \partial \delta} \log \phi \left( \frac{y^* + d + \delta - (\mu - k) \eta(z)}{\eta(z)} \right) = \frac{1}{\eta(z)^2} \left( \log \phi \left( \frac{y^* + d + \delta - (\mu - k) \eta(z)}{\eta(z)} \right) \right)'' < 0$$

as required. This argument also shows that $\delta$ MLRP-increases in $k$.

To see that $\delta$ MLRP-decreases in $z$, we wish to show that $g(\delta; d, z, k)$ is log-submodular in $\delta$ and $z$. We write $\psi(\delta, z) = y^* + d + \delta + (\mu - k) \eta(z)$, note the derivatives $\frac{\partial \psi}{\partial \delta} > 0, \frac{\partial \psi}{\partial z} > 0, \frac{\partial^2 \psi}{\partial \delta \partial z} > 0$, and compute

$$\frac{\partial^2(\log g)}{\partial \delta \partial z} = \frac{\partial^2}{\partial \delta \partial z} \log \phi(\psi(\delta, z)) = \frac{\partial \psi}{\partial \delta} \cdot \frac{\partial \psi}{\partial z} \cdot [\log \phi(\psi(\delta, z))]'' + \frac{\partial^2 \psi}{\partial \delta \partial z} \cdot [\log \phi(\psi(\delta, z))]' < 0.$$

\[\square\]

**Lemma 2.** The hazard rate of the standard normal $\lambda(\alpha)$ has the following properties:

(a) $\lambda(\alpha)$ is increasing with derivative $\lambda' = \lambda(\lambda - \alpha) \in (0, 1)$,
(b) $\lambda(\alpha)$ is bounded below by $\lambda(\alpha) > \alpha > \alpha \lambda'(\alpha),$
(c) $\lambda(\alpha)$ is convex with second derivative $\lambda'' = (2\lambda - \alpha)\lambda' - \lambda$.

**Proof.** We prove (a) and (b) out of order. Using (21), first note that $\lambda(\alpha) = E[\zeta | \zeta \geq \alpha] > \alpha$, where $\zeta \sim N(0, 1)$. Using (20), this has slope $\lambda' = \lambda(\lambda - \alpha)$, which is positive given $\lambda > \alpha$. To prove the slope is less than one, we specialize Lemma 1 to the standard normal with cutoff $\alpha$, and excess $\delta = \zeta - \alpha$ with density $g(\delta; \alpha)$, so that

$$\lambda(\alpha) = E[\zeta | \zeta \geq \alpha] = \alpha + \int_{\delta=0}^{\infty} \delta g(\delta; \alpha) d\delta. \quad (24)$$

By the proof of Lemma 1, $\delta$ MLRP-falls in $\alpha$. Hence the second term in (24) falls in $\alpha$, so $\lambda'(\alpha) < 1$. This completes the proof of (a) and (b). The convexity of the hazard rate is well known.\[^{33}\]

\[\square\]

**A.2 Proof of Propositions 1-4**

Propositions 1-3 follow immediately from (2), which we replicate here for convenience

- **Taste-Based Discrimination:** $\frac{\partial p}{\partial d} < 0, \frac{\partial \bar{y}}{\partial d} > 0, \frac{\partial v}{\partial d} < 0, \frac{\partial x}{\partial d} > 0$
- **Screening Discrimination:** $\frac{\partial p}{\partial z} < 0, \frac{\partial \bar{y}}{\partial z} < 0, \frac{\partial v}{\partial z} > 0, \frac{\partial x}{\partial z} > 0$
- **Complementary Production:** $\frac{\partial p}{\partial k} < 0, \frac{\partial \bar{y}}{\partial k} < 0, \frac{\partial v}{\partial k} < 0, \frac{\partial x}{\partial k} > 0$

While the propositions are organized by theory (the rows in the matrix), it is convenient to organize the proof by the outcome variables (i.e. the columns of the matrix).

\[^{33}\]E.g., https://math.stackexchange.com/questions/1349555/standard-normal-distribution-hazard-rate
**Hiring probability:** The hiring probability $p = \Pr(\hat{y} > y^* + d) = 1 - \Phi(\alpha)$ clearly falls in $\alpha$ and thus, by (23), in $d$, $z$, and $k$.\(^{34}\)

**Turnover:** Recalling the excess productivity $\delta := \hat{y} - (y^* + d) \geq 0$ from Lemma 1, and writing $\tilde{\epsilon} := \hat{y} - y \sim N(0, \gamma^2)$ for the difference between estimated and actual productivity, the worker is fired iff $\tilde{\epsilon}$ exceeds the sum of the two safety margins $\delta$ and $\tau$. Conditional turnover “$x|\delta$” thus equals $\Pr(\tilde{\epsilon} \geq \delta + \tau) = 1 - \Phi((\delta + \tau)/\gamma(z))$, which falls in $\delta$ and rises in the residual variance $\gamma^2(z) = \sigma^2_\delta \sigma^2_\epsilon (1 + z)^2 / (\sigma^2_\delta + \sigma^2_\epsilon (1 + z)^2)$, which in turn rises in $z$. In expectation over $\delta$, turnover becomes

$$x = \Pr(\tilde{\epsilon} \geq \delta + \tau|\delta \geq 0) = \int_{\delta \geq 0} (1 - \Phi((\delta + \tau)/\gamma(z)))g(\delta; d, z, k)d\delta. \quad (25)$$

By Lemma 1, a rise in $d$ or $k$ MLRP-decreases excess productivity $\delta$ and so MLRP-raises conditional turnover $x|\delta$. A rise in $z$ also MLRP-raises $x|\delta$ , and additionally raises $x|\delta = 1 - \Phi((\delta + \tau)/\gamma(z))$ directly, so turnover rises.

**Expected Productivity:** Differentiating (21) $\bar{y} = E[\hat{y}|\hat{y} \geq y^*] = \mu - k + \lambda(\alpha)\eta$

$$\frac{\partial \bar{y}}{\partial d} = \lambda' > 0 \quad \frac{\partial \bar{y}}{\partial z} = (\lambda - \alpha\lambda') \eta' < 0 \quad \frac{\partial \bar{y}}{\partial k} = -1 + \lambda' < 0, \quad (26)$$

where the derivatives rely on (23), and the inequalities on Lemma 2.

**Productivity Variance:** Applying the law of total variance, using (22), and Lemma 2(a), productivity variance equals

$$v = \text{Var}(y|\hat{y} \geq y^*) = E[\text{Var}(y|\hat{y})|\hat{y} \geq y^*] + \text{Var}(E[y|\hat{y}]|\hat{y} \geq y^*)$$

$$= \text{Var}(y|\hat{y}) + \text{Var}(\hat{y}|\hat{y} \geq y^*) = \gamma^2 + \eta^2(1 + \alpha\lambda(\alpha) - \lambda(\alpha)^2) = \sigma^2_\delta - \eta^2\lambda'(\alpha). \quad (27)$$

Differentiating and using (23) and Lemma 2(b,c)

$$\frac{\partial v}{\partial d} = -\eta\lambda'' = -\eta((2\lambda - \alpha)\lambda' - \lambda) < 0 \quad \frac{\partial v}{\partial z} = \eta(2\lambda - \alpha)(\alpha\lambda' - \lambda) \eta' > 0 \quad \frac{\partial v}{\partial k} = \frac{\partial v}{\partial d} < 0 \quad (28)$$

**Proof of Proposition 4:** Dividing (28) by (26), we get

$$0 < -\frac{\partial v}{\partial \bar{y}}/\partial d = \eta[(2\lambda - \alpha) - \lambda/\lambda'] < \eta(2\lambda - \alpha) = \frac{\partial v/\partial z}{-\partial \bar{y}/\partial z}. \quad (29)$$

\(^{34}\)Here we use the assumption that managers hire less than half of the applicants, $y^* > \mu$, which guarantees $\alpha = y^* + d - (\mu - k) > 0$, which in turn underlies $\partial \alpha/\partial z > 0$ in (23).
References


Online Appendix

B Omitted Proofs

B.1 Maximizing Expected Sales

In this appendix we analyze the model variant in which managers maximize expected sales, as discussed in Section 2.5. Let $Y$ be a worker’s sales, and assume a manager wishes to hire workers with $Y \geq Y^*$. As in the main text, let $\hat{y} = E[y|\tilde{y}]$ be expected log-sales, with mean $\mu$, estimator variance $\eta^2 = \sigma_0^2/\sigma_0^2 + \sigma^2$, and estimation error $y - \hat{y}$ with residual variance $\gamma^2 := \sigma_0^2\sigma^2/\sigma_0^2 + \sigma^2$.

The manager’s expected utility from a worker given estimate $\hat{y}$ is thus

$$E[Y|\hat{y}] = e^{\hat{y}} E[e^{y-\hat{y}}|\hat{y}] = e^{\hat{y} + \gamma^2/2}.$$

The manager thus wishes to hire a worker with estimate $\hat{y} \geq y^* := \log Y^* - \gamma^2/2$. Intuitively, managers lower their standard, $y^* < \log Y^*$, since they are risk-loving with respect to uncertainty in log-productivity.

**Taste-Based Discrimination.** In this model, biased managers gain extra utility from same-race candidates. Suppose they gain utility $Y/D$ from a cross-race candidate, and $Y$ from a same-race candidate, where $D > 1$ is the same-race bias. Taking logs, the cross-race cutoff $y^*_c = \log Y^* + d - \gamma^2/2$ thus exceeds the same-race cutoff $y^*_s = \log Y^* - \gamma^2/2$ by the bias $d = \log D$. The analysis of hiring is thus identical to that in Section 2.1, and thus the proofs of Propositions 1(a)-(c) carry over unchanged. When it comes to turnover, suppose that the manager fires the worker if their realized productivity $y$ drops more than $\tilde{\tau}$ below the cutoff $y^*_i$. Thus the turnover rate for $i \in \{c, s\}$ is

$$x_i := \Pr(y \leq y^*_i - \tilde{\tau}|\hat{y} \geq y^*_i) = \Pr(y \leq y^*_i - \tau|\hat{y} \geq y^*_i).$$

where $\tau := \tilde{\tau} - \gamma^2/2$. This is the same as in Section 2.1, and Proposition 1(d) immediately applies.

**Complementary Production.** As with taste-based discrimination, the analysis of the baseline case (where managers maximize expected log-sales) is essentially unchanged. Since the residual variance is fixed at $\gamma^2$, the cross-race cutoff $y^*_c$ exceeds the same-race cutoff $y^*_s$ by $k$, and the firing threshold is $\tau$ below the hiring threshold. Proposition 3 immediately generalizes.

**Screening Discrimination.** In this model, the signal variance $\sigma^2_i$ is lower for same-race applicants than for cross-race applicants, $\sigma^2_s \leq \sigma^2_c$. This implies that cross-race hires have more residual variance than same-race hires, $\gamma^2_s \leq \gamma^2_c$. Thus, managers set a lower threshold for cross-race hires,
\( y_c^* = \log Y^* - \gamma_c^2/2 < \log Y^* - \gamma_s^2/2 = y_s^* \). Intuitively, managers are risk-loving with respect to the log-productivity, and thus like cross-race candidates because of their high residual variance. We first verify that managers still hire more same-race workers.

**Proposition 2(a)**. Assume \( \log Y^* > \mu + \sigma_0^2 \). In the screening model, the hiring probability is higher for same-race applicants than cross-race applicants, \( p_s > p_c \).

**Proof.** For part (a), let \( \Phi \) and \( \phi \) be the cdf and pdf of the standard normal distribution. Write the precision of the prior as \( h_0 := 1/\sigma_0^2 \) and the precision of the raw signal as \( h_\epsilon = 1/\sigma_\epsilon^2 \). The estimator variance and residual variance are then \( \eta^2 = h_\epsilon/(h_0 + h_\epsilon)h_0 \) and \( \gamma^2 = 1/(h_0 + h_\epsilon) \). Now, the acceptance rate can be written as a function of the signal precision

\[
p(h_\epsilon) = \Pr \left( \hat{y} \geq \log Y^* - \frac{\gamma^2}{2} \right) = 1 - \Phi \left( \frac{\log Y^* - \mu - \frac{\gamma^2}{2}}{\eta} \right) = 1 - \Phi \left( \sqrt{\frac{h_0 + h_\epsilon}{h_\epsilon}} (\log Y^* - \mu) - \frac{1}{2} \sqrt{\frac{h_0}{(h_0 + h_\epsilon)h_\epsilon}} \right)
\]

Simple computations show that the derivative equals

\[
p'(h_\epsilon) = -\phi(\cdot) \times \sqrt{h_0} \left( -\frac{h_0}{2h_\epsilon} \sqrt{\frac{1}{h_\epsilon(h_0 + h_\epsilon)}} (\log Y^* - \mu) + \frac{1}{4} \left( \frac{1}{h_\epsilon} + \frac{1}{h_0 + h_\epsilon} \right) \sqrt{\frac{1}{(h_0 + h_\epsilon)h_\epsilon}} \right)
\]

\[
= \phi(\cdot) \times \frac{1}{2} \sqrt{\frac{h_0}{(h_0 + h_\epsilon)h_\epsilon}} \frac{h_0}{h_\epsilon} \left( (\log Y^* - \mu) - \frac{1}{2} \frac{1}{h_0} \left( 1 + \frac{h_\epsilon}{h_0 + h_\epsilon} \right) \right)
\]

Given that \( h_\epsilon/(h_0 + h_\epsilon) \leq 1 \), this last term is positive if \( \log Y^* \geq \mu + \sigma_0^2 \). Since same-race candidates have greater signal precision \( h_\epsilon \), they have a higher acceptance rate, and so managers hire relatively more same-race workers. \( \square \)

Turning to productivity, the proof of Proposition 2(b),(c) shows that productivity of cross-race hires has lower mean and greater variance when using the same bar, \( y_c^* = y_s^* \). When risk-neutral managers lower the bar for cross-race recruits, this further lowers their expected productivity and raises productivity variance. Finally, when we consider turnover, the proof of Proposition 2(d) shows that cross-race hires have higher turnover when using the same bar, \( y_c^* = y_s^* \). When we lower the bar for cross-race recruits, this further raises their turnover.

**B.2 Proof of Proposition 6**

Let \( p_s \) and \( p_c \) be the probability same-race and cross-race candidates are hired. Let \( M_t \in [0, 1] \) and \( N_t \in [0, 1] \) be the number of \( A \) managers and workers at a firm, and consider the dynamics \( (N_t, M_t) \) in \((N, M)\) space (see Figure B1).
First, consider the dynamics of workers, given the number of managers. Let \( \psi \) be the rate at which workers and managers leave; this could be due to exit or promotion. A race-A manager hires race-A workers over race-B workers in the ratio \( p_s n / p_c (1 - n) \). Hence the number of race-A workers evolves according to

\[
\dot{N} = \psi \left[ (1 - M) \left( \frac{p_b n}{p_c n + p_s (1 - n)} \right) + M \left( \frac{p_s n}{p_s n + p_c (1 - n)} \right) \right] - N. \tag{30}
\]

Thus, in \((N, M)\) space, \( \dot{N} = 0 \) is a straight line from \((\frac{p_s n}{p_c n + p_s (1 - n)}, 0)\) to \((\frac{p_s n}{p_s n + p_c (1 - n)}, 1)\), so with slope above one.

Next, consider the dynamics of managers, given the number of workers. Managers retire at rate \( \psi \) and replacements are chosen randomly from the population of workers, so \( \dot{M} = \psi (N - M) \). Thus, \( \dot{M} = 0 \) along the diagonal where \( N = M \).

Figure B1 illustrates the dynamics of \((N, M)\). It is evident that there is a unique steady state where \( \dot{M} = 0 \) and \( \dot{N} = 0 \), and that the system converges to this steady state from any initial condition. When \( N = M = n \), then

\[
\dot{N} = \psi \left[ (1 - n) \left( \frac{p_b n}{p_c n + p_s (1 - n)} \right) - n \left( \frac{p_c (1 - n)}{p_s n + p_c (1 - n)} \right) \right] = (p_s - p_c) (2n - 1) > 0
\]

if \( n > 1/2 \). Thus the steady state must satisfy \( N^* = M^* > n \).

Finally, we consider how many managers of a particular race are required to obtain equitable hiring. Rearranging (30), if we have \( \dot{N} = 0 \) when \( N = n \), then \( M \) must solve

\[
M \left( \frac{(p_s^2 - p_c^2) n (1 - n)}{(p_s n + p_c (1 - n)) (p_c n + p_s (1 - n))} \right) = \frac{(p_s - p_c) (1 - n) n}{p_s n + p_c (1 - n)}. \]
Letting $1 - n = \epsilon \approx 0$, we obtain

$$M \left( \frac{p_s^2 - p_c^2}{p_s p_c} \right) \epsilon \approx \frac{(p_s - p_c)}{p_c} \epsilon,$$

so that the number of race-B managers is $1 - M \approx \frac{p_c}{p_s + p_c}$. 
C Omitted Tables and Figures

![Histograms of SPH by Manager Race, Worker Race](image)

**Figure C1: Histograms of SPH by Manager Race, Worker Race**

*Notes:* The left hand figure shows the distribution of productivity for white workers; the black line is productivity under white managers, the green line is productivity under black managers, and the blue line productivity under Hispanic managers. The statistical power of the data afford us the ability to plot the frequency distributions for every combination of racial pair at 5%-wide intervals of SPH without the use of kernel densities or other smoothing.
Figure C2: Supermodularity of Productivity by Quantile

Notes: This figure recreates Figure 3 with data above/below six months. Specifically, we partition each worker’s performance before/after 6 months of tenure. On each partition, it then shows $\hat{\Delta}_q^q$ as defined in (18) across 98 one-percentile bins, after we have thrown away the top and bottom bins. The Lowess curve has bandwidth 0.8.
### Table C1: Turnover Regression: Sensitivity

<table>
<thead>
<tr>
<th>Pairwise Relations</th>
<th>White-Black ($\hat{\Delta}^{\tau}_{WB}$)</th>
<th>White-Hispanic ($\hat{\Delta}^{\tau}_{WH}$)</th>
<th>Black-Hispanic ($\hat{\Delta}^{\tau}_{BH}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Salespeople</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 months</td>
<td>-0.0144 (0.0150)</td>
<td>-0.0130 (0.0155)</td>
<td>-0.0316 (0.0267)</td>
</tr>
<tr>
<td>6 months</td>
<td>-0.0154 (0.0171)</td>
<td>-0.0354 (0.0175)</td>
<td>0.0009 (0.0305)</td>
</tr>
<tr>
<td>9 months</td>
<td>-0.0221 (0.0182)</td>
<td>-0.0444 (0.0185)</td>
<td>-0.0293 (0.0322)</td>
</tr>
<tr>
<td>12 months</td>
<td>-0.0224 (0.0186)</td>
<td>-0.0442 (0.0189)</td>
<td>-0.0417 (0.0330)</td>
</tr>
<tr>
<td>(b) Salespeople, controlling for SPH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 months</td>
<td>-0.0259 (0.0160)</td>
<td>0.0024 (0.0170)</td>
<td>-0.0272 (0.0285)</td>
</tr>
<tr>
<td>6 months</td>
<td>-0.0191 (0.0182)</td>
<td>-0.0023 (0.0192)</td>
<td>0.0222 (0.0324)</td>
</tr>
<tr>
<td>9 months</td>
<td>-0.0326 (0.0194)</td>
<td>-0.0069 (0.0203)</td>
<td>-0.0082 (0.0344)</td>
</tr>
<tr>
<td>12 months</td>
<td>-0.0282 (0.0205)</td>
<td>-0.0102 (0.0216)</td>
<td>-0.0184 (0.0365)</td>
</tr>
<tr>
<td>(c) All workers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 months</td>
<td>-0.0289 (0.0039)</td>
<td>-0.0566 (0.0040)</td>
<td>-0.0491 (0.0066)</td>
</tr>
<tr>
<td>6 months</td>
<td>-0.0382 (0.0046)</td>
<td>-0.0730 (0.0048)</td>
<td>-0.0641 (0.0078)</td>
</tr>
<tr>
<td>9 months</td>
<td>-0.0379 (0.0045)</td>
<td>-0.0887 (0.0047)</td>
<td>-0.0789 (0.0077)</td>
</tr>
<tr>
<td>12 months</td>
<td>-0.0314 (0.0043)</td>
<td>-0.0945 (0.0045)</td>
<td>-0.0800 (0.0073)</td>
</tr>
</tbody>
</table>

**Notes:** This table shows how the supermodularity coefficient for the turnover regression (15) depends on our assumptions. Part (a) shows the regression for salespeople where $TURN_i$ is classified as leaving after 3, 6, 9 and 12 months. The “6 month” row corresponds to Table 6. Part (b) shows the same regressions for salespeople, where we add average productivity, $SPH_i$ as a control variable. In the “6 month” model the coefficient on $SPH_i$ is $-0.0835$ with 95% confidence interval $[-0.0948, -0.0722]$. Part (c) shows the same regression as part (a), for all workers rather than just salespeople.
### Table C2: Hiring Tests for All Workers

<table>
<thead>
<tr>
<th></th>
<th>White worker</th>
<th>Black worker</th>
<th>Hispanic worker</th>
</tr>
</thead>
<tbody>
<tr>
<td>White manager</td>
<td>57.5</td>
<td>25.2</td>
<td>17.3</td>
</tr>
<tr>
<td></td>
<td>[57.4, 57.6]</td>
<td>[25.1, 25.3]</td>
<td>[17.2, 17.4]</td>
</tr>
<tr>
<td>Black manager</td>
<td>54.5</td>
<td>28.0</td>
<td>17.5</td>
</tr>
<tr>
<td></td>
<td>[54.1, 54.9]</td>
<td>[27.7, 28.4]</td>
<td>[17.2, 17.8]</td>
</tr>
<tr>
<td>Hispanic manager</td>
<td>55.6</td>
<td>24.5</td>
<td>19.8</td>
</tr>
<tr>
<td></td>
<td>[55.2, 56.0]</td>
<td>[24.2, 24.9]</td>
<td>[19.5, 20.1]</td>
</tr>
</tbody>
</table>

**Notes:** This table shows the hiring regression (9) for 812,399 newly hired WBH workers hired by WBH managers, holding constant location, department, and month. Below each coefficient is the 95% confidence interval.

### Table C3: Turnover Regression for All Workers

<table>
<thead>
<tr>
<th>6 month turnover</th>
<th>Point estimate</th>
<th>95% confidence interval</th>
<th>Number of workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>White-White ($\hat{\beta}_{WW}^x$)</td>
<td>reference</td>
<td></td>
<td>331,665</td>
</tr>
<tr>
<td>White-Black ($\hat{\beta}_{WB}^x$)</td>
<td>0.0548</td>
<td>[ 0.0515, 0.0580]</td>
<td>119,984</td>
</tr>
<tr>
<td>White-Hispanic ($\hat{\beta}_{WH}^x$)</td>
<td>-0.0045</td>
<td>[-0.0086, -0.0004]</td>
<td>66,017</td>
</tr>
<tr>
<td>Black-White ($\hat{\beta}_{BW}^x$)</td>
<td>0.0008</td>
<td>[-0.0058, 0.0074]</td>
<td>22,250</td>
</tr>
<tr>
<td>Black-Black ($\hat{\beta}_{BB}^x$)</td>
<td>0.0174</td>
<td>[ 0.0116, 0.0231]</td>
<td>29,996</td>
</tr>
<tr>
<td>Black-Hispanic ($\hat{\beta}_{BH}^x$)</td>
<td>-0.0378</td>
<td>[-0.0485, -0.0275]</td>
<td>8,490</td>
</tr>
<tr>
<td>Hispanic-White ($\hat{\beta}_{HW}^x$)</td>
<td>-0.0100</td>
<td>[-0.0170, -0.0029]</td>
<td>19,489</td>
</tr>
<tr>
<td>Hispanic-Black ($\hat{\beta}_{HB}^x$)</td>
<td>0.0320</td>
<td>[ 0.0233, 0.0408]</td>
<td>12,364</td>
</tr>
<tr>
<td>Hispanic-Hispanic ($\hat{\beta}_{HH}^x$)</td>
<td>-0.0874</td>
<td>[-0.0927, -0.0822]</td>
<td>37,693</td>
</tr>
<tr>
<td>Location FEs</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Department FEs</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>648,036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean turnover</td>
<td>0.489</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** This table shows the turnover regression (15) for all newly hired WBH workers hired by WBH managers.
Table C4: **Hiring Tests with Peer Race**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White manager</td>
<td>ref.</td>
<td>-2.55</td>
<td>-3.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.62)</td>
<td>(0.58)</td>
</tr>
<tr>
<td>Black</td>
<td>-2.71</td>
<td>ref.</td>
<td>-3.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.74)</td>
<td>(0.74)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-3.93</td>
<td>-2.52</td>
<td>ref.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.75)</td>
<td>(0.81)</td>
</tr>
<tr>
<td>White peers</td>
<td>5.88</td>
<td>-3.80</td>
<td>-2.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.39)</td>
<td>(2.86)</td>
</tr>
<tr>
<td>Black peers</td>
<td>-1.78</td>
<td>5.36</td>
<td>-3.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.79)</td>
<td>(3.20)</td>
</tr>
<tr>
<td>Hispanic peers</td>
<td>2.45</td>
<td>-2.95</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.90)</td>
<td>(3.28)</td>
</tr>
<tr>
<td>Location FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Department FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>48576</td>
<td>48576</td>
<td>48576</td>
</tr>
</tbody>
</table>

Notes: This table expands the hiring regression (9) to allow to peer race. That is, 

\[ 100 \times \text{Worker}_{t} = \text{const.} + \beta_{\theta}^{\text{Manager}}_{t} + \text{Peer}_{t} + \epsilon_{t}^{P}, \]

where “Peer” is the fraction of a manager’s team who are of race $\theta$ when the worker is hired. Below each coefficient is the standard error.
Table C5: **Productivity Regression with Peer Race**

<table>
<thead>
<tr>
<th>Sales performance</th>
<th>Point estimate</th>
<th>95% confidence interval</th>
<th>Number of workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>White-White ($\beta_{Ww}$)</td>
<td>reference</td>
<td></td>
<td>26,803</td>
</tr>
<tr>
<td>White-Black ($\beta_{Wb}$)</td>
<td>0.0089</td>
<td>[-0.0000, 0.0178]</td>
<td>2,284</td>
</tr>
<tr>
<td>White-Hispanic ($\beta_{Wh}$)</td>
<td>-0.0197</td>
<td>[-0.0292, -0.0101]</td>
<td>2,120</td>
</tr>
<tr>
<td>Black-White ($\beta_{Bw}$)</td>
<td>-0.0152</td>
<td>[-0.0220, -0.0085]</td>
<td>6,165</td>
</tr>
<tr>
<td>Black-Black ($\beta_{Bb}$)</td>
<td>0.0008</td>
<td>[-0.0093, 0.0110]</td>
<td>2,338</td>
</tr>
<tr>
<td>Black-Hispanic ($\beta_{Bh}$)</td>
<td>-0.0218</td>
<td>[-0.0358, -0.0078]</td>
<td>875</td>
</tr>
<tr>
<td>Hispanic-White ($\beta_{Hw}$)</td>
<td>-0.0293</td>
<td>[-0.0370, -0.0217]</td>
<td>4,548</td>
</tr>
<tr>
<td>Hispanic-Black ($\beta_{Hb}$)</td>
<td>-0.0423</td>
<td>[-0.0582, -0.0265]</td>
<td>708</td>
</tr>
<tr>
<td>Hispanic-Hispanic ($\beta_{Hh}$)</td>
<td>-0.0344</td>
<td>[-0.0452, -0.0236]</td>
<td>2,839</td>
</tr>
<tr>
<td>Share of Peers of Same Race</td>
<td>-0.0097</td>
<td>[-0.0191, -0.0003]</td>
<td></td>
</tr>
</tbody>
</table>

Location FEs: Yes  
Month FEs: Yes  
Department FEs: Yes  
Observations: 335,400  
Pairwise relations  
White-Black ($\Delta_{WB}$): 0.0072 [-0.0058, 0.0202]  
White-Hispanic ($\Delta_{WH}$): 0.0146 [0.0006, 0.0287]  
Black-Hispanic ($\Delta_{BH}$): 0.0306 [0.0074, 0.0538]  

Notes: This table recreates Table 5 and adds the share of peers who share the same race that month (taking into account non-WBH races). These parameters are estimated via WLS, inversely weighted to tenure. Standard errors are clustered at the worker level.
Table C6: Hiring Test for Internal Transfers

<table>
<thead>
<tr>
<th></th>
<th>White worker</th>
<th>Black worker</th>
<th>Hispanic worker</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>65.5</td>
<td>15.1</td>
<td>19.4</td>
</tr>
<tr>
<td>manager</td>
<td>[64.9, 66.1]</td>
<td>[14.6, 15.5]</td>
<td>[18.9, 19.9]</td>
</tr>
<tr>
<td>Black</td>
<td>61.5</td>
<td>19.6</td>
<td>19.0</td>
</tr>
<tr>
<td>manager</td>
<td>[59.5, 63.4]</td>
<td>[18.1, 21.1]</td>
<td>[17.3, 20.6]</td>
</tr>
<tr>
<td>Hispanic</td>
<td>61.1</td>
<td>15.4</td>
<td>23.5</td>
</tr>
<tr>
<td>manager</td>
<td>[59.4, 62.7]</td>
<td>[14.1, 16.7]</td>
<td>[22.1, 24.9]</td>
</tr>
</tbody>
</table>

Notes: This table shows the hiring regression (9) for 16,878 internally hired WBH salespeople hired by WBH managers, holding constant location, department, and month. Below each coefficient is the 95% confidence interval.

Table C7: Productivity Regression for Internal Transfers Working Under Non-Hiring Managers

<table>
<thead>
<tr>
<th>Sales performance</th>
<th>Point estimate</th>
<th>95% confidence interval</th>
<th>Number of workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>White-White</td>
<td>reference</td>
<td></td>
<td>10,707</td>
</tr>
<tr>
<td>White-Black</td>
<td>0.0265</td>
<td>[-0.0037, 0.0567]</td>
<td>751</td>
</tr>
<tr>
<td>White-Hispanic</td>
<td>-0.0028</td>
<td>[-0.0304, 0.0248]</td>
<td>921</td>
</tr>
<tr>
<td>Black-White</td>
<td>-0.0102</td>
<td>[-0.0315, 0.0111]</td>
<td>2,097</td>
</tr>
<tr>
<td>Black-Hispanic</td>
<td>0.0034</td>
<td>[-0.0344, 0.0412]</td>
<td>727</td>
</tr>
<tr>
<td>Black-Black</td>
<td>0.0311</td>
<td>[-0.0152, 0.0774]</td>
<td>322</td>
</tr>
<tr>
<td>Hispanic-White</td>
<td>-0.0137</td>
<td>[-0.0332, 0.0058]</td>
<td>2,098</td>
</tr>
<tr>
<td>Hispanic-Black</td>
<td>-0.0276</td>
<td>[-0.0770, 0.0218]</td>
<td>285</td>
</tr>
<tr>
<td>Hispanic-Hispanic</td>
<td>-0.0032</td>
<td>[-0.0326, 0.0262]</td>
<td>1,356</td>
</tr>
<tr>
<td>Location FEs</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Department FEs</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>207,532</td>
<td></td>
<td>19,935</td>
</tr>
</tbody>
</table>

Notes: This table shows the productivity regression (11) but restricts the sample to 19,935 workers who are not working under the manager who initially hired them. These parameters are estimated via WLS, inversely weighted to tenure as in Table 5. The supermodularity coefficient $\Delta_{WB}$ is defined by (12). Confidence intervals are clustered at the worker level. The number of workers is larger than in Table C6 since the hiring regression only include workers who are hired into a new team, whereas this also included managerial switches.
### Table C8: Age and Sex Differences by Racial Pair

<table>
<thead>
<tr>
<th>Age difference in years</th>
<th>White worker</th>
<th>Black worker</th>
<th>Hispanic worker</th>
<th>White worker</th>
<th>Black worker</th>
<th>Hispanic worker</th>
</tr>
</thead>
<tbody>
<tr>
<td>White manager</td>
<td>17.20</td>
<td>17.53</td>
<td>18.41</td>
<td>0.383</td>
<td>0.381</td>
<td>0.386</td>
</tr>
<tr>
<td></td>
<td>(11.30)</td>
<td>(11.50)</td>
<td>(11.51)</td>
<td>(0.486)</td>
<td>(0.486)</td>
<td>(0.487)</td>
</tr>
<tr>
<td>Black manager</td>
<td>16.67</td>
<td>17.04</td>
<td>17.32</td>
<td>0.382</td>
<td>0.411</td>
<td>0.392</td>
</tr>
<tr>
<td></td>
<td>(10.78)</td>
<td>(11.17)</td>
<td>(10.16)</td>
<td>(0.486)</td>
<td>(0.492)</td>
<td>(0.486)</td>
</tr>
<tr>
<td>Hispanic manager</td>
<td>14.82</td>
<td>13.75</td>
<td>14.66</td>
<td>0.386</td>
<td>0.411</td>
<td>0.426</td>
</tr>
<tr>
<td></td>
<td>(10.24)</td>
<td>(9.44)</td>
<td>(9.14)</td>
<td>(0.487)</td>
<td>(0.492)</td>
<td>(0.495)</td>
</tr>
<tr>
<td>Cross-race mean</td>
<td>17.12</td>
<td></td>
<td></td>
<td>Cross-race mean</td>
<td>0.385</td>
<td></td>
</tr>
<tr>
<td>Same-race mean</td>
<td>16.96</td>
<td></td>
<td></td>
<td>Same-race mean</td>
<td>0.389</td>
<td></td>
</tr>
<tr>
<td>p-value, same means</td>
<td>0.160</td>
<td></td>
<td></td>
<td>p-value, same means</td>
<td>0.454</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports differences in ages and gender by racial pair, for salespeople in the main sample, given the race of the worker and manager. The **left panel** reports the means and standard deviations for the absolute value of difference in ages in years. For instance, the top left cell restricts the sample to white workers hired by white managers, then reports the mean absolute value of their age differences. The “cross-race mean” reports the average of the “cross-race” cells, weighted by sample size, while the “same-race” mean reports the average of the same-race cells. The p-values correspond to two-sample t-test for equal means, where the dependent variable (the absolute value of the difference in ages) depends on whether the manager is same-race or cross-race. The **right panel** reports the mean and standard deviation of an indicator that equals one if the manager and worker have different genders. A majority of employees are men, so the probability that sexes differ is less than 0.5. Data in this table include 47,094 newly hired sales workers included in the main sample, and exclude approximately 3.5% of workers for whom manager gender or age are missing.