STATISTICAL DISCRIMINATION AND THE DISTRIBUTION OF WAGES

Prashant Bharadwaj [Q]
Rahul Deb [◊]
Ludovic Renou [‡]

March 25, 2024

ABSTRACT

We characterize wage distributions that are consistent with a general model of statistical discrimination. We adapt this theoretical characterization to develop an empirical test, the rejection of which we interpret as evidence of taste-based discrimination. We demonstrate how this test can be applied on cross-sectional and panel data. Results from Census and NLSY data suggest taste-based discrimination at work against Blacks in the labor force.

[Q] Department of Economics, University of California, San Diego prbharadwaj@ucsd.edu
[◊] Department of Economics, University of Toronto, rahul.deb@utoronto.ca
[‡] School of Economics and Finance, Queen Mary University of London, lrenou.econ@gmail.com
This paper builds on a previous version titled "Which wage distributions are consistent with statistical discrimination?" by the second and third authors. We are very grateful to Paula Onuchic for engaging in a conversation with us that inspired this paper and to her, Aislinn Bohren, Michael Grubb, Peter Hull, Alex Imas, Rob McMillan and Konstantinos Tatsiramos for detailed discussions. We would also like to thank Juan Eberhard, Namrata Kala, Kory Kroft, Aditya Kuvalekar, Marco Manacorda, Mallesh Pai, Debraj Ray, Maher Said and numerous seminar participants for insightful comments. This paper has greatly improved thanks to the comments of the co-editor and the referees. We would like to thank the gracious hospitality of the Studienzentrum Gerzensee where part of this research was conducted. Deb thanks the SSHRC for their continued and generous financial support without which this research would not be possible.

1. INTRODUCTION

While it is well established that economic outcomes for observationally identical individuals can differ based on their group identity, it is significantly harder to determine the reason for these disparate outcomes. Discrimination is one possible explanation. At a broad level, economists characterize discrimination as either *statistical* (outcomes differ because of differences in information) or *taste-based* (bias or animus towards one group drives outcome differences). While both forms of discrimination are problematic, taste-based discrimination is particularly pernicious because, unlike statistical discrimination, it is unresponsive to information. Establishing the form of discrimination is important both for accountability and to devise corrective policies.

Consider, for instance, the differences in the wage distributions for Black and white workers of identical age and education who work the same jobs. Suppose that, despite being observationally identical, whites have higher average wages. Is such a wage gap the result of taste-based discrimination? Not necessarily. Even if whites and Blacks have the same average productivities, their productivity distributions and/or the information that employers receive about these workers need not be identical. Thus, if wages were a nonlinear function of the worker's productivity (as perceived by employers), wage gaps can arise even without any taste-based discrimination.

In this paper, we propose a general model of statistical discrimination in the labor market and theoretically characterize the set of wage distributions that are consistent with this model. This characterization in turn yields a nonparametric test for statistical discrimination, rejections of which we interpret as evidence of taste-based discrimination. A strength of this approach is that it can be applied to commonly available cross-sectional data (such as Census data) and it provides a framework for interpreting when the "unexplained" part of a wage decomposition can be interpreted as arising from statistical discrimination alone. Our empirical results provide evidence of taste-based discrimination against Blacks in the labor force.

Our model is in the spirit of Phelps (1972). There are two groups whose productivity distributions differ. The group identity is observable to employers, but productivities are not. Instead, employers learn about the workers' productivity from signals whose distributions may vary across the groups. For example, these signals could be the information that employers receive from the job screening process that includes interviews, tests, curricula vitae etc. Signal realizations induce posterior productivity distributions (via Bayes' rule) and, in particular, these can be used to compute posterior estimates (the mean of the productivity conditional on the signal realization) of the unobserved productivity. Therefore, each group's signal generates a distribution over posterior productivity estimates. Wages are then determined via a strictly increasing, continuous function of the posterior productivity estimate that, importantly, does not depend on the group. The combination of assumptions make our model more general than

others in the literature: we do not require the productivity distributions or statistical experiments to be Gaussian and we allow for nonlinear wage functions to capture imperfectly competitive labor markets. Our theory aims to characterize the pairs of wage distributions that are rationalizable by this model under different assumptions about the set of permissible productivity distributions.

We first consider the baseline case where we assume that both groups have equal mean productivities but, apart from this, the distributions can differ arbitrarily. We show that a necessary and sufficient condition for a pair of wage distributions to be rationalizable under this assumption of equal mean productivities is that neither wage distribution first-order stochastically dominates the other. Thus, if wages are ordered by first-order stochastic dominance, then they cannot be explained by *statistical discrimination alone*. We then argue that an immediate consequence of this result is a characterization of rationalizable wage distributions assuming that one group has a weakly higher mean productivity: the wage distribution for the group with lower mean productivity should not be first-order stochastically dominant. If we observe the latter in the data, we interpret this as evidence of taste-based discrimination.

First-order stochastic dominance is a simple condition that easy to visualize and, we argue, to test. We first discuss how out test can be applied on cross-sectional data. We consider Black and white, prime age males in a given occupation. The distribution over covariates for both groups differ. We use the approach of Chernozhukov, Fernández-Val, and Melly (2013) to test whether the wage distribution of Blacks is first-order stochastically dominated by the counterfactual wage distribution generated by assuming that the wages for Blacks are determined by the same process as those of whites (in language familiar to labor economists, this is the "wage structure" effect). In words, this compares the observed wage distribution for Blacks with what they would receive if they were treated as whites. The assumption mapping this approach to our theory is that, conditioning on observables, the mean productivity of Blacks is not lower than that of whites. There is reason to think this assumption is reasonable. As Bohren, Hull, and Imas (2022) argue, members of disadvantaged groups face barriers or (in their language) systemic discrimination in achieving certain characteristics. Thus, we might expect that, facing such systemic discrimination, Black workers who achieve a given level of education have higher average productivity than their white counterparts.[1] As a proof of concept, we apply our method to Census data. The empirical results provide evidence of taste-based discrimination: wages of Blacks are consistently first-order stochastically dominated by whites across occupations.

While we stress that our main contribution is theoretical, we show that our suggestive evidence is compelling by conducting two additional analyses as robustness checks. First, we compare the wage distributions of white workers with high school degrees in a given occupation against Blacks with two years of college. The assumption of higher average productivity for Blacks is only violated if whites are more

[1] Ashraf, Bandiera, Minni, and Quintas-Martınez (2022) demonstrate that women who work at a multinational firm have higher productivities relative to comparable men and the gap widens as it becomes harder for women to enter the labor force.

productive on average despite having fewer years of schooling (which we do not believe to be true). Second, we implement our test on panel data from the NLSY which allows us to condition on past wages in addition to other covariates. Then, the assumption that Blacks have higher average productivities is only violated if, in the previous period, Blacks were paid more than whites given the same expected productivity (which we view to be implausible). In both robustness checks, we show that the wages of Blacks remain first-order stochastically dominated by those of whites.

The penultimate section of the paper extends the theory in several directions. Notably, we examine the case where we do not assume average productivities are ordered. We derive a tight lower bound on the average productivity differences required to rationalize a given pair of wage distributions. We discuss how this bound can be used to uncover evidence for taste-based discrimination. We then argue that deriving a similar bound for percentage differences in average productivities is not possible unless we make further assumptions about the relationship between wages and productivities.

RELATION TO THE LITERATURE

There are large insightful literatures in economics, psychology and sociology studying discrimination and we will not attempt to provide a comprehensive description here. Instead, we refer the reader to several excellent recent surveys in economics—Fang and Moro (2011), Lang and Lehmann (2012), Bertrand and Duflo (2017), Lang and Spitzer (2020), Onuchic (2022)—that cover both the theory and the empirical evidence in a variety of different settings. But with that said, we would like to situate our paper within the broader literature.

In recent years, considerable progress have been made towards credibly establishing the *presence* of discrimination. Field experiments (of both the audit and correspondence variety) have been particularly instrumental in evidencing discrimination, as they allow the researcher to finely control the observables. However, as the aforementioned Bertrand and Duflo (2017) observe: "while field experiments have been overall successful at documenting that discrimination exists, they have (with a few exceptions) struggled with linking the patterns of discrimination to a specific theory."

The classic way to test for taste-based discrimination is using what are known as "outcome tests" (in the spirit of Becker, 1957, 1993). These tests require the researcher to have access to not just to the decision (whether or not a loan is granted, a driver is searched by a police officer, etc) but also the post-decision result (whether or not the loan is repaid, contraband is found on the driver, etc). The key insight is that even though the rates at which decisions are made may differ due to group differences, the post-decision results of the *marginal* case should be the same if the decision maker is unbiased. This requires devising empirical strategies to identify the post-decision results of marginal cases or models that provide

a systematic relationship between the average and marginal post-decision result.[2]

Such post-decision outcome data is clearly not available in correspondence studies that typically send fictitious curricula vitae to employers and so interview call back rates are the only information that researchers have to work with.[3] In our context of wage discrimination, a version of an outcome test would require the researcher to observe some information on worker productivity in addition to the wages. Reliable data on productivity is very rarely available. Thus, one way to interpret our contribution is that our theory allows us to devise a test for taste-based discrimination in an important context (the labor market) using easily available cross-sectional data. Moreover, our approach allows us to sidestep many of the difficulties associated with identifying the marginal case in outcome tests.

In our view, an important contribution of our paper is that it provides a theoretical lens to interpret the decompositions of wages. Wage decompositions in labor economics have a rich history, starting with the seminal work of Kitagawa (1955), Oaxaca (1973) and Blinder (1973) who developed the framework to understand whether differences in outcomes were the result of differing characteristics or differential returns to characteristics across groups. The "unexplained portion" of the Kitagawa–Oaxaca–Blinder (henceforth KOB) decomposition has long been a North Star for labor economists aiming to quantify the *amount* of discrimination; early influential papers include Juhn et al. (1993) and Altonji and Blank (1999).[4] This strand of the literature has largely evolved in parallel to the work that aims to determine the *type* of discrimination. Our novel theory, combined with the recent empirical advances made possible by DiNardo, Fortin, and Lemieux (1996) and Chernozhukov, Fernández-Val, and Melly (2013) help these literatures speak to each other by providing a way of interpreting the unexplained portion of the wage decomposition through the lens of the two dominant models of discrimination in labor economics.

2. THE MODEL

This section presents our model of statistical discrimination – a non-parametric generalization of the model of Phelps (1972). Section 6 discusses some extensions.

There are two groups—1 and 2—of workers; examples include female and male, Black and White, junior and senior, or disabled and able bodied. We do not take a stand on which of these two groups is advantaged/disadvantaged, if any. We observe two wage distributions G_1 and G_2, with $G_i(w) \in [0, 1]$ being the fraction of workers in group $i \in \{1, 2\}$ who are paid a wage of $w \geq 0$ or less.[5] We assume that the

[2]See, for instance, Knowles, Persico, and Todd (2001), Anwar and Fang (2006), Arnold, Dobbie, and Yang (2018) and Canay, Mogstad, and Mountjoy (2020).

[3]As we discuss below, the binary decision (interview call back or not) in correspondence studies (as opposed to wages) further complicates testing for both the presence of discrimination and the form it takes.

[4]The aforementioned recent work of Bohren et al. (2022) provides more nuance on how to interpret the unexplained portion and introduces the ideas of "direct" and "systemic" discrimination (which we referred to earlier).

[5]Throughout, all distributions are right-continuous and have limits on the left.

wage distributions are bounded, that is, $G_i(\overline{w}) = 1$ for some $\overline{w} > 0$, $i = 1, 2$.

The question we address is: under what conditions are the observed wage distributions rationalizable by (or consistent with) a general model of statistical discrimination? The purpose is to develop a test for taste-based discrimination: when the wage distributions *cannot* be rationalized, statistical discrimination alone cannot explain the data. In other words, our null hypothesis is that the data is consistent with statistical discrimination alone (which includes no discrimination) and we are interested in rejecting the null hypothesis. We interpret a rejection of the null hypothesis as evidence of taste-based discrimination – we argue the validity of this interpretation later. We now present the model in detail, starting with the productivity distributions.

Productivity distributions: Workers differ in their productivities, with $\theta_i \in [0, \overline{\theta}] =: \Theta$ (for some fixed $\overline{\theta} \geq \overline{w}$) denoting the productivity of a worker in group $i \in \{1, 2\}$, and H_i its (cumulative) distribution. The productivity refers to the marginal product of a worker and is thus measured in the same unit as wages (US dollars in our empirical application).

We hypothesize a set $\widehat{\mathcal{H}}$ of pairs of productivity distributions (H_1, H_2) and derive the testable implications of our model assuming that the productivity distributions $(H_1, H_2) \in \widehat{\mathcal{H}}$. Thus, as we change $\widehat{\mathcal{H}}$, the testable implications may change. As an example, in our baseline model below, we assume that $\widehat{\mathcal{H}}$ is the set of all pairs of productivity distributions that have equal means. As another example, we assume that $\widehat{\mathcal{H}}$ contains all pairs of distributions, whose means differ by at most $d > 0$. The validity of the hypothesized set $\widehat{\mathcal{H}}$ must be argued, either empirically or theoretically.

Information: Employers do not directly observe the productivity of workers, but receive informative signals (from curricula vitae, reference letters, interviews, tests, etc.). Employers then form an expectation of the productivity of workers and pay them accordingly: wages are strictly increasing in expected productivity. Since wages only depend on the expected productivity, it is without loss to restrict attention to signals of the following form.

A *signal* (S_i, π_i) for group $i \in \{1, 2\}$ consists of a set of *signal realizations* $S_i = \Theta$ and a joint distribution π_i over $\Theta \times S_i$, whose marginal distribution over Θ is the productivity distribution H_i. We denote the marginal distribution of π_i over S_i by F_i. We assume that the *posterior estimate* $\mathbb{E}_{\pi_i}[\theta_i \mid s_i]$ of the productivity satisfies

$$s_i = \mathbb{E}_{\pi_i}[\theta_i \mid s_i],$$

for all s_i in the support of F_i. In other words, the signal realization s_i is an accurate estimate of the true productivity θ_i. This is without loss of generality, as we can always relabel signals to guarantee that they are accurate in the above sense. Accordingly, we will write θ_i for the posterior estimate (the signal realization) in what follows.

It is well known that F_i is a distribution of posterior estimates arising from *some* signal if, and only if, the prior distribution H_i is a *mean-preserving spread* of the posterior distribution F_i, which we denote by $F_i \succcurlyeq_2 H_i$ (where the notation reflects second-order stochastic dominance). Formally, the mean-preserving spread condition requires that

$$\int_0^{\theta} H_i(\theta_i)d\theta_i \geq \int_0^{\theta} F_i(\theta_i)d\theta_i \text{ for all } \theta \in [0, \overline{\theta}], \text{ with equality at } \theta = \overline{\theta}.$$

Note that the requirement of equality at $\theta_i = \overline{\theta}$ is the same as ensuring that H_i and F_i have the same mean.[6]

We stress that the above formulation subsumes *all* possible signaling technologies. In particular, this includes the common formulation (as in Phelps, 1972; Aigner and Cain, 1977) of modeling (biased) signals as $s_i = \theta_i + \varepsilon_i$, where ε_i is a noise term whose distribution can depend on the group i and possibly the productivity θ_i as well.

Wage function: If an employer estimates the productivity of a worker to be θ, they pay the worker $W(\theta)$, where the *wage function* $W : \Theta \to \mathbb{R}_+$ is continuous and strictly increasing.[7] Let $\mathcal{W}$ be the set of continuous and strictly increasing functions with domain Θ and range $\mathbb{R}_+$. As with the set of productivity distributions, we hypothesize a set of wage functions $\widehat{\mathcal{W}}$. As an example, employers may pay workers a fixed fraction of their expected productivities, in which case $\widehat{\mathcal{W}}$ is the set of linear functions.

It is worth making a few comments about the wage functions. First, observe that this wage function does not depend on the group identity and, in this sense, our model captures statistical, but not taste-based, discrimination. Second, the labor economics literature frequently assumes perfectly competitive labor markets. In our notation, this amounts to assuming $\widehat{\mathcal{W}} = \{W_{id}\}$ where $W_{id}(\theta) = \theta$ for all $\theta \in \Theta$. As will become clear from our results below, a strength of our framework is that we require no such restrictive assumptions and we can accommodates imperfectly competitive labor markets (via the reduced form wage function) quite generally.

Induced wage distributions: The distribution F_i over posterior estimates *induces the wage distribution* G_i via the wage function W. Formally, for both $i \in \{1, 2\}$, $G_i(w)$ is the measure of the set $\{\theta : W(\theta) \leq w\}$ according to F_i, that is, $G_i(w) = F_i(W^{-1}(w))$ for $w \in [W(0), W(\overline{\theta})]$, $G_i(w) = 0$ for $w < W(0)$ and $G_i(w) = 1$ for $w > W(\overline{\theta})$.[8] Note that, even though the wage function does not depend on group identity, the wage distributions G_1 and G_2 may differ across groups for the simple reason that the

[6] Integration by parts implies that the mean satisfies $\int_0^{\overline{\theta}} \theta_i dF_i(\theta_i) = \theta_i F_i(\theta_i)|_0^{\overline{\theta}} - \int_0^{\overline{\theta}} F_i(\theta_i)d\theta_i = \overline{\theta} - \int_0^{\overline{\theta}} F_i(\theta_i)d\theta_i$.
[7] We could additionally assume that $W(\theta) \leq \theta$ (workers are paid less than their marginal product) and no result in the paper changes. While this is a natural assumption, we do not require it so choose not to impose it.
[8] We define W^{-1} as the inverse of W on the domain $[W(0), W(\overline{\theta})]$.

distributions of posterior estimates F_1 and F_2 may differ.

Rationalizability: We say that the observed wage distributions G_1 and G_2 are *rationalizable* (given $\widehat{\mathcal{H}}$, $\widehat{\mathcal{W}}$) if there exist (i) productivity distributions $(H_1, H_2) \in \widehat{\mathcal{H}}$, (ii) distributions of posterior estimates F_i that satisfy $F_i \succcurlyeq_2 H_i$ for $i \in \{1, 2\}$, and (iii) a wage function $W \in \widehat{\mathcal{W}}$, such that these jointly induce the observed wage distributions.

Before moving on to the analysis, we would like to discuss our model and the main question we address. The model is in the spirit of the seminal models of Phelps (1972) and Aigner and Cain (1977). Phelps considers two populations whose productivities are drawn from a Gaussian distribution. Signals are also normally distributed, differ across groups, and the wage function is linear in the posterior estimate. If the means of the productivity distributions for both groups are the same, then the Phelps' model implies that the average wage for both groups is the same (because the posterior distribution must have the same mean as the prior, and the wage function is linear). In this case, there is no discrimination at the group level even though the wage distributions differ (so there is individual level discrimination). Aigner and Cain (1977) observe it is possible to generate discrimination at the group level via more general wage functions even when the productivity distributions for both groups are identical. In their model, wages depend both on the mean and the variance of the posterior belief. In the normal learning environment, the posterior variance is the same for all signal realizations so they model the wage as just the difference between the posterior mean and some multiple of the (signal independent) variance of the posterior belief. Hence, different normally distributed signals can generate distinct mean wages.

Our model is more general than these seminal papers (and most of the literature) in that we do not assume that the productivity distributions are Gaussian and precisely known by the analyst (they instead lie in set $\widehat{\mathcal{H}}$), and we allow for unrestricted (not necessarily Gaussian) signals. Moreover, we do not restrict wages to be linear in the posterior estimate. That said, unlike Aigner and Cain (1977), wages in our model only depend on the mean of the posterior distribution but not on the variance. This choice is deliberate: our model is very general and our assumptions balance this generality against meaningful testable implications that can be taken to the data. As we discuss in our concluding remarks, allowing wages to depend on both the mean and the variance makes the testable implications of our model vacuous even when relatively restrictive assumptions are imposed on $\widehat{\mathcal{H}}$ (equal mean productivities) and $\widehat{\mathcal{W}}$ (wages that are affine in the posterior mean and variance).

Similar to Phelps (1972) and Aigner and Cain (1977), we do not model the underlying reason that productivities differ across groups. In this sense, we differ from papers like Coate and Loury (1993) whose primary purpose is to explain how stereotypes (that assume the disadvantaged group has lower mean productivity) can be self-fulfilling. Note that we can accommodate self-fulfilling stereotypes against group 1 in our model by hypothesizing the set $\widehat{\mathcal{H}} = \{(H_1, H_2) \,|\, \mathbb{E}_{H_1}[\theta_1] \leq \mathbb{E}_{H_2}[\theta_2]\}$.

Our main theoretical aim is to precisely characterize the set of rationalizable pairs of wage distributions under different assumptions on $\widehat{\mathcal{H}}$ and $\widehat{\mathcal{W}}$. This separates our analysis from most theoretical papers on discrimination that aim to derive the main implications of their models, but do not provide a complete characterization of the testable implications. In this sense, our analysis is closer to the theoretical literature on decision theory and revealed preference.

3. CHARACTERIZING RATIONALIZABLE WAGE DISTRIBUTIONS

In this section, we derive the main theoretical results that we take to the data. We begin with a baseline result that assumes both groups have equal mean productivities. We then generalize to a setting where we make the weaker assumption that one group has weakly higher mean productivity than the other – this generalization is the basis of our empirical application.

3.1. BASELINE RESULT: EQUAL MEAN PRODUCTIVITIES

As we have emphasized earlier, the key novelty of our framework is that we allow for (arbitrary) non-Gaussian priors, signals and nonlinear wages. In this subsection, we isolate the implications of these modeling features by assuming that both groups have equal mean productivities but, apart from this, we impose no further restrictions. In other words, we characterize how the wage distributions can differ due to a combination of statistical discrimination and nonlinear wages for two groups that are identical on average. We weaken this assumption in subsequent sections.

We define

$$\mathcal{H}_{=} := \{(H_1, H_2) \,|\, \mathbb{E}_{H_1}[\theta_1] = \mathbb{E}_{H_2}[\theta_2]\}$$

to be the set of productivity distribution pairs that have equal means. Note that, barring the equal means requirement, the productivity distributions can differ arbitrarily.[9] Formally, we characterize the set of rationalizable wage distributions given that the productivity distributions lie in the set $\mathcal{H}_{=}$ but the wage function is unrestricted (that is, we take $\widehat{\mathcal{W}} = \mathcal{W}$).

Despite the equal means assumption, the model is very general as the signals and the wage function are unrestricted. Given this generality, the first natural question to ask is: are there *any* wage distributions that are *not* rationalizable (given $\mathcal{H}_{=}$, $\mathcal{W}$)? To this point, note that our model allows the posterior estimate distribution of group 1 to be a strict mean-preserving spread of group 2,[10] in which case a strictly convex wage function W will generate higher mean wages for group 1. In other words, differences in mean wages (a wage gap) can arise purely via statistical discrimination, even though both groups have equal mean productivities. So, to find inconsistent distributions, we need to consider higher moments

[9] Distributions can be discrete, continuous or a mixture of the two.
[10] That is, $F_2 \succcurlyeq_2 F_1$ and $F_2 \neq F_1$.

of the wage distribution. In fact, as we now argue, we need to consider *all* moments via the following order.

The wage distribution G_i *strictly first-order stochastically dominates* the wage distribution G_j, which we denote $G_i \succ_1 G_j$, if $G_i(w) \leq G_j(w)$ for all $w \in \mathbb{R}_+$, with the inequality strict for some w.

Now suppose that the wage distribution of group i strictly first-order stochastically dominates that of group j. We now argue that these distributions are *not* rationalizable. For contradiction, assume that these distributions are rationalizable (given $\mathcal{H}_=$, $\mathcal{W}$). This implies that there exist posterior estimate distributions F_i and F_j, and a wage function W, such that

$$F_i(\theta) = G_i(W(\theta)) \leq G_j(W(\theta)) = F_j(\theta) \qquad \text{for all } \theta \in [0, \overline{\theta}],$$

with the inequality strict for some θ, that is, $F_i \succ_1 F_j$. It follows that F_i has a strictly higher mean than F_j, which is a contradiction since F_i and F_j are *mean-preserving* contractions of some productivity distributions $(H_i, H_j) \in \mathcal{H}_=$, which both have the same mean.

The above argument shows that a necessary condition for a pair of wage distributions to be rationalizable (given $\mathcal{H}_=$, $\mathcal{W}$) is that neither strictly first-order stochastically dominates the other. Our first result shows that this condition is also sufficient.

THEOREM 1. *Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_=$, $\mathcal{W}$) if, and only if, neither G_1 nor G_2 strictly first-order stochastically dominates the other.*

We now offer an in-depth discussion of Theorem 1. To start with, we note that that our model has in-build redundant generality. In what follows, we explain what they are and then discuss their economic implications. Denote

$$\widetilde{\mathcal{H}}_{\equiv} := \{(H_1, H_2) \in \mathcal{H}_= \mid (H_1, H_2) \text{ are supported on } \{0, \overline{\theta}\}\}$$

to be the subset of $\mathcal{H}_=$ containing the pairs of distributions supported on the points 0 and $\overline{\theta}$. (Recall that the supports of $(H_1, H_2) \in \mathcal{H}_=$ are included in $[0, \overline{\theta}]$.) Note that $(H_1, H_2) \in \widetilde{\mathcal{H}}_{\equiv}$ implies that $H_1 = H_2$ since there is only a single distribution with binary support $\{0, \overline{\theta}\}$ that has a given mean.

Take any distribution H_i and let $\widetilde{H}_i$ be the (discrete) distribution supported on $\{0, \overline{\theta}\}$ such that both distributions have equal means $\mathbb{E}_{H_i}[\theta_i] = \mathbb{E}_{\tilde{H}_i}[\theta_i]$. Now, observe that every mean-preserving contraction F_i of H_i, i.e., $F_i \succcurlyeq_2 H_i$, is also a mean-preserving contraction of $\widetilde{H}_i$, i.e., $F_i \succcurlyeq_2 \widetilde{H}_i$. This is because, $\widetilde{H}_i$ is the distribution that is the "most spread" (in that all the mass is at both end points of the interval 0 and

$\overline{\theta}$) amongst all distributions with mean $\mathbb{E}_{H_i}[\theta_i]$ that are supported on a subset of $[0, \overline{\theta}]$. Consequently,

$$\{(F_1, F_2) \mid F_1 \succcurlyeq_2 H_1,\ F_2 \succcurlyeq_2 H_2 \text{ and } (H_1, H_2) \in \widetilde{\mathcal{H}}_{\equiv}\} = \{(F_1, F_2) \mid F_1 \succcurlyeq_2 H_1,\ F_2 \succcurlyeq_2 H_2 \text{ and } (H_1, H_2) \in \mathcal{H}_{=}\}.$$

If we denote $\mathcal{H}_{\equiv} := \{(H_1, H_2) \mid H_1 = H_2\}$ the set of all pairs of identical productivity distributions, the above equality implies that

$$\{(F_1, F_2) \mid F_1 \succcurlyeq_2 H_1,\ F_2 \succcurlyeq_2 H_2 \text{ and } (H_1, H_2) \in \mathcal{H}_{\equiv}\} = \{(F_1, F_2) \mid F_1 \succcurlyeq_2 H_1,\ F_2 \succcurlyeq_2 H_2 \text{ and } (H_1, H_2) \in \mathcal{H}_{=}\},$$

because $\widetilde{\mathcal{H}}_{\equiv} \subset \mathcal{H}_{\equiv} \subset \mathcal{H}_{=}$. This shows is that the set of rationalizable wage distributions given $\mathcal{H}_{=}$, $\mathcal{W}$ is the same as the set of rationalizable wage distributions given $\mathcal{H}_{\equiv}$, $\mathcal{W}$. In words, the testable implications of our model are the same whether we assume equal mean productivities or *identical productivity distributions*.

THEOREM 1 (CONTINUED). *The following statements are equivalent.*

(i) Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_{=}$, $\mathcal{W}$).

(ii) Neither G_1 nor G_2 strictly first-order stochastically dominates the other.

(iii) Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_{\equiv}$, $\mathcal{W}$).

It has been argued that, for the distributions of certain traits, men and women have the same mean, but the former have a higher variance. This is sometimes referred to as the "variability hypothesis." The third statement of Theorem 1 implies that any two wage distributions that are not ordered by strict first-order stochastic dominance, no matter how different, could have resulted from statistical discrimination on *identical* populations. In other words, allowing for different variances of the productivity distributions leads to no additional explanatory power of statistical discrimination.

The redundant generality in our model can be rephrased in an additional way. Observe that, given any pair of productivity distributions $(H_1, H_2) \in \mathcal{H}_{=}$, every pair of mean-preserving contractions, $F_1 \succcurlyeq_2 H_1$ and $F_2 \succcurlyeq_2 H_2$ also belong to the set $\mathcal{H}_{=}$, that is $(F_1, F_2) \in \mathcal{H}_{=}$. This is simply because F_1 and F_2 have the same means and $\mathcal{H}_{=}$ contains every pair of distributions whose means are equal. Consequently, the set $\mathcal{H}_{=}$ equals the set

$$\{(F_1, F_2) \mid F_1 \succcurlyeq_2 H_1,\ F_2 \succcurlyeq_2 H_2 \text{ and } (H_1, H_2) \in \mathcal{H}_{=}\}.$$

This equality of sets has an important economic implication. It says that we cannot distinguish the model where employers have the prior beliefs (F_1, F_2) and perfectly observe the productivity of workers from

the model where employers have the prior beliefs (H_1, H_2) and form the posterior beliefs (F_1, F_2) via noisy signals.

We say that two wage distributions G_1 and G_2 are *rationalizable without signal discrimination* (given $\widehat{\mathcal{H}}$, $\widehat{\mathcal{W}}$) if there exist (i) productivity distributions $(H_1, H_2) \in \widehat{\mathcal{H}}$, (ii) perfectly informative signals ($F_1 = H_1$, $F_2 = H_2$) and (iii) a wage function $W \in \widehat{\mathcal{W}}$, such that these jointly induce the observed wage distributions.[11] The above argument thus implies the following result.

THEOREM 1 (CONTINUED). *The following statements are equivalent.*

(i) Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_=$, $\mathcal{W}$).

(ii) Neither G_1 nor G_2 strictly first-order stochastically dominates the other.

(iii) Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_\equiv$, $\mathcal{W}$).

(iv) Wage distributions G_1 and G_2 are rationalizable without signal discrimination (given $\mathcal{H}_=$, $\mathcal{W}$).

The fourth statement says that, when condition (ii) holds, we cannot conclude that *discrimination in either informativeness of the signals or wage payments is present*. In other words, it is possible that the differences in wages arise from statistical discrimination on identical populations or, simply, from heterogeneous populations (with identical mean productivities) facing identical and unbiased hiring processes. Naturally, we cannot exclude the possibility that the heterogenous populations are themselves the product of past discrimination, be it statistical or taste-based. The empirical application we develop partially address that issue. We employ decomposition methods to test for *direct* discrimination, that is, we compare the wage distribution of the disadvantaged group with the wage distribution we would have observed had the disadvantaged group been treated as the advantaged group – we thus fix the distribution of observable characteristics. An implication of the statement is that when condition (ii) is rejected in the data, this not only shows discrimination is present, it says that discrimination cannot be statistical alone!

In the setting of correspondence studies (where researchers send fictitious CVs to employers), a differential callback rate is typically interpreted as evidence of discrimination (a statistical versus taste-based conclusion is typically not made). But consider the following situation. Suppose that, despite having the same mean, the variance of productivities might differ for workers from both groups despite having an identical CVs. One possible reason is that universities have differential admissions policies across groups and so the same educational qualification might nonetheless imply different productivity distributions. If employers only call back for interviews, those applicants whose expected productivities are above a

[11] Note that it would be equally natural to assume that employers do not perfectly observe productivities, but learn about them via a single group-independent signal. The statement of our result applies verbatim with this definition of rationalizability without discrimination.

threshold, the callback rates might differ. In other words, differential callback rates might occur even if the employer has the same group-independent signal, that is, there is no discrimination (as we define it). A version of this critique has been made by Heckman (1998) and Neumark (2012).

It is possible to derive a version of Theorem 1 for binary outcomes (for example, an interview callback or not in a correspondence study); a formal statement is in an older version Deb and Renou (2022) of this paper. It is easy to show, that unless *every* candidate from one group and *none* from the other is offered an interview, any other pair of callback rates are rationalizable even if both groups have the same mean productivity. Of course, in practice, the callback rates for both groups are always non-zero. Thus, it is theoretically possible that the empirical results from many correspondence studies evidence no discrimination (in our sense) whatsoever! A second, and in our view important, implication of point (iv) of Theorem 1 is that studying a non-binary outcome such as wages allows for not just the conclusive evidence of the presence of discrimination but also allows us to detect whether taste-based discrimination is present.

It is worth reiterating that the aim of our theory is to characterize wage distributions that are consistent with statistical discrimination alone. This allows us to develop a test, rejections of which imply that, while statistical discrimination might be present, it alone cannot explain the data. It is easy to show that if wages are group dependent—that is, wages for group 1, 2 are determined by potentially different functions W_1, W_2—then every pair of wage distributions are rationalizable with group-dependent wage functions (given $\mathcal{H}_=$, $\mathcal{W}$). Following the definition of Becker (1957), we interpret such group dependent wages as *taste-based discrimination* because two workers with identical expected productivities in the eyes of the employer are paid differently. Importantly, this implies that if wages are ordered by strict first-order stochastic dominance (thus violating condition ii of Theorem 1), this can be interpreted as evidence of bias or animus. This interpretation would not always be correct if there existed pairs of wage distributions that were not rationalizable by group-dependent wages.

Finally, note that the above result does not require employers to have *accurate* beliefs about either the productivity distributions or about the signal. Employers may believe that the productivity distribution of the advantaged group has higher variance or that the signals for this group are more accurate. The only assumption we require is that these beliefs (whether accurate or inaccurate) satisfy the assumption that they assign the same mean productivity to both groups. Thus, a rejection of the test in Theorem 1 is robust evidence of taste-based discrimination in that it rules out statistical discrimination arising from either accurate or inaccurate information that differs across groups (this feature differentiates our setting from that Bohren, Haggag, Imas, and Pope, 2019). Of course, employers may have inaccurate beliefs that assign a lower mean productivity to the disadvantaged group.[12] We interpret this to be taste-based

[12] The beliefs are inaccurate in that there are inconsistent with our hypothesis of equal productivity means.

discrimination since we cannot distinguish such incorrect beliefs from group dependent wage functions. Since all the restrictions we impose in this paper (via the set $\widehat{\mathcal{H}}$) are only on the means of the productivity distributions, the argument in this paragraph applies to all theoretical results in the paper.

3.2. ORDERED MEAN PRODUCTIVITIES

While assuming equal mean productivities is a natural starting point for a theoretical study on discrimination, it is an extremely restrictive assumption if one wants to take the theory to the data, as we do. In this subsection, we weaken this assumption and show that Theorem 1 extends immediately to an environment where we assume mean productivities are not equal but ordered. We first present the result formally and then briefly discuss why this generalization is helpful for empirical applications.

We define

$$\mathcal{H}_{\geq} := \{(H_1, H_2) \,|\, \mathbb{E}_{H_1}[\theta_1] \geq \mathbb{E}_{H_2}[\theta_2]\}$$

to be the set of productivity distribution pairs in which the mean productivity of group 1 is weakly greater than that of group 2. Of course, $\mathcal{H}_{\geq} \supset \mathcal{H}_{=}$ so this is a weaker assumption than that imposed in Theorem 1 but the following result follows immediately.

THEOREM 2. *Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_{\geq}$, $\mathcal{W}$) if, and only if, G_2 does not strictly first-order stochastically dominate G_1.*

The discussion in this paragraph serves as a proof of this result. Suppose $G_2 \succ_1 G_1$. Then the argument preceding (the first statement of) Theorem 1 implies that, for any $W \in \mathcal{W}$, the resulting distributions $F_1(\theta_1) = G_1(W(\theta_1))$, $F_2(\theta_2) = G_2(W(\theta_2))$ of posterior estimates satisfy $\mathbb{E}_{F_1}[\theta_1] < \mathbb{E}_{F_2}[\theta_2]$. This is a contradiction since every pair (H_1, H_2) that satisfies $F_i \succcurlyeq_2 H_i$ has mean $\mathbb{E}_{F_i}[\theta_i] = \mathbb{E}_{H_i}[\theta_i]$ for $i \in \{1, 2\}$, and so $(H_1, H_2) \notin \mathcal{H}_{\geq}$. If G_1 and G_2 are not ordered by strict first-order stochastic dominance, then Theorem 1 shows that these wage distributions are rationalizable (given $\mathcal{H}_{=}$, $\mathcal{W}$) and so also rationalizable (given $\mathcal{H}_{\geq}$, $\mathcal{W}$). Lastly, if $G_1 \succ_1 G_2$, then with the (perfectly competitive) wage function $W(\theta) = \theta$, the resulting distributions of posterior estimates, and therefore true productivity distributions, are ordered as needed.

In the next section, we derive and implement an empirical test for taste-based discrimination based on Theorem 2. In a nutshell, this result is more useful than Theorem 1 because it is hard to ensure that mean productivities are equal even with fine controls. It is, however, possible to stack the deck in favor of the disadvantaged group (to ensure they have higher mean productivity) by, for instance, comparing higher education disadvantaged group workers to those with lower education in the advantaged group. Evidence for taste-based discrimination is more compelling if despite this, we nonetheless find that the wages for the advantaged group are strictly first-order stochastically dominant. In other words, we do not compare

groups with the same individual characteristics – we compare individuals from the disadvantaged group with characteristics, which are typically associated with higher productivities, e.g., years of schooling or AFQT scores, with individuals from the advantaged group without these characteristics.

4. EMPIRICAL APPLICATION

In this section, we describe how Theorem 2 yields a simple empirical test to uncover taste-based discrimination on commonly available cross-sectional data. We first describe the methodology and then, as a proof of concept, apply our test to Census and NLSY data.

It is worth presenting a high level motivation for our approach before we provide specific details. There is a tradition in labor economics of using the KOB decomposition to determine both the presence of discrimination and to measure its magnitude. As Guryan and Charles (2013) explain, this method "separates differences in average wages, for example, into the part that is explained by differences in characteristics (e.g., education), the part that is explained by differences in returns to those characteristics (e.g., returns to education) and unexplained differences. Many in this literature have called the unexplained differences, or both the unexplained and the differences in returns, the result of discrimination."

As we have argued in Theorem 1, even if the mean productivities are identical, arbitrary wage gaps can arise absent discrimination if the productivity distributions of both groups are not the same. In other words, we genuinely need to decompose the entire wage distributions, and not only their means, as typical KOB decompositions do.

We decompose the wage distributions controlling for individual and job characteristics. In common with the large empirical literature on discrimination, we may be controlling for too little (omitted variables) or too much. As Guryan and Charles (2013) explain: "the variables the researcher controls for might themselves be affected by discrimination. Controlling for such variables can cause the unexplained differences to understate the role that discrimination in general plays in determining wage gaps."

While our analysis suffers from these problems, we think that this is less of a concern. To start with, we are only interested in detecting taste-based discrimination, and not in measuring its magnitude. In addition, even if the variables we control for have been affected by discrimination, we can use this to our advantage. For instance, if we think there is discrimination in education, this implies that, on average, Blacks (having to overcome more barriers) will be of higher ability compared to similarly educated whites. The assumption of Theorem 2 is, therefore, more likely to be true. In fact, we go one step further. We compare Blacks with more schooling to lesser educated whites. If the wages of whites still strictly first-order stochastically dominate those of Blacks, then a conclusion that taste-based discrimination is present is only invalid if we believe that whites with less schooling are more productive on average (which we do

not).

4.1. THE METHODOLOGY

We compare the wage distributions for Black (group 1) and white (group 2) workers within a given occupation. The assumption here is that wages in a given occupation are governed by a single wage function and the different characteristics of workers affect their expected productivity (and hence their wage). Our goal is to test Theorem 2: that is, under the assumption that Blacks are more productive on average, we want to test whether the wage distribution G_2 of whites strictly first-order stochastically dominates the wage distribution G_1 of Blacks. Recall that, if $G_2 \succ_1 G_1$, then the wage distributions cannot be the result of statistical discrimination alone. Thus, our null hypothesis is that statistical discrimination alone rationalizes the data, and we are interested in rejecting the null. We interpret the rejection of the null as evidence of taste-based discrimination.

There are two challenges to overcome. The first is that the econometrician does not observe the true wage distributions G_1 and G_2 but instead only observes empirical samples. Second, we assume that (unobserved) individual productivities are correlated with (observed) individual characteristics (education, age, experience, etc). However, the distributions of individual characteristics in any occupation vary between Black and white workers, perhaps due to pre-labor market discrimination. We need to control for these differences to ensure we can apply Theorem 2 (which, again, requires Blacks to be more productive on average).

The first challenge can be easily dealt with. We can test whether $G_2 \succ_1 G_1$ by Kolmogorov-Smirnov-type tests. This is equivalent to testing whether G_2 first-order stochastically dominates G_1 but not vice versa. Examples of such tests can be found in McFadden (1989) and Barrett and Donald (2003).

The second difficulty is more challenging. We observe neither the true productivity distributions of the workers nor the universe of signals that employers get. To overcome this difficulty, we employ a version of the KOB decomposition. We first compute the (counterfactual) distribution $\widehat{G}_1$ of wages for Blacks that we would have observed had they faced the wage setting process of whites that is, had employers perceived them to be white. This counterfactual distribution captures both potential statistical discrimination (via the different signals for each group) and taste-based discrimination. We then test the null hypothesis of $\widehat{G}_1 \not\succ_1 G_1$: if the null is rejected, statistical discrimination alone cannot explain the difference between the observed wage distribution G_1 and the counterfactual distribution $\widehat{G}_1$.

More precisely, let X be a vector of observable individual characteristics such as years of schooling, age and state of residence. Let $G_i(\cdot|X)$ be the observed distribution of wages conditional on characteristics X and F_i° the joint probability over expected productivity (in the eyers of the employer) and individual

characteristics. We perform the following decomposition:

$$\begin{aligned} G_1(w) - G_2(w) &= \int G_1(w|X)dF_1^\circ(X) - \int G_2(w|X)dF_2^\circ(X), \\ &= \left[\int G_1(w|X)dF_1^\circ(X) - \int G_2(w|X)dF_1^\circ(X)\right] \\ &\qquad + \left[\int G_2(w|X)dF_1^\circ(X) - \int G_2(w|X)dF_2^\circ(X)\right], \\ &= \left[G_1(w) - \widehat{G}_1(w)\right] + \left[\widehat{G}_1(w) - G_2(w)\right], \end{aligned}$$

where $\widehat{G}_1(w) := \int G_2(w|X)dF_1^\circ(X)$. A recent interpretation of this decomposition is provided by Bohren, Hull, and Imas (2022). They interpret the term on the left $G_1(w) - G_2(w)$ as *total* discrimination, the first term on the right $G_1(w) - \widehat{G}_1(w)$ as *direct* discrimination and the final term $\widehat{G}_1(w) - G_2(w)$ as *systemic* discrimination. Total discrimination compounds the differential treatment of Blacks and whites before entering the job market (for instance, due to barriers in educational attainment) with the differential treatment after entering the job market. We are interested in the direct discrimination term, that is, the comparison of G_1 and $\widehat{G}_1$. The comparison is analogous to a correspondence study since we compare the wages of Blacks with the wages they would have obtained had they been treated as whites (but retaining their individual characteristics). Conceptually, it is similar to exposing employers to resumes which differ only in names.

To compute the counterfactual distribution $\widehat{G}_1$, we follow the method of Chernozhukov, Fernández-Val, and Melly (2013). In a nutshell, this consists of estimating the marginal $F_i(X)$ over individual characteristics from the empirical distribution and the conditional $G_i(\cdot|X)$ from either distribution or quantile regressions (for each w, $X \mapsto G_i(w|X)$ is viewed as a function of the covariates X). We then test whether $\widehat{G}_1 \succ_1 G_1$, which the approach of Chernozhukov, Fernández-Val, and Melly (2013) permits. Importantly, and unlike DiNardo, Fortin, and Lemieux (1996), the method is flexible enough to make statistical inference on the entire distributions, which is needed to test for first-order stochastic dominance.

Now, suppose we conclude that $\widehat{G}_1 \succ_1 G_1$. We can then infer that the wage distributions *cannot* be rationalized by statistical discrimination alone if

$$\int \mathbb{E}_{F_1^\circ}[\theta_1|X]dF_1^\circ(X) \geq \int \mathbb{E}_{F_2^\circ}[\theta_2|X]dF_1^\circ(X).$$

For instance, this hypothesis is satisfied whenever the productivity distribution θ_i is independent of race i, conditional on characteristics X. It is, however, weaker. The assumption allows for the expected productivity $\mathbb{E}_{F_1^\circ}[\theta_1|X] \geq \mathbb{E}_{F_2^\circ}[\theta_2|X]$ of Blacks to be higher for some individual characteristics X and lower for others. Economic theories support both scenarios. On the one hand, barriers to entry into the labor

market and obtaining education suggest that Blacks are positively selected into the labor market (which is our belief). While we are not aware of empirical evidence supporting positive selection of Blacks, Ashraf, Bandiera, Minni, and Quintas-Martınez (2022) document that women are positively selected into the labor market (and that women are more comparatively productive than men in countries where female labor force participation is lower). On the other hand, classic theories of statistical discrimination (such as Coate and Loury, 1993) argue that Blacks may under-invest in productive skills and consequently make stereotypes self-fulfilling. To deal with this latter hypothesis, we also compare Blacks with more schooling to whites with less; here, it is significantly harder (and, in our view, implausible) to argue that Blacks have lower mean productivity.

We end this subsection with a brief discussion on the choice of the covariates X that should be included in the computation of the counterfactual distribution $\widehat{G}_1$. We suggest including as many covariates as necessary to make the assumption of ordered means assumption as plausible as possible, subject to data and computational limitations. It is not essential that the employers observe all the variables in X when setting wages. Employers may not observe certain covariates, but may receive other information (unobserved by the analyst) that is correlated with the unobserved variables in X. In other words, since the individual characteristics employers observe may be arbitrarily correlated with the the individual characteristics the analyst observes, it makes no differences to the empirical analysis whether we assume that employers observe the individual covariates X we use or not. A similar observation appears in Altonji and Pierret (2001) who employ NLSY data and use AFQT scores, among other controls, to proxy the information employers obtain over time. We explicitly make this remark because one of our robustness checks uses NLSY data and we include AFQT scores and past wages—variables that may not be observed by employers—as controls.

4.2. EVIDENCE OF TASTE-BASED DISCRIMINATION

We employ two data sets to apply our theoretical insights. We conduct our main analysis on Census data since there are more observations and the data is recent. For a robustness check, we exploit the fact that the NLSY is a panel data set. We begin by describing the sample construction for each dataset.

4.2.1. The data

Census: We use the 5 year American Community Survey 2021 (ACS) that contains all households and persons from the 1% ACS samples from 2017-2021. We restrict the 2021 data to men aged 30-55 years (to capture their prime labor market years) and to those who are employed, working full time (52 weeks), and working for a wage. The main income measure we use is the INCWAGE variable capturing "Wage and Salary Income," although we also consider alternative measures such as total personal income and total personal earned income.

We conduct analysis at the occupation level. We create eleven general categories of occupation status from the four digit codes (2010 basis) available in the census.[13] This aggregation is based on documentation from the Census after 2020. The 11 groups are: (1) Management, Business, and Financial, (2) Professional, (3) Service, (4) Sales and related, (5) Office and administrative support, (6) Farming, fishing, and forestry, (7) Construction and extraction, (8) Installation, maintenance, and repair, (9) Production, (10) Transportation, and material, and (11) Armed forces occupations. We choose to aggregate in this way because we conduct our analysis at the occupation level and the sample becomes too small if we disaggregate further.

Race is measured from the self reported general race category in the census. In our richest model, we include highest grade attained, marital status, field of degree (general version), age, and state of residence as controls.

NLSY: From the NLSY79 we restrict our sample to men who worked full time (52 weeks) in 1998 and 2000 and who reported positive wages in those years. We selected those years as individuals in the NLSY were between the ages of 14-22 at the time of their first interview in 1979. Hence, in 2000 these individuals are in the age range of 35-43, which were their prime labor market years. The main wage variable we use from the NLSY is about the respondent's "amount of wages, salary, and tips" in the past year (so survey year 2000 relates to wages in 1999, etc.).

In addition to variables capturing occupation in 2000 and 1998 (3 digit CPS 1980 codes), we also have the highest grade completed (as of calendar year 2000), and the AFQT percentile score (all respondents took the AFQT in 1981 and we use the 2006 revised percentiles of this measure from the data).

4.2.2. Results

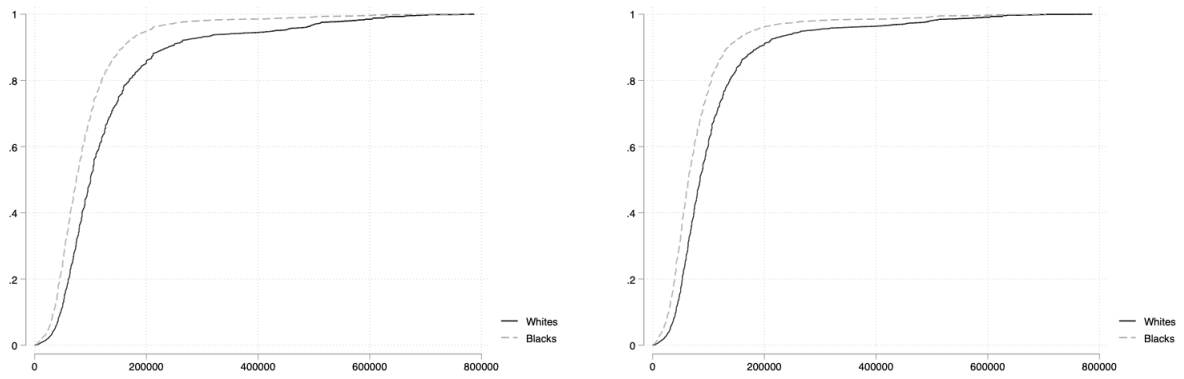
The one sentence summary of our results is that, across specifications, the wage distribution of Blacks is strictly first-order stochastically dominated by the counterfactual distribution of wages they would receive were they to be treated as whites. We interpret this as evidence of taste-based discrimination via the lens of Theorem 2. Over the course of this subsection, we will attempt to convince the reader that this interpretation is warranted.

We begin by discussing our main results from the Census data. As mentioned above, we categorize workers into 11 occupation categories and conduct our analysis at the occupation level. So that we do not have to display 11 figures, we choose 2 occupation categories to display where we believe taste-based discrimination might be present: (i) professional and related occupations and (ii) management, business and financial occupations. We briefly discuss results from the remaining occupations at the end of this section and in further detail in the appendix.

[13] Details are in the Appendix.

In Figure 1, we plot the distributions of wages for Black and white workers in these two occupation categories. These plots are meant to be descriptive and first-order stochastic dominance is clearly visible.

Figure 1: Raw Wage Distributions by Race



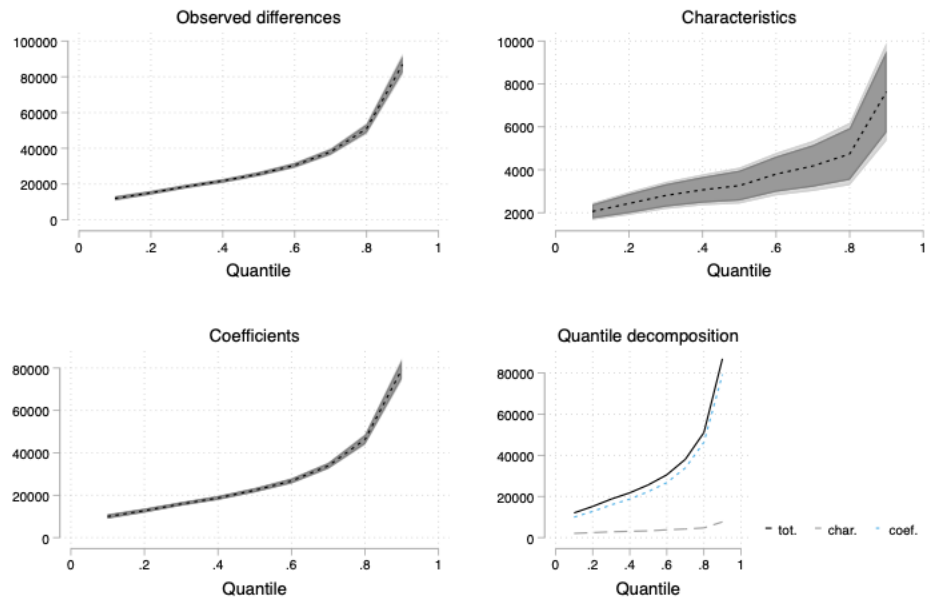
(a) Management, Finance, and Business Occupations (b) Professional Occupations

Data is from 5 year ACS 2017-2022. Men aged 30-55, working full time (52 weeks), and working for wages.

We then turn to our main specification: the quantile decomposition discussed earlier using the methodology of Chernozhukov, Fernández-Val, and Melly (2013). We conduct the decomposition for both occupations with the following rich set of controls: dummies for education, dummies for degree attained, census region of residence (9 regions), age, and age squared. Figures 2 and 3 display the results for both occupation categories. Each figure contains four panels: quantiles are on the x-axis and differences between the annual wages of whites and Blacks are on the y-axis. Note that an equivalent way of stating that wages for a group strictly first-order stochastically dominates the other is that the wages for the first group are weakly higher than the second at *all* quantiles and strictly higher at some. For the decompositions we use the default CDECO settings, which are that estimation is based on linear quantile regressions based on Koenker and Bassett (1978), hundred bootstrap replications are performed for inference, and both pointwise and uniform confidence intervals at 95% are constructed.

The bottom right panel in each figure summarizes the remaining three so it suffices to describe those. The top left panel plots the difference in wages at each quantile of the wage distribution; this is the total discrimination we described above. The top right panel is the share of that difference attributed to the different characteristics of Black and white workers, the aforementioned systemic discrimination. The main panel of interest for us is the bottom left panel which captures the direct discrimination (the difference between $\widehat{G}_1$ and G_1). Observe that the difference is significantly greater than 0 at all quantiles of the wage distribution. In other words, if we were to believe that conditioning on the above mentioned observables, the average productivity of Blacks is at least as high as whites, then this is evidence of taste-based

Figure 2: Quantile Decompositions: Management, Finance, and Business



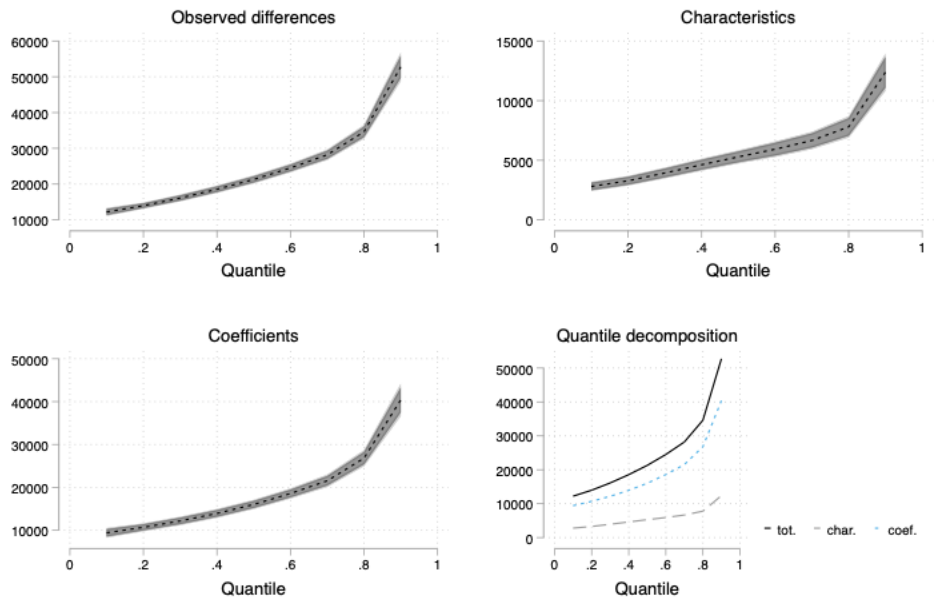
Data is from 5 year ACS 2017-2022. Men aged 30-55, working full time (52 weeks), and working for wages. CDECO command based on Chernozhukov, Fernández-Val, and Melly (2013). Controls include age, age-squared, dummies for educational attainment, dummies for census region, and dummies for type of degree obtained. 95% confidence bands.

discrimination as established in Theorem 2.

To make this interpretation even starker, we compare Black workers with an associate's degree and white workers with a high school degree. Specifically, we consider the subsample of Black and white workers with associate's and high school degrees respectively and conduct the quantile decomposition with all the controls—marital status, state of residence, age and age squared—barring education (since education has already been accounted for). We display the results in Figures 4 and 5; once again, the bottom left panels show that the wages of Blacks with associate's degrees are first-order stochastically dominated by the wages that they would receive if employers treated them as whites with high school degrees. Similar results are obtained if, instead of associate's degrees we consider Blacks with some college (this measures 1 or more years of college credit, but not a *completed* Bachelors or Associate's degree), figures are in the appendix. As mentioned earlier, we view this to be compelling evidence for taste-based discrimination since it seems implausible to us that Blacks with more education are perceived less productive on average.

It is worth mentioning that we compare associate's and high school degrees since this is the highest education disparity between Blacks and whites under which we get $\widehat{G}_1 \succ_1 G_1$. If we were to compare Blacks with completed college degrees with whites who only have high school degrees, we get the opposite relation $G_1 \succ_1 \widehat{G}_1$ (see figures in the Appendix) which is reassuring but (in light of already mentioned

Figure 3: Quantile Decompositions: Professional Occupations



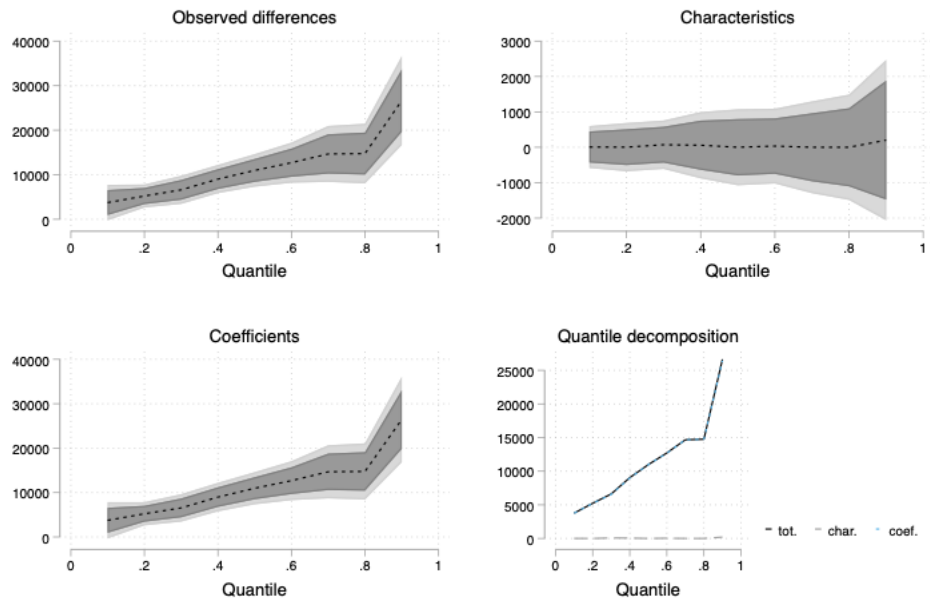
Data is from 5 year ACS 2017-2022. Men aged 30-55, working full time (52 weeks), and working for wages. CDECO command based on Chernozhukov, Fernández-Val, and Melly (2013). Controls include age, age-squared, dummies for educational attainment, dummies for census region, and dummies for type of degree obtained. 95% confidence bands.

results) only slightly.

The results from the remaining occupations are qualitatively similar. Details can be found in the Online Appendix.

We now describe out results from the NLSY. These data are not ideal since the sample is smaller and the data are not recent but we choose to include them nonetheless since the NLSY, unlike the Census, is a panel data set. In particular, this allows us to control on past wages and occupations (note that we do not cut the sample by occupation category as this will reduce sample sizes even further). These controls allow us to compare Black and white workers who worked the same job in the previous period at the same wage. Viewed through the lens of our model, this says that unless employers engage in taste-based discrimination, the expected productivity of these two groups of workers should be the same – as wages only depend on expected productivities. If employers engage in taste-based discrimination against Blacks, then this implies that, conditioning on past wages and occupations, Blacks should be more productive than whites as required. The remaining possibility is that Blacks are of lower mean productivity, but got paid the same as whites (with higher mean productivity) in the previous period. We do not view this to be probable.

Figure 4: Quantile Decompositions: Whites with HS and Blacks with Assoc Degrees working in Management, Finance, and Business occupations



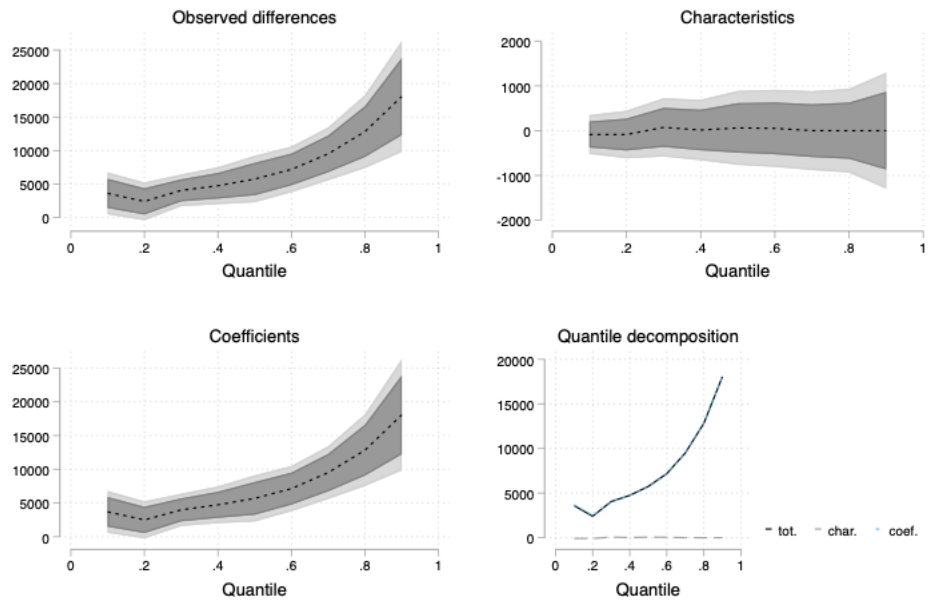
Data is from 5 year ACS 2017-2022. Men aged 30-55, working full time (52 weeks), and working for wages. Sample consists of Blacks with Associates Degrees and Whites with High School Degrees in Management, Finance, and Business occupations. CDECO command based on Chernozhukov, Fernández-Val, and Melly (2013). Controls include age and age-squared.

Figure 6 only displays the wage structure effect (required to conclude $\widehat{G}_1 \succ_1 G_1$) under increasingly stringent controls. The bottom right figure displays the full specification that controls not only for past wages and occupations but also AFQT scores. Strict first-order stochastic dominance is clearly visible in all plots; in the bottom right plot $\widehat{G}_1$ and G_1 are not significantly different at low quantiles but at high quantiles $\widehat{G}_1$ is significantly higher. The KS statistic p-values confirm the visual evidence.

5. EXTENDING THE THEORY

In this section, we generalize the results in Section 3 along several directions. We first impose natural shape restrictions on the set of permissible wage functions and derive the stronger testable implications on the set of wage distribution pairs. We then characterize the set of rationalizable wage distribution pairs when we do not assume mean productivities are ordered. We show how these results can be inverted to derive bounds on the productivity differences required to rationalize the wage distributions.

Figure 5: Quantile Decompositions: Whites with HS and Blacks with Assoc Degrees working in Professional Occupations



Data is from 5 year ACS 2017-2022. Men aged 30-55, working full time (52 weeks), and working for wages. Sample consists of Blacks with Associates Degrees and Whites with High School Degrees in Professional occupations. CDECO command based on Chernozhukov, Fernández-Val, and Melly (2013). Controls include age and age-squared.

5.1. CONCAVE AND CONVEX WAGES

The results from Section 3 allow for arbitrary wage functions. There might be settings where there is natural structure on the wage functions. Imposing such structure has the advantage of strengthening the testable implications of the model.

Motivated by the increasing inequality, one natural assumption to impose is that wage functions are convex in the expected productivity of the worker. We denote

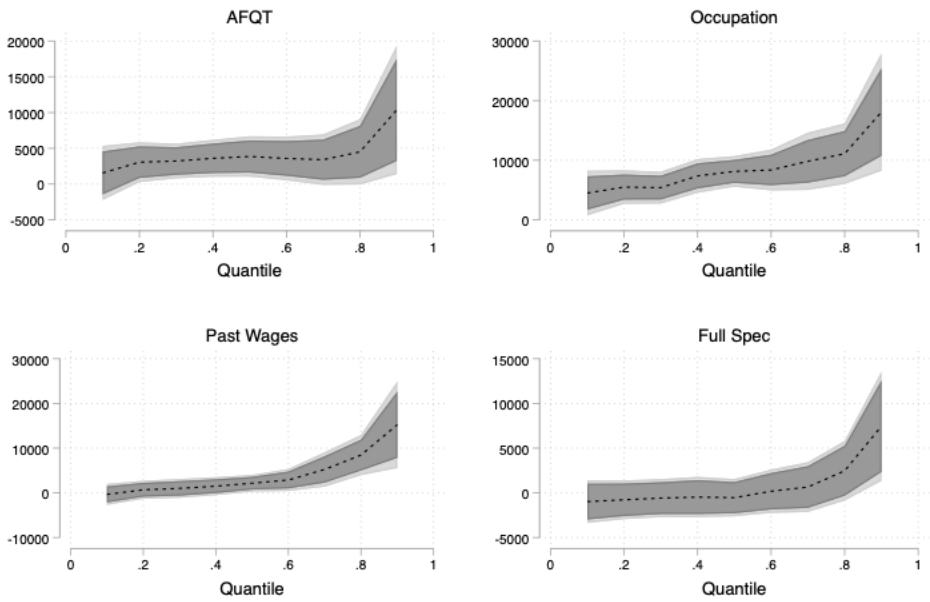
$$\mathcal{W}_{conv} := \{W \in \mathcal{W} \,|\, W \text{ is strictly convex}\}$$

to be the set of strictly increasing and strictly convex (and therefore continuous) functions. Likewise, we use $\mathcal{W}_{conc}$ to denote the set of strictly increasing and strictly concave functions.

We say that wage distribution G_i dominates distribution G_j in the *strict concave order*, if

$$\int_0^{\overline{w}} M(W)dG_i(w) > \int_0^{\overline{w}} M(W)dG_j(w)$$

Figure 6: Quantile Decompositions: Wage Structure Effects from the NLSY



Data is from the NLSY-79. Wages recorded in calendar year 2000 for men working full time (52 weeks). CDECO command based on Chernozhukov et al (2013). Only wage structure effect (coefficient plot from CDECO) is displayed. All estimates control for highest grade completed, age, and age-squared. In addition, the top left panel controls for AFQT percentile scores, top right panel controls for dummies for occupation in 2000, bottom left controls for wages in 1998, and the bottom right controls for all of these as well as dummies for occupation in 1998. The first three panels clearly show that White wages FOSD Black wages. For the full specification estimate, the KS statistic p-values for the null that Whites FOSD Blacks is 0.66 and the p-values for Blacks FOSD Whites is 0.03.

for every strictly increasing, strictly concave function M. This order is closely related to the mean-preserving spread order $\succcurlyeq_2$. The latter implies second-order stochastic dominance but not vice versa since second-order stochastic dominance does not require equal means. The strict concave order implies strict second-order stochastic dominance but does not require both distributions to have equal means. The *strict convex order* can be analogously defined when the above inequality holds for all strictly increasing, strictly concave function M.

We derive a version of Theorem 1 assuming wages are convex.

THEOREM 3. *Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_=$, $\mathcal{W}_{conv}$) if, and only if, neither G_1 nor G_2 dominates the other in the strict concave order.*

This result also yields the analogous version of Theorem 2 when we assume ordered means. If we were to instead assume wages were concave, the statement would be verbatim with $\mathcal{W}_{conc}$ replacing $\mathcal{W}_{conv}$ and the strict convex order replacing the strict concave order.

If $G_i \succ_1 G_j$, then G_i also dominates distribution G_j in both the strict concave and convex orders. Thus, Theorem 3 shows that the testable implications are stronger under the assumption of convex wages but, once again, they take a form that is easy to take to the data. There are well-known tests for higher orders of stochastic dominance developed in the econometrics literature (once again, see Barrett and Donald, 2003).

5.2. NON-ORDERED MEANS

All of the above results required mean productivities to be either equal or ordered. Instead, suppose we assume that

$$\mathcal{H}_{|1-2|\leq d} := \{(H_1, H_2) \mid |\mathbb{E}_{H_1}[\theta_1] - \mathbb{E}_{H_2}[\theta_2]| \leq d\},$$

that is, both groups have productivity distributions whose mean differs by at most $d \in \mathbb{R}_+$. Additionally, we assume

$$\mathcal{W}_{L1} := \{W \in \mathcal{W} \mid |W(\theta) - W(\theta')| \leq |\theta - \theta'| \text{ for all } \theta, \theta' \in \Theta\},$$

that is, wage functions are 1-Lipschitz. This technical restriction imposes some discipline on the wage function, that is, it provides a bridge between changes in wages and changes in productivities. It is weaker than assuming that wage functions are differentiable with slope less than 1.

We now characterize wage distributions that are rationalizable under these assumptions.

THEOREM 4. *Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_{|1-2|\leq d}$, $\mathcal{W}_{L1}$) if, and only if, either*

(i) the wage gap is less than d, that is, $|\mathbb{E}_{G_1}[w] - \mathbb{E}_{G_2}[w]| \leq d$, or

(ii) neither G_1 nor G_2 strictly first-order stochastically dominates the other.

As Theorem 1 shows, if wages are not ordered by first-order stochastic dominance, they can be rationalized with equal mean productivities. Consequently, they will be rationalized (given $\mathcal{H}_{|1-2|\leq d}$, $\mathcal{W}_{L1}$) for any $d > 0$; the additional restriction on the wage function $\mathcal{W}_{L1} \subset \mathcal{W}$ does not affect Theorem 1 (the older version Deb and Renou, 2022 has a formal statement of this). So the real contribution of the above result is the first condition (i). This shows that the wage gap can be a useful statistic to test for the presence of taste-based discrimination, but *only* when the wage distributions are ordered by strict first-order stochastic dominance.

An issue with applying the result in Theorem 4 to data is that the researcher has to choose the appropriate mean productivity difference d to test for. However, the statement of this theorem can be inverted to show that the wage gap is a *tight* lower bound for mean productivity differences required to rationalize

the wage distributions. In other words, if we want statistical discrimination alone to rationalize the data, the wage gap is the smallest difference in productivity means required, whenever one wage distribution first-order stochastically dominates the other.

THEOREM 4 (CONTINUED). *Suppose $G_2 \succ_1 G_1$. Then, there exist (H_1, F_1, H_2, F_2, W) that jointly induce G_1, G_2 such that $W \in \mathcal{W}_{L1}$ and $\mathbb{E}_{H_2}[\theta] - \mathbb{E}_{H_1}[\theta] = \mathbb{E}_{G_2}[w] - \mathbb{E}_{G_1}[w]$.*

Moreover, every (H_1, F_1, H_2, F_2, W) with $W \in \mathcal{W}_{L1}$, that jointly induce G_1, G_2 satisfy $\mathbb{E}_{H_2}[\theta] - \mathbb{E}_{H_1}[\theta] \geq \mathbb{E}_{G_2}[w] - \mathbb{E}_{G_1}[w]$.

In words, this result states every (H_1, F_1, H_2, F_2, W) with $W \in \mathcal{W}_{L1}$ that induce wage distributions $G_2 \succ_1 G_1$ have the feature that the differences in mean productivities is at least the wage gap and that this bound is tight. Thus, the wage gap is a useful measure of the minimum productivity difference required to rationalize the wage distributions assuming that there is no taste-based discrimination at work. The data can be used to contextualize this bound and, in a sense, we already conduct such an exercise in Section 4.2.

To see this, lets revisit our empirical application. Compare white and Black male workers with high school degrees who work in professional occupations. If we were convinced that there was no taste-based discrimination, then a lower bound of mean productivity differences between the groups is the wage gap (-11,563 USD). Now lets compare Black workers with high-school degrees to Black workers with associate's degrees. We can once again use Theorem 4 to conclude that a lower bound of the mean productivity differences between the groups is the difference in mean wages (-8,664 USD). In other words, in the absence of taste-based discrimination, the lower bound on productivity differences between white and Black high-school educated worker corresponds to more than two additional years of schooling for Black workers.

Instead of bounding absolute productivity differences, one could also try to bound percentage productivity differences. Formally, let

$$\mathcal{H}_{1/2 \geq \alpha} := \{(H_1, H_2) \,|\, \mathbb{E}_{H_1}[\theta_1] \geq \alpha \mathbb{E}_{H_2}[\theta_2]\},$$

where $\alpha \in (0, 1)$ denote the set of productivity distribution pairs such that group 1's mean productivity is at least a fraction α of that of group 2. As with Theorem 4, one could first characterize the set of rationalizable wage distributions and then invert result to derive the bound. Unfortunately, there is too little structure in the model to bound percentage differences.

THEOREM 5. *Suppose $G_1(0) = G_2(0) = 0$. Then, for every $\alpha \in (0, 1)$, every pair of wage distributions*

is rationalizable (given $\mathcal{H}_{1/2\geq\alpha}$, $\mathcal{W}_{L1}$)

This result states that, for any fraction α, every pair of wage distributions are rationalizable with productivity distributions whose means are within a fraction α of each other. Loosely speaking, this is because we can consider wage functions that are very flat for low wages but have slope one for higher wages. With such a wage function, absolute differences in higher wages correspond to absolute differences in productivities but because the values of the latter are high, the percentage difference between mean productivities becomes small. The additional assumption $G_1(0) = G_2(0) = 0$ is trivially satisfied in any application since we restrict attention to working adults and no one works for zero wages.

6. CONCLUDING REMARKS

In this paper, we developed a simple but general framework that lends itself to testing for taste-based discrimination in widely available cross-sectional wage data. We view our contributions to be several. First, unlike a bulk of the literature, our modeling choices allow for unrestricted signals and can flexibly capture imperfectly competitive labor markets. Despite this generality, the testable implications of the model are easy to describe and test. We demonstrate how an ordered means assumption can be validated on either cross sectional or panel data by comparing less educated whites to more educated Blacks or by conditioning on past wages and occupations respectively. Our theoretical results provide a lens through which standard wage decompositions can be interpreted and show that these can be used to uncover evidence of taste-based discrimination. Finally, we demonstrate the flexibility of our framework by deriving the testable implications of the model under difference assumptions on the set of permissible productivity distribution pairs and wage functions.

We end the paper by discussing a few remaining assumptions of the model and suggest some directions for future research. Throughout, we assumed that there were two groups. One could, in principle, test whether the wage distributions for three or more groups—each with their own productivity distribution and signal but with there being a common wage function—are rationalizable. In contexts where there are two or more advantaged groups, contrasting the wage distribution of a disadvantaged group against all advantaged groups simultaneously (as opposed to each one individually) can result in stronger testable restrictions. In fact, we show (in the Online Appendix) that Theorem 1 and Theorem 2 generalize to such a setting. The wage distributions are rationalizable if, and only if, the wage distribution of the disadvantaged group (with higher mean productivity) is not first-order stochastically dominated by a convex combination of the wage distributions of the advantaged groups.

A critical assumption in our model is that wages only depend on the posterior mean which is, of course, an assumption that is commonly made when assuming perfectly competitive labor markets. With that said, one could allow wages to depend on higher moments of the posterior distribution inferred by employers

upon observing a signal realization. Indeed, the seminal work of Aigner and Cain (1977) allows wages to depend on both the mean and the variance of the posterior. We show (in the Online Appendix) that such additionally generality renders vacuous the testable implications of our model. Specifically, even if wages are linear in the mean and variance of the posterior, every pair of wage distributions can be rationalized under the assumption of equal mean productivities.

While we considered many combinations of assumptions on the sets of productivity distributions and wage functions, there are of course other restrictions that one could impose based on the context. For instance, information or estimates about labor market competition may allow the researcher to bound slopes of the wage function. This in turn, could allow the researcher to infer bounds on mean productivity differences (absolute or percentage) required to rationalize the wage distributions absent taste-based discrimination. In the Online Appendix, we describe a general procedure for either testing for rationalizability given arbitrary restrictions $\widehat{\mathcal{H}}$ and $\widehat{\mathcal{W}}$ or for deriving bounds on mean productivity differences given arbitrary $\widehat{\mathcal{W}}$.

Lastly, our empirical application demonstrated (by comparing less educated whites to more educated Blacks) how the data can validate assumptions on the set $\widehat{\mathcal{H}}$ (in this case, $\widehat{\mathcal{H}} = \mathcal{H}_{\geq}$) of productivity distributions. Our approach can be adapted to richer data sets where there may be additional information about worker productivity. In a sense, the robustness check we conduct using NLSY data does precisely this by exploiting the panel data structure to control for past wages and occupations. But this of course, does not employ the full richness of the panel as we only use information in one previous period. To do so, we would need to introduce dynamics into our framework which is something we hope to do in future research.

REFERENCES

Aigner, D. J. and G. G. Cain (1977): "Statistical theories of discrimination in labor markets," *Industrial and Labor Relations Review*, 30, 175–187.

Altonji, J. G. and R. M. Blank (1999): "Race and gender in the labor market," *Handbook of labor economics*, 3, 3143–3259.

Altonji, J. G. and C. R. Pierret (2001): "Employer learning and statistical discrimination," *Quarterly Journal of Economics*, 116, 313–350.

Anwar, S. and H. Fang (2006): "An alternative test of racial prejudice in motor vehicle searches: Theory and evidence," *American Economic Review*, 96, 127–151.

Arnold, D., W. Dobbie, and C. S. Yang (2018): "Racial bias in bail decisions," *Quarterly Journal of Economics*, 133, 1885–1932.

Ashraf, N., O. Bandiera, V. Minni, and V. Quintas-Martinez (2022): "Gender roles and the misallocation of labour across countries," *Unpublished Manuscript*, 2.

Barrett, G. F. and S. G. Donald (2003): "Consistent tests for stochastic dominance," *Econometrica*, 71, 71–104.

Becker, G. S. (1957): *The economics of discrimination*, University of Chicago press.

——— (1993): "Nobel lecture: The economic way of looking at behavior," *Journal of Political Economy*, 101, 385–409.

Bertrand, M. and E. Duflo (2017): "Field experiments on discrimination," *Handbook of Economic Field Experiments*, 1, 309–393.

Blinder, A. S. (1973): "Wage discrimination: reduced form and structural estimates," *Journal of Human resources*, 436–455.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2019): "Inaccurate statistical discrimination: An identification problem," Tech. rep., National Bureau of Economic Research.

Bohren, J. A., P. Hull, and A. Imas (2022): "Systemic discrimination: Theory and measurement," Tech. rep., National Bureau of Economic Research.

Canay, I. A., M. Mogstad, and J. Mountjoy (2020): "On the use of outcome tests for detecting bias in decision making," Tech. rep., National Bureau of Economic Research.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on counterfactual distributions," *Econometrica*, 81, 2205–2268.

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

Deb, R. and L. Renou (2022): "Which wage distributions are consistent with statistical discrimination?" *CEPR Discussion Paper No. 17676*.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," *Econometrica*, 64, 1001–1044.

Fang, H. and A. Moro (2011): "Theories of statistical discrimination and affirmative action: A survey," *Handbook of Social Economics*, 1, 133–200.

Guryan, J. and K. K. Charles (2013): "Taste-based or statistical discrimination: the economics of discrimination returns to its roots," *Economic Journal*, 123, F417–F432.

Heckman, J. J. (1998): "Detecting discrimination," *Journal of Economic Perspectives*, 12, 101–116.

Juhn, C., K. M. Murphy, and B. Pierce (1993): "Wage inequality and the rise in returns to skill," *Journal of Political Economy*, 101, 410–442.

Kitagawa, E. M. (1955): "Components of a difference between two rates," *Journal of the American Statistical Association*, 50, 1168–1194.

Knowles, J., N. Persico, and P. Todd (2001): "Racial bias in motor vehicle searches: Theory and evidence," *Journal of Political Economy*, 109, 203–229.

Lang, K. and J.-Y. K. Lehmann (2012): "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Literature*, 50, 959–1006.

Lang, K. and A. K.-L. Spitzer (2020): "Race discrimination: An economic perspective," *Journal of Economic Perspectives*, 34, 68–89.

McFadden, D. (1989): "Testing for stochastic dominance," in *Studies in the economics of uncertainty: In honor of Josef Hadar*, Springer, 113–134.

Neumark, D. (2012): "Detecting discrimination in audit and correspondence studies," *Journal of Human Resources*, 47, 1128–1157.

Oaxaca, R. (1973): "Male-female wage differentials in urban labor markets," *International Economic Review*, 693–709.

Onuchic, P. (2022): "Recent contributions to theories of discrimination," *arXiv preprint arXiv:2205.05994*.

Phelps, E. S. (1972): "The statistical theory of racism and sexism," *American Economic Review*, 62, 659–661.