University of Toronto
Department of Economics

Working Paper 736

Which wage distributions are consistent with statistical discrimination?

By Rahul Deb and Ludovic Renou

September 10, 2022

WHICH WAGE DISTRIBUTIONS ARE CONSISTENT WITH STATISTICAL DISCRIMINATION?

Rahul Deb[◊]
Ludovic Renou[†]

September 7, 2022

ABSTRACT

When are the wage distributions for two groups consistent with a general reduced form model of statistical discrimination? In our model, each group's productivities are drawn from different distributions with common means. Productivities are unobserved but inferred from noisy signals. Wages are determined by a strictly increasing (but otherwise unrestricted) function of the posterior expectation of the productivities (computed from the signal). We show that a pair of wage distributions are consistent with this model of statistical discrimination if, and only if, neither wage distribution first-order stochastically dominates the other. A rejection of this condition thus provides evidence of bias.

[◊]Department of Economics, University of Toronto, rahul.deb@utoronto.ca
[†] School of Economics and Finance, Queen Mary University of London, lrenou.econ@gmail.com
We are very grateful to Paula Onuchic for engaging in a conversation with us that inspired this paper, to Prashant Bharadwaj for several detailed discussions and to Namrata Kala, Mallesh Pai, Maher Said for insightful comments. 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

Wages for members of different groups (women versus men, blacks versus whites) frequently differ even when their observable characteristics are the same. Is this the result of differential information that employers have across groups or instead, the consequence of employer bias? The answer to this question is important as, in particular, the ability to distinguish between these explanations is needed to devise appropriate corrective policies. The purpose of this paper is to theoretically establish the properties of wage distributions that can possibly arise from statistical discrimination due to unbiased employers receiving different signals about the productivity of the members of the two groups. Our theoretical condition naturally lends itself to statistical testing and this in turn is useful, because a rejection of this test can be interpreted as (micro-founded) evidence that wage differences are the result of bias.

Specifically, suppose a researcher observes the wage distributions (appropriately controlling for observables) for two groups. Is it ever possible to conclude, from such basic observational data, that these wage distributions did not arise from statistical discrimination? If so, when? We provide precise answers to these questions. To do so, we develop a simple but general reduced-form model of statistical discrimination in the spirit of Phelps (1972). There are two groups whose productivity distributions have identical means, but can otherwise be different. The group identity is observable to employers, but productivities are not. Instead, there are group-dependent "statistical experiments" that generate signals about the underlying productivity. As an example, signals could be the information that employers receive from the job screening process that includes interviews, tests, curricula vitae etc. Signals 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) of the unobserved productivity. Therefore, each group's statistical experiment generates a (generically different) distribution over posterior productivity estimates. Wages are determined via a strictly increasing, continuous function of the posterior productivity estimate that, importantly, does not depend on the group. The model is reduced form in that we do not microfound the statistical experiments or the wage function (although foundations can easily be provided) but very general in that both are completely unrestricted (as long as the wage function is strictly increasing and continuous).

Formally, our question is the following. Suppose a researcher observes the wage distributions for two different groups. When can we find productivity distributions, statistical experiments that can differ for each group and a wage function that is common across groups (so, in other words, the ingredients of the model) such that these generate the observed wage distributions? Our main result (Theorem 1) shows that a researcher can conclude that two distinct wage distributions can possibly be the result of statistical discrimination if, and only if, neither distribution first-order stochastically dominates the other. Importantly, this result is easy to describe and visualize. Consequently our hope is that non-specialists (such as bureaucrats, journalists and administrators) who frequently use wage gaps as evidence of discrimination will instead consider implementing this dominance test (for instance, by simply plotting the wage distributions).

Before discussing the applied relevance of this result, it is worth situating it in the broader literature on discrimination. One strand of this literature aims to cleanly empirically identify the *presence* of discrimination. Field experiments (of both the audit and correspondence variety) are frequently employed to uncover discrimination because the experimental methodology allows the researcher to fix all other observables and only vary the group characteristic. However, as Bertrand and Duflo (2017) observe in their survey of the literature: "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."

One reason for this is pointed out in Heckman and Siegelman (1993) and Heckman (1998). As an example, consider correspondence studies that send fictitious curricula vitae to employers measure whether or not the candidate gets invited for an interview; a difference in the call back rates by group status is interpreted as evidence of discrimination. Now suppose that the employer believes that the two groups have the same mean productivity but that the variance of the advantaged group is higher (a feature that our model allows). If employers only call back for interviews those candidates whose productivities they think are above a certain threshold, the differential variance can lead to higher call back rate for the advantaged group. Of course, this could also be the result of taste-based discrimination but this cannot be differentiated using this binary outcome.[1]

There is an alternate strand of the literature that aims to devise tests for statistical discrimination using richer outcome data (than our test requires). Motivated by an insight of Becker (1957, 1993), papers in this strand consider settings where the researcher has access not just to the decision (whether or not a loan is granted, a driver is searcher by a police offer etc.) but also the post-decision result (whether or not the loan is repaid, contraband is found on the driver etc.). Analogous data in our setting would correspond to the researcher observing the productivity of the worker in addition to their wage. 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]

[1] Partly motivated by this difficulty, there is a nascent experimental literature that exploits dynamics (see, for instance Bohren, Imas, and Rosenberg, 2019) to tease out the sources of discrimination. The key observation is that dynamics help because beliefs respond to information whereas preferences do not.

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

One way to view our main insight is that we show statistical discrimination can be tested on basic wage data without having access to any further productivity information. The reason we can do so is because we exploit the fact that the outcome variable we study (the wage) is *not binary*.[3] In Section 4.2, we make this point more formally by showing that a similar test is not possible with binary outcomes. Conversely, while outcome tests are not our focus, a strength of our methodology is that it can naturally be extended to devise tests on such richer data that, importantly, do not require the identification of the marginal case. We view this to be a conceptual contribution of our paper that we demonstrate with Theorem 4.

Our tests rest on the assumption that the mean of the true unobserved type distributions of the two group are identical. Of course, there is a large literature on statistical discrimination following Arrow (1973) which studies how the different prior beliefs (or stereotypes) about unobservable characteristics by group affect investments and therefore outcomes. In these "equilibrium" models of discrimination, the disadvantaged group recognizes that prior beliefs are such that unobserved investments are not likely to be rewarded and, therefore, they rationally invest less (so have lower average productivity) than the advantaged group (in turn justifying the discriminatory beliefs). Thus, testing for discrimination based on our model is only appropriate in settings where wage distributions are estimated subject to enough control variables that net out the effect of differentiable investments.

While our result is appealing because of the minimal data requirements and the fact that our model is free of parametric assumptions, some readers might be concerned that first-order stochastic dominance is a strong condition that is unlikely to ever be found in wage data. On the contrary, there is abundant evidence that men's wages are higher than those of women at *all* quantiles of the wage distributions (which is an equivalent way of stating first-order stochastic dominance). See, for instance, Table 1 where entries are the difference between log wages of men and women (positive values imply men's wages are higher). This table is taken from Arulampalam, Booth, and Bryan (2007) (Table 4 in their paper). They analyze data from several European countries and use quantile regressions to show that men earn more than women not just in average (the second column) but at different quantiles of the wage distribution (columns three to seven). Similar evidence of wage gaps across the distribution in Europe is also found by De la Rica, Dolado, and Llorens (2008) and Christofides, Polycarpou, and Vrachimis (2013). Most recently, Maasoumi and Wang (2019) also find this pattern in most years in US data (1976-2013) even after correcting for selection into employment. Our main result says that such wage distributions are precisely the type of distributions that cannot arise from statistical discrimination alone and therefore are evidence of biased employment practices.

Importantly, our test can be taken directly to data (without having to separately estimate wage gaps at dif-

[3]In influential work, Altonji and Pierret (2001) develop a test for statistical discrimination using the properties of wage evolution due to employer learning. Unlike our setting, their model makes several functional form restrictions and their test requires panel data.

Country	Mean	10%	25%	50%	75%	90%
Public Sector						
Austria	0.227	0.191	0.163	0.191	0.221	0.266
Belgium	0.09	0.046	0.05	0.07	0.109	0.169
Britain	0.134	0.091	0.116	0.135	0.144	0.205
Denmark	0.07	0.058	0.051	0.059	0.086	0.136
Finland	0.216	0.115	0.14	0.203	0.269	0.319
France	0.096	0.092	0.077	0.078	0.108	0.167
Germany	0.122	0.111	0.11	0.118	0.14	0.147
Ireland	0.184	0.186	0.169	0.177	0.165	0.181
Italy	0.097	0.041	0.047	0.081	0.138	0.169
Netherlands	0.121	0.039	0.07	0.112	0.16	0.218
Spain	0.083	0.09	0.079	0.095	0.069	0.076
Private Sector						
Austria	0.214	0.182	0.177	0.188	0.21	0.247
Belgium	0.132	0.1	0.121	0.131	0.148	0.185
Britain	0.19	0.155	0.172	0.188	0.213	0.227
Denmark	0.088	0.032	0.065	0.088	0.123	0.161
Finland	0.151	0.068	0.112	0.154	0.188	0.205
France	0.163	0.146	0.126	0.132	0.152	0.19
Germany	0.143	0.088	0.109	0.137	0.166	0.213
Ireland	0.163	0.081	0.143	0.184	0.195	0.206
Italy	0.173	0.148	0.135	0.152	0.179	0.22
Netherlands	0.131	0.059	0.091	0.123	0.168	0.222
Spain	0.181	0.173	0.178	0.184	0.189	0.176

Table 1: Estimated wage gap between men and women using data from 1995-2001 (Arulampalam et al., 2007). All estimates are statistically significant at the 1% level. Models include dummies for whether training was received in the last year, age, education, tenure, marital status, health status, contracts, private sector firm size, any experience of unemployment since 1989, part-time status, fixed term and casual size, region (where possible), year, industry and occupation.

ferent quantiles) because there are well known non-parametric statistical tests of stochastic dominance between two distributions. Our result provides a structural interpretation (in terms of statistical discrimination) to such a test conducted on the wage distributions of an advantaged and disadvantaged group. Recent important econometric developments (but are not limited to) Barrett and Donald (2003), Linton, Maasoumi, and Whang (2005), Linton, Song, and Whang (2010) and Davidson and Duclos (2013). In fact, the aforementioned paper Maasoumi and Wang (2019) conducts precisely such an analysis and concludes that "beyond the early 1990s (except for 2010), men's earnings first-order dominate women's in the majority of the cases to a high degree of statistical confidence." Our framework and result provide a theory-driven interpretation of this result (that Maasoumi and Wang, 2019 do not) as discrimination that is not statistical alone.[4]

[4]Our insight also uncovers a connection between the literature on discrimination and other empirical literatures that apply tests of stochastic dominance (of first and higher orders). Examples are the literature that compares income distributions to infer whether poverty, inequality, or social welfare is greater in one distribution than in another (see Anderson, 1996, Davidson and Duclos, 2000) and the literature on efficient portfolio choice (see Post, 2003, Kuosmanen, 2004).

While we have described our model and results above in terms of the labor market, it is important to stress that our framework (in its current form or natural extensions along the lines of Theorem 4) can capture a variety of distinct settings. For instance, in the criminal justice system, one outcome is the monetary bail amount assigned to different defendants. In this case, the unknown type is the likelihood of fleeing or pre-trial misconduct if released (see Rehavi and Starr, 2014). A different outcome in this context is the sentence duration. Here, the unknown type is the true severity of the crime (for instance, the extent to which a murder/robbery was premeditated). In both cases, the signals are the arguments presented in court. Statistical discrimination can arise because the disadvantaged group may have access to fewer resources which could lead to worse legal representation and a lower ability to navigate the system (see Abrams, Bertrand, and Mullainathan, 2012).

Before proceeding to our model, it is worth acknowledging that, in addition to the papers already cited, 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.

2. THE MODEL

To streamline the exposition, we present the model in the context of discrimination in the labor market. However, as mentioned in the introduction, other applications, such as discrimination in policing or in the justice system, also fit our model.

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 $G_i(w) \in [0,1]$ the fraction of workers in group $i \in \{1,2\}$ being paid a (hourly) wage of $w \geq 0$ or less.[5] The question we address is: are the observed wage distributions consistent with a reduced-form model of statistical discrimination? The model is simple, non-parametric, and general. In a nutshell, the model assumes that workers differ in their productivities, but that there are no significant differences between the two groups; that is, the average productivity is the same in both groups. Employers do not perfectly observe the productivity of workers, acquire some information about it (for instance, through tests, interviews, or referrals), and then pay workers accordingly. We only require that the higher the expected productivity, the higher the wage.

We stress that the only source of discrimination is information. Hiring tools such as personality and

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

aptitude tests or algorithmic resume screeners are all examples of techniques, which may advantage one group over another in signaling their productivity.

We now present the model in detail, starting with the productivity distributions.

Productivity distributions: Workers differ in their productivities, with $\theta_i \in [0,1] =: \Theta$ denoting the productivity of a worker in group $i \in \{1,2\}$.[6] In group i, the (cumulative) productivity distribution is H_i. We assume that $\int_0^1 \theta_1 dH_1(\theta_1) = \int_0^1 \theta_2 dH_2(\theta_2)$. In words, the average productivity is the same in the two groups. A case of particular interest is when the two distributions are identical, i.e., $H_1 = H_2$. As we shall see, there are no differences between a model that assumes identical distributions and another that assumes different distributions, but with identical means. It is, however, easier to test the hypothesis of identical means than identical distributions. Note that we make no additional restrictions, so that we can accommodate discrete distributions, continuous distributions, or mixtures of the two.

Information: Employers do not directly observe the productivity of workers, but receive informative signals. For instance, employers read curricula vitae, interview job applicants, or conduct tests. Employers then form an expectation of the productivity of workers and pay them accordingly: the higher the expected productivity, the higher the wage. Since wages only depend on the expected productivity, it is without loss to restrict attention to *unbiased statistical experiments*.

An unbiased statistical experiment (S_i, π_i) for group $i \in \{1,2\}$ consists of a set of signals $S_i = \Theta$ and a joint distribution π_i over $\Theta \times S_i$, whose marginal distribution over Θ is H_i. Denote the marginal distribution of π_i over S_i as F_i. Moreover (to reflect the "unbiased" terminology), we require that the *posterior estimate* $\mathbb{E}_{\pi_i}[\theta_i \mid s_i]$ of the productivity satisfy

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

for all s_i in the support of F_i; that is, s_i is an unbiased estimate of the true mean $\mathbb{E}_{H_i}[\theta_i]$. This is without loss of generality, as we can always relabel signals to guarantee that they are unbiased in the above sense. Accordingly, we will write θ_i for the posterior estimate (the signal) in what follows.

It is well-known that F_i is a distribution of posterior estimates arising from some statistical experiment if, and only if, the prior distribution H_i is a *mean-preserving spread* of the posterior distribution F_i, which we denote by $H_i \succcurlyeq_2 F_i$. Formally, the mean-preserving spread condition requires that

$$\int_0^\theta H_i(\theta_i)d\theta_i \geq \int_0^\theta F_i(\theta_i)d\theta_i \text{ for all } \theta \in [0,1], \text{ with equality at } \theta = 1.$$

Note that the requirement of equality at $\theta_i = 1$ is the same as ensuring that H_i and F_i have the same

[6] The restriction of productivities to the set $[0,1]$ is a normalization.

mean.[7] If this inequality is strict for any $\theta \in (0,1)$, we say H_i is a *strict mean-preserving spread* of F_i, which we denote by $H_i \succ_2 F_i$.

Wage function: If an employer estimates the productivity of a worker to be θ, the employer pays the worker $W(\theta)$, where the *wage function* $W : [0,1] \to \mathbb{R}_+$ is continuous and strictly increasing. Observe that this wage function does not depend on the group identity and, in this sense, there is no taste-based bias.

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 $\{\theta : W(\theta) \leq w\}$, that is, $G_i(w) = F_i(W^{-1}(w))$ for $w \in [W(0), W(1)]$, $G_i(w) = 0$ for $w < W(0)$ and $G_i(w) = 1$ for $w > W(1)$.[8] Note that, even though the wage function does not depend on group identity, the wage distributions G_1 and G_2 may vary across groups because the distributions of posterior estimates F_1 and F_2 may differ. Moreover, because W is an arbitrary increasing function, G_1 and G_2 may not have the same mean. In other words, the two groups may get different average wages.

Consistency with statistical discrimination: We say that the observed wage distributions G_1 and G_2 are *consistent with statistical discrimination* if there exist prior distributions H_i, distributions of posterior estimates F_i that satisfy $H_i \succcurlyeq_2 F_i$ for $i \in \{1,2\}$, and a continuous and strictly increasing wage function W, such that these jointly induce the observed wage distributions. (We assume throughout that the wage distributions are bounded, i.e., $G_i(\overline{w}) = 1$ for some $\overline{w}$, $i = 1,2$.)

Before stating our main result, we comment on our reduced-form model. While our model is natural and general, in the sense that we allow any statistical experiments and wage functions, we make two key assumptions: (i) the prior distributions for both groups have identical means and (ii) the wages are a function of the posterior estimate alone.

In the introduction, we discussed the first of these. To reiterate, we are implicitly assuming that the wage distributions G_1 and G_2 are estimated controlling for enough observables (and/or with additional corrections for selection) to make identical mean productivity a reasonable assumption. That said, we can in principle allow the productivity distributions to have different means. However, without additional structure, this results in all wage distributions being consistent with statistical discrimination since the model then becomes too general.[9] A natural next step is to impose more structure on the productivity distributions in order to restore non-trivial testable implications. One intuitive candidate is to restrict productivity distributions that are shifted but otherwise identical; that is, $H_i(\theta) = H_j(\theta + \kappa)$ for some

[7] Integration by parts implies that the mean satisfies $\int_0^1 \theta_i dF_i(\theta_i) = \theta_i F_i(\theta_i)|_0^1 - \int_0^1 F_i(\theta_i)d\theta_i = 1 - \int_0^1 F_i(\theta_i)d\theta_i$.
[8] We define W^{-1} as the inverse of W on the domain $[W(0), W(1)]$. None of our results depend on the continuity of W. It would be enough to consider left-continuous and strictly increasing wage functions with generalized inverse $\sup\{\theta : W(\theta) \leq w\}$ at w.
[9] Take any $\overline{w}$ such that $G_1(\overline{w}) = G_2(\overline{w}) = 1$ and define $H_i(\theta) = F_i(\theta) = G_i(\overline{w}\theta)$ for $\theta \in [0,1]$ and $i \in \{1,2\}$.

constant $\kappa \geq 0$ (that is unknown to the researcher), for $i \neq j$ and all θ. It is possible to show that, even in this case, all wage distributions are consistent with statistical discrimination.[10]

We end this section with a brief discussion of the second assumption. Our model is in the spirit of Phelps (1972). He considers two populations, whose productivities are drawn from a normal distribution. Signals are also normally distributed, differ across groups, and the wage function is linear in the posterior mean. If the means of the prior distributions for both groups are the same then, in this model, the expected wage for both groups will be the same because the posterior distribution must have the same mean as the prior. 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 prior 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 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.

Recall that we do not allow wages to depend on the higher moments of F_i. This is for two reasons. First, we allow for any statistical experiments and so any further generality in the wage function might make the testable implications of our model vacuous. Second, a general wage function will lead to additional technical complications in the analysis. This is because it would require us to work with the entire joint distribution π_i as opposed to just the (marginal) distributions F_i of the posterior estimates.

3. THE MAIN CHARACTERIZATION RESULT

Given the generality of our model, the first natural question to ask is: are there *any* wage distributions that are *not* consistent with statistical discrimination? To this point, note that our model allows the posterior estimate distribution of group one to be a strict mean-preserving spread of group two (or $F_1 \succ_2 F_2$ in our notation), in which case a strictly convex wage function W will generate higher mean wages for group one. In other words, differences in mean wages can arise purely via statistical discrimination even though the prior distributions have the same mean. So, to find inconsistent distributions, we need to consider higher moments. In fact, as we now argue, we need to consider *all* moments.

The wage distribution G_i *strictly first-order stochastically dominates* the wage distribution G_j (or that $G_i \succ_1 G_j$) 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

[10] A proof is available upon request.

group j (so $G_i \succ_1 G_j$). We now argue that these distributions are *not* consistent with statistical discrimination. For contradiction, assume that these distributions are consistent. This implies that there exist posterior estimate distributions F_i and F_j, and a wage function W, such that

$$F_i(\theta) = G_i(W(\theta)) \leq G_j(W(\theta)) = F_j(\theta) \qquad \text{for all } \theta \in [0, 1]$$

with the inequality strict for some θ. Note that this implies that $F_i \succ_1 F_j$ (so F_i has a strictly higher mean), which is a contradiction since F_i and F_j are *mean-preserving* contractions of some prior distributions H_i and H_j, which both have the same mean.

The above argument shows that a necessary condition for a pair of wage distributions to be consistent with statistical discrimination is that neither strictly first-order stochastically dominates the other. Our main result shows that this condition is also sufficient. In fact, we show a stronger result, that is, if the wage distributions are consistent with statistical discrimination, then they are consistent with statistical discrimination and identical productivity distributions, that is, $H_1 = H_2$.

THEOREM 1. *The following statements are equivalent.*

(i) The wage distributions G_1 and G_2 are consistent with statistical discrimination.

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

(iii) The wage distributions G_1 and G_2 are consistent with statistical discrimination and identical productivity distributions.

Before presenting a sketch of the proof, it is worth discussing the implication of the third statement. 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." 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, different variances of the productivity distributions have no additional explanatory power.

We now sketch the proof of the statement $[(ii) \implies (iii)]$, with the help of a simple example. (Recall that we have already argued that $[(i) \implies (ii)]$, and $[(iii) \implies (i)]$ is trivially true.) In Table 2, we have two wage distributions G_1 and G_2, with neither first-order stochastically dominating the other (since $G_1(10) > G_2(10)$, while $G_1(15) < G_2(15)$).

The idea of the proof is to construct a wage function W such that the two distributions F_1 and F_2, defined by $F_i(\theta) := G_i(W(\theta))$ for all $\theta, i \in \{1, 2\}$, have the same mean.

Table 2: Sketch of proof: An example

wage/hour	\$10	\$15	\$20
G_1	$1/3$	$5/12$	1
G_2	$1/6$	$1/2$	1

In this simple example, it suffices to find three points $0 \leq \theta^1 < \theta^2 < \theta^3 \leq 1$ (on which F_1 and F_2 are supported) such that $W(\theta^1) = 10$, $W(\theta^2) = 15$, $W(\theta^3) = 20$, and

$$\theta^1 \underbrace{\left(\frac{1}{3} - \frac{1}{6}\right)}_{>0} + \theta^2 \underbrace{\left(\frac{1}{12} - \frac{1}{3}\right)}_{<0} + \theta^3 \underbrace{\left(\frac{7}{12} - \frac{1}{2}\right)}_{>0} = 0,$$

which, in words, ensures that F_1 and F_2 have the same mean. A possible solution is $\theta^1 = 0, \theta^2 = 1/3$ and $\theta^3 = 1$. Notice that a solution exists precisely because neither G_1 nor G_2 strictly first-order dominates the other, which is reflected in the alternating signs in the above expression.

To complete the argument, we need to construct a distribution H such that $H \succcurlyeq_2 F_i, i \in \{1, 2\}$. We now argue that the distribution H, defined by $H(\theta) = 7/18$ for all $\theta < 1$ and $H(1) = 1$, is one such distribution. Note that H is supported on $\{0, 1\}$ with a probability of $11/18$ on $[\theta = 1]$. The mean of H is $11/18$, as is the mean of F_1 and F_2. Moreover,

$$\int_0^\theta H(x)dx - \int_0^\theta F_1(x)dx = \begin{cases} \left(\frac{7}{18} - \frac{1}{3}\right)\theta & \text{if } 0 \leq \theta \leq 1/3 \\ \left(\frac{7}{18} - \frac{1}{3}\right)\frac{1}{3} + \left(\frac{7}{18} - \frac{5}{12}\right)\left(\theta - \frac{1}{3}\right) & \text{if } 1/3 < \theta \leq 1 \end{cases}.$$

It is easy to check that this is indeed positive for all θ, hence $H \succcurlyeq_2 F_1$. Intuitively, to generate F_1 from H, we construct an experiment with three signals, which generate the posterior beliefs about the event $[\theta = 1]$ of 0, $1/3$ and 1 with probability $1/3$, $1/12$, $7/12$, respectively. Since the expectation of the posterior beliefs is equal to the prior belief ($11/18$), such a construction is possible. A similar argument shows that $H \succcurlyeq_2 F_2$.

While the general construction for arbitrary distributions G_1 and G_2 is more elaborate, the same idea works. We first construct W such that $F_1 := G_1 \circ W$ has the same mean as $F_2 := G_2 \circ W$ by transporting "mass" from the region $\{w : G_1(w) \geq G_2(w)\}$ to the region $\{w : G_1(w) < G_2(w)\}$. We then construct a common distribution H as the right derivative of the convex function

$$\theta \mapsto \max\left(\int_0^\theta F_1(x)dx, \int_0^\theta F_2(x)dx\right).$$

The proof is in the Appendix.

4. DISCUSSION

As mentioned earlier, we view Theorem 1 to be the main insight of the paper since, from an applied perspective, rejecting statistical discrimination provides suggestive evidence that differences in group outcomes are the result of bias. In this discussion section, we explore three theoretical extensions of our framework. In Section 4.1, we demonstrate how taste-based discrimination can be modeled within our general reduced-form framework and we characterize wage distributions consistent with taste-based discrimination. In Section 4.2, we consider a slight variation of our model, where the outcome is binary and we show that statistical discrimination has almost no empirical bite. Finally, in Section 4.3, we illustrate the versatility of our approach by revisiting outcome tests à la Becker.

Those readers who are not interested in these theoretical extensions can directly proceed to our concluding remarks (Section 5).

4.1. TASTE-BASED DISCRIMINATION

In this section, we model taste-based discrimination within our framework and we show that taste-based discrimination imposes qualitatively different testable restrictions on observed wage distributions. We begin with our definition.

Wage distributions G_1 and G_2 are *consistent with taste-based discrimination* if there exist productivity distributions H_1 and H_2 (that have identical means) and continuous and strictly increasing wage functions W_1 and W_2 such that these jointly yield the observed wage distributions, i.e., $G_i(w) = H_i(W_i^{-1}(w))$ for all $w \in [W_i(0), W_i(1)]$, $G_i(w) = 0$ for all $w < W_i(0)$ and $G_i(w) = 1$ for all $w > W_i(1)$.

If we furthermore require the two groups to have the same productivity distribution, i.e., $H_1 = H_2$, we say that G_1 and G_2 are *consistent with taste-based discrimination and identical productivity distributions*.

Note the differences of this notion with that of statistical discrimination. We have now removed the noisy signal (the source of statistical discrimination) and instead discrimination is directly introduced via the different wage functions. We do not impose any structure on these wage functions; discrimination is "taste-based" because two workers from different groups with the same expected productivity can be offered different wages.

We need one last piece of notation to present our next result. Let $\underline{w}_i = \inf\{w \in \mathbb{R}_+ \mid G_i(w) > 0\}$ and $\overline{w}_i = \sup\{w \in \mathbb{R}_+ \mid G_i(w) < 1\}$. In words, $[\underline{w}_i, \overline{w}_i]$ is the smallest closed interval that contains the support of the wage distribution G_i. We are now ready to present the theorem characterizing consistency with taste-based discrimination.

THEOREM 2. *(i) Every pair of wage distributions G_1 and G_2 is consistent with taste-based discrimination.*

(ii) Wage distributions G_1 and G_2 are consistent with taste-based discrimination and identical productivity distributions if, and only if, there exists a strictly increasing, continuous bijection $\varphi : [\underline{w}_1, \overline{w}_1] \to [\underline{w}_2, \overline{w}_2]$ such that $G_1(w) = G_2(\varphi(w))$ for all $w \in [\underline{w}_1, \overline{w}_1]$.

The first part of Theorem 2 shows that, if we allow productivity distributions of both groups to differ (while maintaining the assumption of identical means), then all wage distributions could be the result of taste-based discrimination. This is unsurprising since allowing for different wage functions in addition to distinct productivity distributions introduces a lot of freedom into the model. But importantly, this implies that a rejection of the first-order stochastic dominance condition of Theorem 1 can be in fact interpreted as evidence of bias. This interpretation would not always be correct if there existed wage distributions that were consistent with neither statistical nor taste-based discrimination.

This result also implies that all wage distributions can be explained with a combination of statistical and taste-based discrimination even if the productivity distributions are *identical*. This is because the statistical experiments can first introduce heterogeneity into the posterior estimate distributions and then we can apply the part (i) of Theorem 2.

When the productivity distributions of both groups are the same, the empirical content of taste-based discrimination is not vacuous. Part (ii) of Theorem 2 shows that the wage distributions must be related via the monotone transformation φ. The necessity is clear since $G_1(w) = H(W_1^{-1}(w)) = G_2(W_2(W_1^{-1}(w))$ when the two distributions are consistent with tasted-based discrimination and identical productivity distribution, so that $\varphi := W_2 \circ W_1^{-1}$. Moreover, if the two distributions G_1 and G_2 are strictly increasing and continuous, i.e., have no atoms and no null sets, then the existence of φ is guaranteed. We can simply choose $\varphi := G_2^{-1} \circ G_1$. The proof of Theorem 2 is in the Appendix.

This result has two important empirical implications. First, if the range of wages (w_i, w_i') belongs to the same quantile q, i.e., $G_i(w_i) = G_i(w_i') = q$, then there exists a range of wages (w_j, w_j') belonging to the q-th quantile of G_j. In other words, if a range of wages is not observed for group i, a corresponding range is not observed for group j at the *same quantile*. Second, if the distribution G_i has an atom at wage w_i of size $s > 0$, i.e., $s = G_i(w_i) - \lim_{w \uparrow w_i} G_i(w)$, then the distribution G_j has an atom at the wage $w_j = \varphi(w_i)$ of the *same size* s. In other words, atoms of G_i are in bijection with the atoms of G_j: if G_i has an atom of size s, so does G_j, and conversely. Geometrically, the flat parts and the jumps of G_i are in bijection with the flat parts and the jumps of G_j.

As an illustration, consider the two distributions in Table 3. Clearly, G_2 first-order stochastically dominates G_1, so that the two distributions are not consistent with statistical discrimination. Moreover, G_1 and G_2 are neither consistent with taste-based discrimination and identical productivity distributions. Indeed, G_1 has an atom at \$15 of size $5/12$, but G_2 has no atoms of size $5/12$. Intuitively, since the fraction $5/12$ of workers from group 1 are paid \$15 per hour, there must exist some productivity level corresponding to this wage. Moreover, the fraction of workers with that productivity level must be $5/12$. However, since the productivity distribution is identical for group 2, the same fraction of workers from group 2 must appear in the wage distribution for group 2, possibly at a different wage. This is not the case.

Table 3: Inconsistency with statistical and taste-based discrimination

wage/hour	\$10	\$15	\$20	mean
G_1	$1/3$	$3/4$	1	$175/12$
G_2	$1/6$	$1/2$	1	$200/12$

Finally, note that our definition of taste-based discrimination does not imply that one group is systematically advantaged over another, i.e., we didn't impose $W_i \leq W_j$. Group i may be advantaged at low productivities, while group j may be at higher productivities. Imposing the restriction $W_1 \leq W_2$ would translate into the bijection $\varphi : [\underline{w}_1, \overline{w}_1] \to [\underline{w}_2, \overline{w}_2]$ satisfying $\varphi(w) \geq w$ for all $w \in [\underline{w}_1, \overline{w}_1]$. In words, at all quantile q, workers from group 2 are paid more than workers from group 1.

4.2. BINARY OUTCOMES

In the introduction, we noted that research studies frequently rely on the differences in *binary* outcomes to document discrimination. For instance, call-back rates from correspondence studies are often used to document discrimination in the labor market. Other instances include mortgage approval rates, credit card approval rates, job promotion and university admission. We stated that statistical discrimination has little bite in such binary setting. We now formalize this statement in the context of our model. Throughout, we use the same notation as in the previous section, but replace the term "wage" with the term "outcome."

There are two outcomes, labelled $w = 0$ and $w = 1$. The distribution G_i is thus a binary distribution, with $G_i(0)$ the probability of outcome $w = 0$. We say that the binary distributions G_1 and G_2 are consistent with statistical discrimination if there exist prior distributions H_1 and H_2, distributions of posterior estimates F_1 and F_2 that satisfy $H_i \succcurlyeq_2 F_i$ and a cutoff $\underline{\theta}$ such that $F_i(\underline{\theta}) = G_i(0)$ for $i \in \{1, 2\}$.

In words, the only difference between this binary outcome setting and the model in Section 2 is that instead of a strictly increasing wage function, there is a group-independent cutoff $\underline{\theta} \in [0, 1]$ such that outcome $w = 1$ occurs only if the posterior estimate is strictly above it. In the context of the labor market, this says that an employer calls back a job candidate only if the candidate's expected productivity is sufficiently high.

The next result characterizes binary outcome distributions that are consistent with statistical discrimination.

THEOREM 3. *Binary outcome distributions G_1, G_2 are consistent with statistical discrimination if, and only if, it is not the case that $G_i(0) = 1$ and $G_j(0) = 0$ for $i \neq j$, $i, j \in \{1, 2\}$.*

In words, this simply says that the outcomes for the two groups are 0 and 1 respectively, with probability 1. Again in the context of the labor market, this says that either all job candidates from group 1 are called back and none from group 2 are, or vice versa. Needless to say, such extreme discrimination is seldom observed and therefore, in practice, statistical discrimination cannot be disentangled from bias in a setting with binary outcomes.

We end the section by providing a simple argument for this result. By contradiction, suppose that distributions $G_1(0) = 1$ and $G_2(0) = 0$ are consistent with statistical discrimination. Then, the mean of F_1 must be less than or equal to $\underline{\theta}$ (since $F_1(\underline{\theta}) = 1$), whereas the mean of F_2 must be strictly greater (since $F_2(\underline{\theta}) = 0$). This is, of course, not possible since both distributions must have the same mean. A symmetric argument applies when $G_1(0) = 0$ and $G_2(0) = 1$.

Conversely, suppose that $0 < G_1(0) < G_2(0) < 1$. (It is easy to adapt the arguments to treat the other cases.) Let F_1 have binary support $\{0, 1\}$ and assign probability $G_1(0)$ to 0 and therefore $1 - G_1(0)$ to 1. Note that F_1 has mean $1 \times (1 - G_1(0)) + 0 \times G(0) = 1 - G_1(0)$. Let F_2 have binary support $\left\{1 - G_1(0) - \frac{\varepsilon}{G_2(0)}, 1 - G_1(0) + \frac{\varepsilon}{1-G_2(0)}\right\}$ where $\varepsilon > 0$ is sufficiently small to ensure both points of the support lie in [0,1]. Assign probability $G_2(0)$ to $1 - G_1(0) - \frac{\varepsilon}{G_2(0)}$ and therefore $1 - G_2(0)$ to $1 - G_1(0) + \frac{\varepsilon}{1-G_2(0)}$. Note that F_2 has mean $G_2(0) \times \left(1 - G_1(0) - \frac{\varepsilon}{G_2(0)}\right) + (1 - G_2(0)) \times \left(1 - G_1(0) + \frac{\varepsilon}{1-G_2(0)}\right) = 1 - G_1(0)$. Thus, the two distributions F_1 and F_2 have the same mean. The second step in the proof of Theorem 1 then shows how to construct a prior H such that $H \succcurlyeq_2 F_i$, as required. The argument is completed by setting the threshold $\underline{\theta} = 1 - G_1(0)$.

4.3. OUTCOME TESTS

The methodology we have introduced is versatile enough to study discrimination in a wide range of settings. As a "proof of concept," we now present one such application — the Becker outcome test — and hope to examine others in future work. To ease the presentation, we frame the application in the context of bail decisions and closely follow Arnold, Dobbie, and Hull (2022). See also Arnold, Dobbie, and Yang (2018), Canay, Mogstad, and Mountjoy (2020), Hull (2021) and Simoiu, Corbett-Davies, and Goel (2017). Several other contexts such as police search decisions and loan decisions also fit the model.

In the context of bail decisions, judges have to decide whether or not to release defendants prior to trials, with $w = 1$ denoting the decision to release a defendant. Upon being released, the defendant may subsequently fail to appear in court or commit another crime, which we model with a binary variable $Y \in \{0, 1\}$; $Y = 1$ indicates a pre-trial misconduct. We–the analyst–observe the fraction $r_i \in (0, 1)$ of defendants from group i released by a judge, and $q_i \in (0, 1)$ the fraction of the released defendants, who committed a pre-trial misconduct.

Unlike the analysis in Section 4.2 where only a single piece of information was observed, the analyst now observes two pieces of information: the release decision along with the post-decision result, that is, if the released defendant committed pre-trial misconduct. Now suppose that the judge releases a defendant if, and only if, the information she possesses signals that the likelihood of pre-trial misconduct is smaller than a given threshold. The Becker outcome test is based on the observation that if the cutoff is the same across groups, the rate of misconduct of the *marginal* defendant of each group should be the same (and equal to the cutoff). Of course, the problem with implementing this test in practice is that the analyst does not observe the identity of the marginal defendant.

So instead, is it possible to detect bias using the averages (r_1, q_1) and (r_2, q_2)? Since the shapes of the signal distributions may differ, the release rates r_1 and r_2 may not be the same even if the judge uses the identical cutoff for both groups. This is the issue with "benchmarking tests" that we discussed in Section 4.2. What if we consider the rates q_1 and q_2 at which misconduct occurs conditional on being released on bail? These too can depend on the shape of the signal distributions and a higher rate of misconduct for either group is possible even if the judge uses a group-neutral cutoff. This is a well known problem with such "outcome tests" (often referred to as the infra-marginality problem). In the remainder of this section, we show that it is possible to derive a test for statistical discrimination that depends jointly on both r_i and q_i provided we assume that, were all defendants to be released, both groups would commit pre-trial misconduct at the same average rate.

We assume that Y_i is distributed with (unknown) probability $\theta_i \in [0, 1]$ in group i, with H_i the prior distribution of θ_i. Thus, the ex-ante probability of pre-trial misconduct is $\mathbb{E}_{H_i}[\theta_i]$. Prior to deciding whether to bail a defendant, the judge obtains some information about the likelihood of pre-trial misconduct, and grants the bail only if the perceived probability of misconduct is smaller than the group-independent threshold $\underline{\theta}$. As in previous sections, we assume that the judge receives an unbiased signal about θ_i, with F_i being the distribution of the signal. The distribution F_i is a mean-preserving contraction of H_i. Therefore, the release rate in group i is $F_i(\underline{\theta})$, while the pre-trial misconduct rate conditional on release is $\mathbb{E}_{F_i}[\theta_i | \theta_i \leq \underline{\theta}]$.

Analogous to our previous definitions, we say that the outcomes (r_1, q_1) and (r_2, q_2) are consistent with statistical discrimination if there exist prior distributions H_1 and H_2, posterior distributions F_1 and F_2, and a threshold $\underline{\theta}$ such that (i) $\mathbb{E}_{H_1}[\theta_1] = \mathbb{E}_{H_2}[\theta_2]$, (ii) $H_i \succcurlyeq_2 F_i$, and (iii) $r_i = F_i(\underline{\theta})$ and $q_i =$

$\mathbb{E}_{F_i}[\theta_i|\theta_i \leq \underline{\theta}]$ for $i \in \{1, 2\}$. With this definition in hand, we have the following characterization.

THEOREM 4. *The outcomes* (r_1, q_1) *and* (r_2, q_2) *are consistent with statistical discrimination if, and only if,* $q_1 < r_2 q_2 + (1 - r_2)1$ *and* $q_2 < r_1 q_1 + (1 - r_1)1$.

The above result shows that, under the assumption that both groups commit pre-trial misconduct at the same rate on average, we can precisely identify the conditions under which the outcomes could have arisen from statistical discrimination. As with the rest of this paper, this result requires no further assumptions on the prior distributions or the signals. It is fully non-parametric and easy to implement. It is worth noting that the aforementioned Simoiu, Corbett-Davies, and Goel (2017) take a different approach. They estimate a parametric model (all distributions lie in certain families parametrized by variables that they estimate), but allow the means of the prior distributions to differ.

5. CONCLUDING REMARKS

In this paper, we introduced a new non-parametric methodology to test whether two distinct wage distributions could have been generated by statistical discrimination alone. A rejection of our test—that one wage distribution first-order stochastically dominates the other—implies that the differences in wage distributions are possibly the result of bias. Our model is a significantly generalized version of Phelps (1972) and Aigner and Cain (1977): we only require the unobserved productivity distributions to have identical means, the signals from which employers get information about the workers' productivities are unrestricted and wages can be determined by any continuous, strictly increasing function of the posterior productivity estimate. Because our main assumption is that the productivity distributions have identical means, the wage distributions on which our test is conducted should be estimated with enough control variables to make this assumption realistic.

To the best of our knowledge, this paper is the first to analyze the problem of testing for discrimination without resorting to any functional form assumptions and we view this to be one of our main contributions. In casual discussions of wage gaps, the mean wages of two groups are typically compared and differences are often interpreted (without justification) as taste-based discrimination. Our test is microfounded and uses the entire wage distribution but importantly, is just as easy to visualize and implement by non-specialists.

Finally, we hope one consequence of this paper is that the methodology is extended to study discrimination in other settings, possibly with richer data. We have already sketched some possible extensions such as the Becker outcome test, and hope to study more in future works.

A. PROOFS

We start with a preliminary remark. The distribution G_i is supported on a closed subset of $[\underline{w}_i, \overline{w}_i]$, where $0 \leq \underline{w}_i = \inf\{w \in \mathbb{R}_+ \mid G_i(w) > 0\}$ and $\overline{w}_i = \sup\{w \in \mathbb{R}_+ \mid G_i(w) < 1\} < \infty$ (since the wage distributions have bounded support). Let W_i be the random variable with distribution G_i. The random variable $W_i' := \frac{1}{\overline{w}_i} W_i$ is then supported on a closed subset of $[0, 1]$ with distribution G_i', where $G_i'(w) = G_i(\overline{w}_i w)$. Hence, we can assume without loss of generality that the wage distributions are supported on a subset of $[0, 1]$. To ease notation, we will do so throughout the proofs.

PROOF OF THEOREM 1. We have already argued that the first statement implies the second statement. To prove the theorem, we only need to show that the second statement implies the third (since the third obviously implies the first).

The proof consists of two steps. In the first, we show that there exists a strictly increasing and continuous wage function W such that the two distributions defined by $F_i(\theta) = G_i(W(\theta))$ for $i \in \{1, 2\}$ have the same mean. In the second step, we show that for any two distributions with a common mean, there exists a common productivity distribution H such that $H \succcurlyeq_2 F_i$ for both $i \in \{1, 2\}$.

Step 1: There exists a strictly increasing and continuous function $W : [0, 1] \to [0, 1]$ such that the two distributions defined by $F_i(\theta) = G_i(W(\theta))$ for $\theta \in [0, 1]$ and $i \in \{1, 2\}$ satisfy $\int_0^1 F_1(\theta)d\theta = \int_0^1 F_2(\theta)d\theta$.

First, observe that if

$$\int_0^1 G_1(w)dw = \int_0^1 G_2(w)dw,$$

then $W(\theta) = \theta$ is trivially the requisite function.

So, without loss, suppose that

$$\int_0^1 G_1(w)dw > \int_0^1 G_2(w)dw.$$

(A symmetric argument applies if we interchange 1 and 2.) Define the function

$$\Delta_G(w) = G_1(w) - G_2(w)$$

and note the above inequality is simply $\int_0^1 \Delta_G(w)dw > 0$.

Since G_2 does not strictly first-order stochastically dominate G_1, there exists a non-empty interval $[\underline{w}, \overline{w}] \subset (0, 1)$ such that $\int_{\underline{w}}^{\overline{w}} \Delta_G(w)dw < 0$. (This follows from the right-continuity of G_1 and G_2.) Therefore, there must exist strictly positive constants $\gamma^+ > 0$ and $\gamma^- > 0$ such that

$$\frac{1}{\gamma^+}\int_0^{\underline{w}} \Delta_G(w)dw + \frac{1}{\gamma^-}\int_{\underline{w}}^{\overline{w}} \Delta_G(w)dw + \frac{1}{\gamma^+}\int_{\overline{w}}^{1} \Delta_G(w)dw = 0.$$

Define

$$\kappa = \frac{\overline{w} - \underline{w}(1-(\gamma^-/\gamma^+))}{\gamma^-} + \frac{1-\overline{w}}{\gamma^+} = \frac{\gamma^+(\overline{w}-\underline{w}) + \gamma^-(1-(\overline{w}-\underline{w}))}{\gamma^+\gamma^-} > 0.$$

Note the inequality follows from the fact that $\overline{w} > \underline{w}$, $\overline{w} - \underline{w} < 1$, $\gamma^- > 0$ and $\gamma^+ > 0$.

Using this κ, we define

$$\underline{\theta} = \frac{\underline{w}}{\kappa\gamma^+} > 0,$$

and

$$\overline{\theta} = \frac{\overline{w} - \underline{w}(1-(\gamma^-/\gamma^+))}{\kappa\gamma^-} = \frac{\overline{w} - \underline{w}}{\kappa\gamma^-} + \underline{\theta} > \underline{\theta}.$$

Also note that

$$\overline{\theta} < 1,$$

since

$$\kappa\gamma^- > \overline{w} - \underline{w}(1-(\gamma^-/\gamma^+)).$$

Now consider the following piecewise linear wage function

$$W(\theta) = \begin{cases} \kappa\gamma^+\theta & \text{if } 0 \leq \theta < \underline{\theta}, \\ \kappa\gamma^-\theta + \underline{w}(1-(\gamma^-/\gamma^+)) & \text{if } \underline{\theta} \leq \theta \leq \overline{\theta}, \\ \kappa\gamma^+\theta + 1 - \kappa\gamma^+ & \text{if } \overline{\theta} < \theta \leq 1. \end{cases}$$

Few observations are worth making. First, W is continuous because

$$\lim_{\theta \uparrow \underline{\theta}} W(\theta) = \kappa\gamma^+\underline{\theta} = \underline{w} = \kappa\gamma^-\underline{\theta} + \underline{w}(1-(\gamma^-/\gamma^+)) = W(\underline{\theta}),$$

$$W(\overline{\theta}) = \kappa\gamma^-\overline{\theta} + \underline{w}(1-\gamma^-/\gamma^+) = \kappa\gamma^-\frac{\overline{w} - \underline{w}(1-\gamma^-/\gamma^+)}{\kappa\gamma^-} + \underline{w}(1-\gamma^-/\gamma^+) = \overline{w}$$

and

$$\lim_{\theta \downarrow \overline{\theta}} W(\theta) = \kappa\gamma^+\overline{\theta} + 1 - \kappa\gamma^+$$

$$= \kappa\gamma^+ \frac{\overline{w} - \underline{w}(1-\gamma^-/\gamma^+)}{\kappa\gamma^-} + 1 - \kappa\gamma^+$$
$$= \gamma^+ \frac{\overline{w} - \underline{w}(1-\gamma^-/\gamma^+)}{\gamma^-} + 1 - \gamma^+ \left(\frac{\overline{w} - \underline{w}(1-\gamma^-/\gamma^+)}{\gamma^-} + \frac{1-\overline{w}}{\gamma^+} \right) = \overline{w}.$$

Second, $W(0) = 0$ and $W(1) = 1$. To summarize, the piecewise linear wage function $W : [0,1] \to [0,1]$ is bijective, strictly increasing and continuous.

Using the constructed W, define $F_i(\theta) = G_i(W(\theta))$ for $i \in \{1,2\}$. Define $\Delta_F(\theta) = F_1(\theta) - F_2(\theta)$.

Finally, observe that

$$\begin{aligned}
\int_0^1 \Delta_F(\theta)d\theta &= \int_0^{\underline{\theta}} \Delta_F(\theta)d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \Delta_F(\theta)d\theta + \int_{\overline{\theta}}^1 \Delta_F(\theta)d\theta \\
&= \int_0^{\underline{\theta}} \Delta_G(W(\theta))d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \Delta_G(W(\theta))d\theta + \int_{\overline{\theta}}^1 \Delta_G(W(\theta))d\theta \\
&= \frac{1}{\kappa\gamma^+} \int_0^{\underline{w}} \Delta_G(w)dw + \frac{1}{\kappa\gamma^-} \int_{\underline{w}}^{\overline{w}} \Delta_G(w)dw + \frac{1}{\kappa\gamma^+} \int_{\overline{w}}^1 \Delta_G(w)dw \\
&= 0,
\end{aligned}$$

where the second last equality follows from the change of variables from θ to w. Therefore, the constructed distributions F_1 and F_2 have the same mean as required, which completes the proof of this step.

Step 2: Suppose $\int_0^1 F_1(\theta)d\theta = \int_0^1 F_2(\theta)d\theta$. Then, there exists a prior distribution H such that

$$\int_0^\theta H(x)dx \geq \max\left\{ \int_0^\theta F_1(x)dx,\ \int_0^\theta F_2(x)dx \right\} \textit{ with equality at } \theta = 1.$$

Define the function

$$M(\theta) = \max\left\{ \int_0^\theta F_1(x)dx,\ \int_0^\theta F_2(x)dx \right\}.$$

Observe that M is an increasing, convex function since each $\int_0^\theta F_i(x)dx$ is increasing and convex (because F_i is increasing). Also note that

$$M(1) = \int_0^1 F_1(x)dx = \int_0^1 F_2(x)dx.$$

Let H be the right derivative of M (the right derivative always exists and, moreover, $M(\theta) = M(0) + \int_0^\theta H(x)dx$ since M is convex, hence absolutely continuous). This function is increasing, satisfies $M(0) = 0$, $M(1) = 1$ and is, therefore, the requisite prior distribution. (Recall the the right derivative of a convex function is right continuous and has limits on the left.) This completes the proof. ■

PROOF OF THEOREM 2. (i) Without loss, suppose the mean wage of group 1 is the highest; that is, there exists $\alpha \geq 1$ such that $\int_0^1 wdG_1(w) = \alpha \int_0^1 wdG_2(w)$.

Construct the wage functions W_1 and W_2 as:

$$W_1(\theta_1) = \alpha\,\theta_1 \text{ and } W_2(\theta_2) = \theta_2.$$

Construct the corresponding prior distributions H_1 and H_2 as follows:

$$H_i(\theta_i) = G_i(W_i(\theta_i)) \text{ for } i \in \{1, 2\}.$$

Since $G_1(1) = 1$, $H_1(1/\alpha) = 1$ and, therefore, H_1 is supported on a subset of $[0, 1]$. Similarly, H_2 is supported on $[0, 1]$ since G_2 is.

It remains to verify that the distributions H_1 and H_2 have the same mean. We have:

$$\int_0^1 \theta_1 dH_1(\theta_1) = \frac{1}{\alpha}\int_0^\alpha w_1 dG_1(w_1) = \frac{1}{\alpha}\int_0^1 w_1 dG_1(w_1) = \int_0^1 w_2 dG_2(w_2) = \int_0^1 \theta_2 dH_2(\theta_2).$$

(ii) We begin with the if direction. Assume that there exists a strictly increasing continuous bijection $\varphi : [\underline{w}_1, \overline{w}_1] \to [\underline{w}_2, \overline{w}_2]$ such that $G_1(w) = G_2(\varphi(w))$. For future reference, note that we must have $\varphi(\underline{w}_1) = \underline{w}_2$ and $\varphi(\overline{w}_1) = \overline{w}_2$.

We define the wage functions W_1 and W_2 and the productivity distribution H as follows: for all $\theta \in [0, 1]$,

$$W_1(\theta) = (\overline{w}_1 - \underline{w}_1)\theta + \underline{w}_1, \quad H(\theta) = G_1(W_1(\theta)) \quad \text{and } W_2(\theta) = \varphi((\overline{w}_1 - \underline{w}_1)\theta + \underline{w}_1).$$

Notice that for all $w \in [W_1(0), W_1(1)]$, $W_1^{-1}(w) = \frac{w - \underline{w}_1}{\overline{w}_1 - \underline{w}_1}$, hence $G_1(w) = H(W_1^{-1}(w))$. Moreover, for all $w < W_1(0)$, $G_1(w) = 0$ since $W_1(0) = \underline{w}_1$. Similarly, for all $w > W_1(1)$, $G_1(w) = 1$ since $W_1(0) = \overline{w}_1$.

Similarly, for all $w \in [W_2(0), W_2(1)]$, $W_2^{-1}(w) = \frac{\varphi^{-1}(w) - \underline{w}_1}{\overline{w}_1 - \underline{w}_1}$. Therefore,

$$G_2(w) = G_1(\varphi^{-1}(w)) = H\left(\frac{\varphi^{-1}(w) - \underline{w}_1}{\overline{w}_1 - \underline{w}_1}\right) = H(W_2^{-1}(w)).$$

Moreover, for all $w < W_2(0)$, $G_2(w) = 0$ since $W_2(0) = \varphi(\underline{w}_1) = \underline{w}_2$. Similarly, for all $w > W_2(1)$, $G_1(w) = 1$ since $W_2(1) = \varphi(\overline{w}_1) = \overline{w}_2$.

For the only if direction, we can construct the φ function directly from the definition. For any $w \in [W_1(0), W_1(1)]$, observe that

$$G_2(W_2(W_1^{-1}(w))) = H(W_1^{-1}(w))) = G_1(W_1(W_1^{-1}(w)))) = G_1(w).$$

Since, $[\underline{w}_1, \overline{w}_1] \subseteq [W_1(0), W_1(1)]$, the function $\varphi : [\underline{w}_1, \overline{w}_1] \to [W_2(0), W_1(0)]$ defined by $\varphi(w) = W_2(W_1^{-1}(w))$ is an injection (since W_1 and W_2 are strictly increasing and continuous). Moreover, for any decreasing sequence in $[W_1(0), W_1(1)]$ converging to $\underline{w}_1$, $G_1(w) = G_2(\varphi(w)) > 0$, hence $\lim_{w \downarrow \underline{w}_1} \varphi(w) = \varphi(\underline{w}_1) \geq \underline{w}_2$. Similarly, $\varphi(\overline{w}_1) \leq \overline{w}_2$.

Interchanging the indices, for any $w \in [W_2(0), W_2(1)]$, we have that

$$G_1(W_1(W_2^{-1}(w))) = H(W_2^{-1}(w))) = G_2(W_2(W_2^{-1}(w)))) = G_2(w).$$

Since, $[\underline{w}_2, \overline{w}_2] \subseteq [W_2(0), W_2(1)]$, the function $\varphi^{-1} : [\underline{w}_2, \overline{w}_2] \to [W_1(0), W_2(0)]$ defined by $\varphi^{-1}(w) = W_1(W_2^{-1}(w))$ is an injection (since W_1 and W_2 are strictly increasing and continuous). Moreover, $\varphi^{-1}(\underline{w}_2) \geq \underline{w}_1$ and $\varphi^{-1}(\overline{w}_2) \leq \overline{w}_1$.

Therefore, φ is a bijection from $[\underline{w}_1, \overline{w}_1]$ to $[\underline{w}_2, \overline{w}_2]$ with the property that $\varphi(\underline{w}_1) = \underline{w}_2$ and $\varphi(\overline{w}_1) = \overline{w}_2$. The function φ is clearly strictly increasing and continuous.

■

PROOF OF THEOREM 4. (Only if.) Suppose that the outcomes are consistent with statistical discrimination. Since $\mathbb{E}_{F_i}[\theta_i | \theta_i \leq \underline{\theta}] = q_i$, we must have $\underline{\theta} \geq q_i$. Therefore, the mean of F_i must be at least q_i since

$$\mathbb{E}_{F_i}[\theta_i] = F_i(\overline{\theta})\mathbb{E}_{F_i}[\theta_i | \theta_i \leq \underline{\theta}] + (1 - F_i(\underline{\theta}))\mathbb{E}_{F_i}[\theta_i | \theta_i > \underline{\theta}] > r_i q_i + (1 - r_i)\underline{\theta}.$$

Similarly, the mean of F_j, $j \neq i$, is at most $r_j q_j + (1 - r_j)1$. Finally, since F_i and F_j have the same mean, it must be the case that

$$q_i < r_j q_j + (1 - r_j)1,$$

for all (i, j), $j \neq i$.

(If.) The proof is constructive. Without loss of generality, assume that $q_i \geq q_j$. It follows that $r_i q_i + (1 - r_i)1 > q_j$ is automatically satisfied (since $r_i < 1$ and $q_i < 1$). Assume that $r_j q_j + (1 - r_j)1 > q_i$ is also satisfied.

Let $\underline{\theta} = q_i + \delta$ for some $\delta > 0$, and F_i a binary distribution which takes values q_i and $q_i + \varepsilon$ with probability r_i and $1 - r_i$, respectively, where $\varepsilon > \delta$. By construction, the mean of F_i is $q_i + (1 - r_i)\varepsilon$ and $F_i(\underline{\theta}) = r_i$.

We now construct F_j such that its mean is the same as the mean of F_i. The second step in the proof of Theorem 1 then shows how to construct a prior H such that $H \succcurlyeq_2 F_i$ and $H \succcurlyeq_2 F_j$. The distribution F_j is again binary and takes values q_j and $\frac{q_i - r_j q_j + (1-r_i)\varepsilon}{1-r_j}$ with probability r_j and $(1 - r_j)$, respectively. The mean of F_j is:

$$r_j q_j + (1 - r_j)\frac{q_i - r_j q_j + (1 - r_i)\varepsilon}{1 - r_j} = q_i + (1 - r_i)\varepsilon,$$

the same as the mean of F_i.

Finally, we need to choose δ and ε, with $\varepsilon > \delta$, such that (i) $q_i + \varepsilon \leq 1$, (ii) $\frac{q_i - r_j q_j + (1-r_i)\varepsilon}{1-r_j} \leq 1$, and (iii) $\frac{q_i - r_j q_j + (1-r_i)\varepsilon}{1-r_j} > q_i + \delta$. The conditions (i) and (ii) guarantee that F_i and F_j are supported on a subset of $[0, 1]$, while condition (iii) guarantees that $F_j(\underline{\theta}) = r_j$. Since $q_i \geq q_j$ and $\frac{q_i - r_j q_j}{1-r_j} < 1$, it is routine to verify that we can indeed choose ε and δ as required.

■

REFERENCES

Abrams, D. S., M. Bertrand, and S. Mullainathan (2012): "Do judges vary in their treatment of race?" *Journal of Legal Studies*, 41, 347–383.

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

Anderson, G. (1996): "Nonparametric tests of stochastic dominance in income distributions," *Econometrica*, 1183–1193.

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 P. Hull (2022): "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, 112, 2992–3038.

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

Arrow, K. J. (1973): "The theory of discrimination," in *Discrimination in Labor Markets*, Orley Ashenfelter and Albert Rees, eds, 3–33.

Arulampalam, W., A. L. Booth, and M. L. Bryan (2007): "Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution," *ILR Review*, 60, 163–186.

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.

Bohren, J. A., A. Imas, and M. Rosenberg (2019): "The dynamics of discrimination: Theory and evidence," *American Economic Review*, 109, 3395–3436.

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.

Christofides, L. N., A. Polycarpou, and K. Vrachimis (2013): "Gender wage gaps, 'sticky floors' and 'glass ceilings' in Europe," *Labour Economics*, 21, 86–102.

Davidson, R. and J.-Y. Duclos (2000): "Statistical inference for stochastic dominance and for the measurement of poverty and inequality," *Econometrica*, 68, 1435–1464.

——— (2013): "Testing for restricted stochastic dominance," *Econometric Reviews*, 32, 84–125.

De la Rica, S., J. J. Dolado, and V. Llorens (2008): "Ceilings or floors? Gender wage gaps by education in Spain," *Journal of Population Economics*, 21, 751–776.

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

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

Heckman, J. J. and P. Siegelman (1993): "The Urban Institute audit studies: Their methods and findings," *Clear and convincing evidence: Measurement of discrimination in America*, 187–258.

Hull, P. (2021): "What Marginal Outcome Tests Can Tell Us About Racially Biased Decision-Making," Working Paper 28503, National Bureau of Economic Research.

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

Kuosmanen, T. (2004): "Efficient diversification according to stochastic dominance criteria," *Management Science*, 50, 1390–1406.

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.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005): "Consistent testing for stochastic dominance under general sampling schemes," *Review of Economic Studies*, 72, 735–765.

Linton, O., K. Song, and Y.-J. Whang (2010): "An improved bootstrap test of stochastic dominance," *Journal of Econometrics*, 154, 186–202.

Maasoumi, E. and L. Wang (2019): "The gender gap between earnings distributions," *Journal of Political Economy*, 127, 2438–2504.

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.

Post, T. (2003): "Empirical tests for stochastic dominance efficiency," *Journal of Finance*, 58, 1905–1931.

Rehavi, M. M. and S. B. Starr (2014): "Racial disparity in federal criminal sentences," *Journal of Political Economy*, 122, 1320–1354.

Simoiu, C., S. Corbett-Davies, and S. Goel (2017): "The problem of infra-marginality in outcome tests for discrimination," *Annals of Applied Statistics*, 11, 1193–1216.