WHICH WAGE DISTRIBUTIONS ARE CONSISTENT WITH
STATISTICAL DISCRIMINATION?

Rahul Deb
Ludovic Renou

October 14, 2022

ABSTRACT
We derive a non-parametric test for statistical discrimination that can be applied to cross-sectional wage
data. Specifically, we show that the wage distributions for two groups (with identical observable character-
istics) are consistent with a general reduced-form model of statistical discrimination if, and only if,
neither wage distribution first-order stochastically dominates the other. Our model allows us to interpret
a rejection of this condition as 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 and to her, Prashant
Bharadwaj, Aislinn Bohren, Alex Imas and Rob McMillan for detailed discussions. We would also like to thank Juan Eberhard,
Namrata Kala, Kory Kroft, Aditya Kuvalkar, Mallesh Pai and 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

Observationally identical members from disadvantaged and advantaged groups (for instance, women and men, blacks and whites) are frequently treated differently in labor markets, policing and judicial decisions, housing markets and many other settings. Such discrimination is typically categorized as either “statistical” (outcomes differ because of imperfect information) or “taste-based” (bias or animus towards one group drives outcome differences), and establishing which of these is at work is important both for accountability and to devise corrective policies.

There has been considerable progress 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 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.” Studies that aim to uncover the source of discrimination rely on richer data. One common approach is the “outcome test” (that builds on an insight of Becker, 1957, 1993) which requires the researcher’s data to contain not just to the decision (whether or not a loan is granted, a police officer searches a driver, etc), but also the post-decision result (whether or not the loan is repaid, contraband is found on the driver, etc). Yet, even with these richer data, it is still challenging to tease apart statistical from taste-based discrimination (see Canay, Mogstad, and Mountjoy, 2020).

In this paper, we focus on an important labor market outcome, the differences in which have been, and continue to be, the subject of much public debate: wages. We propose a general reduced-form model and derive a non-parametric test for statistical discrimination that only requires basic administrative wage data. Specifically, we show that two wage distributions are consistent with statistical discrimination if, and only if, neither wage distribution first-order stochastically dominates the other. Through the lens of our model, we can interpret a rejection of this condition as micro-founded evidence that wage differences are the result of bias or animus. Importantly, this condition can be visualized by simply plotting the cumulative wage distributions. We view this feature to be important as it makes our test accessible to non-experts (like administrators, journalists and policy makers), who often incorrectly interpret a wage-gap (which can be the result of statistical discrimination) to be evidence of bias.

Our simple but general reduced-form model of statistical discrimination is 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 micro-found 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, we pose and answer the following question. Suppose a researcher observes the wage distributions for two different groups (appropriately controlling for observables). 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? As mentioned above, 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.

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 different 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.
<table>
<thead>
<tr>
<th>Country</th>
<th>Mean 10%</th>
<th>Mean 25%</th>
<th>Mean 50%</th>
<th>Mean 75%</th>
<th>Mean 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Public Sector</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>0.227</td>
<td>0.191</td>
<td>0.163</td>
<td>0.191</td>
<td>0.221</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.09</td>
<td>0.046</td>
<td>0.05</td>
<td>0.07</td>
<td>0.109</td>
</tr>
<tr>
<td>Britain</td>
<td>0.134</td>
<td>0.091</td>
<td>0.116</td>
<td>0.135</td>
<td>0.144</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.07</td>
<td>0.058</td>
<td>0.051</td>
<td>0.059</td>
<td>0.086</td>
</tr>
<tr>
<td>Finland</td>
<td>0.216</td>
<td>0.115</td>
<td>0.14</td>
<td>0.203</td>
<td>0.269</td>
</tr>
<tr>
<td>France</td>
<td>0.096</td>
<td>0.092</td>
<td>0.077</td>
<td>0.078</td>
<td>0.108</td>
</tr>
<tr>
<td>Germany</td>
<td>0.122</td>
<td>0.111</td>
<td>0.11</td>
<td>0.118</td>
<td>0.14</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.184</td>
<td>0.186</td>
<td>0.169</td>
<td>0.177</td>
<td>0.165</td>
</tr>
<tr>
<td>Italy</td>
<td>0.097</td>
<td>0.041</td>
<td>0.047</td>
<td>0.081</td>
<td>0.138</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.121</td>
<td>0.059</td>
<td>0.07</td>
<td>0.112</td>
<td>0.16</td>
</tr>
<tr>
<td>Spain</td>
<td>0.083</td>
<td>0.09</td>
<td>0.079</td>
<td>0.095</td>
<td>0.069</td>
</tr>
<tr>
<td><strong>Private Sector</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.214</td>
<td>0.182</td>
<td>0.177</td>
<td>0.188</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.132</td>
<td>0.1</td>
<td>0.121</td>
<td>0.131</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>0.155</td>
<td>0.172</td>
<td>0.188</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td>0.088</td>
<td>0.032</td>
<td>0.065</td>
<td>0.088</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>0.151</td>
<td>0.068</td>
<td>0.112</td>
<td>0.154</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>0.163</td>
<td>0.146</td>
<td>0.126</td>
<td>0.132</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>0.143</td>
<td>0.088</td>
<td>0.109</td>
<td>0.137</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>0.163</td>
<td>0.081</td>
<td>0.143</td>
<td>0.184</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td>0.173</td>
<td>0.148</td>
<td>0.135</td>
<td>0.152</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>0.131</td>
<td>0.059</td>
<td>0.091</td>
<td>0.123</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>0.181</td>
<td>0.173</td>
<td>0.178</td>
<td>0.184</td>
<td>0.189</td>
</tr>
</tbody>
</table>

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.

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.¹

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 baseline form or a natural extension that we discuss in Section 4.2) 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 legal system (see Abrams, Bertrand, and Mullainathan, 2012).

¹Our 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).
Our test for statistical discrimination leverages the fact that the outcome variable we study—the wage—is not binary. In fact, we show (in Theorem 3) that a similar test has no empirical bite for binary outcome variables. This result is related to a critique of audit and correspondence studies first made in Heckman and Siegelman (1993) and recently revisited by Neumark (2012). For intuition, consider correspondence studies that send fictitious curricula vitae to employers and 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.

Partly motivated by this difficulty, tests for statistical discrimination in the literature are developed for settings in which the researcher has access to richer data. A classic example is Altonji and Pierret (2001) who test for statistical discrimination in wages on the basis of race. They develop a parametric model which requires panel data and their test is based on the assumption that the researcher has access to information about workers (their AFQT scores) that employers do not. Their insight is that, in a wage regression, the coefficients on variables that employers observe should fall over time but the coefficients on variables that they do not (but the econometrician does) should rise as they learn about the worker. More recently, there is a nascent experimental literature that exploits dynamics (see, for instance Bohren, Imas, and Rosenberg, 2019b) to tease out the sources of discrimination. The key observation is that dynamics help because beliefs respond to information, whereas preferences do not.

As mentioned earlier, there is an alternate strand of the literature that develops outcome tests (in the spirit 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 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). 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

We reiterate that, relative to the above mentioned papers, our main insights are that (i) statistical discrimination can be tested on cross-sectional wage data, (ii) our model allows us to interpret a rejection of our test as evidence of taste-based discrimination and (iii) our test is easy to state and implement so can be used by non-experts. We also view our novel non-parametric approach to be a conceptual contribution to the literature. As mentioned above, we can use it to formalize the difficulty of finding the source of discrimination in experiments that generate binary outcomes. We can also adapt our methodology to outcome tests and we demonstrate (in Theorem 4) how testable implications can be derived without having to identify the marginal case.

The key assumption on which we build our analysis is that both groups have the same mean productivity. In one sense, this is a weak assumption because it allows for the variance of the advantaged group to be higher, and this “variability hypothesis” is sometimes proposed as an explanation for differences in outcomes. Moreover, this assumption is implicit whenever concerns about wage gaps are expressed in public discourse. After all, if one group was systematically more productive than the other, then differences in wages would be perfectly justified. That said, our framework can also be employed to derive the smallest difference in mean productivities that can explain the wage differences in the absence of bias. To be more specific, suppose we believed that employers were not biased and that the distinct wage distributions were the result of both differences in mean productivities and incomplete information. What is the minimum difference in productivities required to explain the wage disparity? We show (in Theorem 5) that the wage gap can be interpreted as such a bound but only when the wage distribution of the advantaged group first-order stochastically dominates that of the disadvantaged group; when this is not the case, a consequence of our main result (Theorem 1) is that this bound is zero.

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.

---

3 There is a literature on statistical discrimination following Arrow (1973) which studies how the different prior beliefs (or stereotypes) affect investments and therefore outcomes. We reiterate that 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.

4 In fact, we show in our main result (Theorem 1) that the empirical content of a model that allows for difference variances in the productivity distribution is exactly the same as a model which requires not just the means but the entire productivity distributions of both groups to be identical.
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. 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 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 $i \in \{1, 2\}$. 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 testable differences between a model that assumes identical distributions and another that assumes different distributions, but with identical means. Testing the hypothesis of identical means is, however, easier than testing the hypothesis of 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

---

5 Throughout, all distributions are right-continuous and have limits on the left.
6 In the context of college admission, academic tests, such as SAT and ACT, have been found to discriminate against low-income, minority and female students. See https://www.forbes.com/sites/markkantrowitz/2021/05/21/how-admissions-tests-discriminate-against-low-income-and-minority-student-admissions-at-selective-colleges/.
7 The restriction of productivities to the set $[0, 1]$ is a normalization.
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 \((S_i, \pi_i)\) for group \(i \in \{1, 2\}\) consists of a set of signals \(S_i = \Theta\) and a joint distribution \(\pi_i\) over \(\Theta \times S_i\), whose marginal distribution over \(\Theta\) is \(H_i\). Denote the marginal distribution of \(\pi_i\) over \(S_i\) as \(F_i\). Moreover, to reflect the “unbiased” terminology, we require that the posterior estimate \(\mathbb{E}_{\pi_i}[\theta_i | s_i]\) of the productivity satisfy

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

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 \(\theta_i\) 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 \(H_i \succcurlyeq F_i\). 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 mean. 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 \(H_i \succ F_i\).

**Wage function:** If an employer estimates the productivity of a worker to be \(\theta\), 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 \(\{\theta : W(\theta) \leq w\}\) according to \(F_i\), 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 \(w > W(1)\). Note that, even though the wage function does not depend on group

---

\(^8\)Integration by parts implies that the mean satisfies \(\int_0^1 \theta_i dF_i(\theta_i) = \theta_i F_i(\theta_i)|_0^1 - \int_0^1 F_i(\theta_i) d\theta_i = 1 - \int_0^1 F_i(\theta_i) d\theta_i\).

\(^9\)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 \(w\).
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 model is consistent with the existence of a wage gap between the two groups.

**Consistency with statistical discrimination:** We say that the observed wage distributions $G_1$ and $G_2$ are consistent with statistical discrimination if there exist productivity distributions $H_i$, distributions of posterior estimates $F_i$ that satisfy $H_i \succ_{\mathbb{P}} F_i$ 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., $G_i(w) = 1$ for some $w$, $i = 1, 2$.)

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 productivity 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. In Section 5, we relax this assumption and derive a lower bound on the mean productivity difference required to explain any two wage distributions as the result of statistical discrimination alone.

We end this section with a brief discussion of the second assumption. Our model is in the spirit of Phelps (1972). Phelps 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 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 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 $\pi_i$ as opposed to just the (marginal) distributions $F_i$ of the posterior estimates.

3. TESTING FOR DISCRIMINATION

In Section 3.1, we present our main result (Theorem 1) that characterizes wage distributions that are consistent with statistical discrimination. In Section 3.2, we then show (in Theorem 2) that taste-based discrimination can also be modeled within our framework. The key observation of that sub-section is that every pair of wage distributions are consistent with taste-based discrimination. This observation is important because it allows us to interpret a rejection of the test for consistency with statistical discrimination as evidence of bias.

3.1. STATISTICAL DISCRIMINATION

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 productivity 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 $G_j$, 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 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

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

with the inequality strict for some $\theta$, 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 $H_i$ 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 two implications of this result. 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 the model.

Note also that the above test for statistical discrimination 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 our test requires 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 bias 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, 2019a). Of course, employers may have inaccurate beliefs that assign a lower mean productivity to the disadvantaged group. We interpret this to be taste-based discrimination or bias.

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

<table>
<thead>
<tr>
<th>wage/hour</th>
<th>$10</th>
<th>$15</th>
<th>$20</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_1$</td>
<td>1/3</td>
<td>5/12</td>
<td>1</td>
</tr>
<tr>
<td>$G_2$</td>
<td>1/6</td>
<td>1/2</td>
<td>1</td>
</tr>
</tbody>
</table>

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

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

which, in words, ensures that $F_1$ and $F_2$ have the same mean. A 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 $H \succ_2 F_i, i \in \{1, 2\}$. 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,

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

It is easy to check that this is indeed positive for all $\theta$, hence $H \succ_2 F_1$. 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 $H \succ_2 F_2$.

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 Appendix A.
3.2. TASTE-BASED DISCRIMINATION

In this section, we model taste-based discrimination within our framework and 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 $w_i = \inf \{ w \in \mathbb{R}_+ \mid G_i(w) > 0 \}$ and $\overline{w}_i = \sup \{ w \in \mathbb{R}_+ \mid G_i(w) < 1 \}$. In words, $[w_i, \overline{w}_i]$ 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 $\varphi : [w_1, \overline{w}_1] \to [\overline{w}_2, \overline{w}_2]$ such that $G_1(w) = G_2(\varphi(w))$ for all $w \in [w_1, \overline{w}_1]$.

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. 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 or animus. 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.

Finally, the second part of Theorem 2 shows that, when the productivity distributions of both groups are the same (and the inferred productivities are not garbled by noisy signals), the empirical content of taste-based discrimination is not vacuous. For applied purposes, we view the assumption of identical productivity distributions to be overly strong, but the theoretical implications of this result are worth discussing. Specifically, this result shows that the wage distributions must be related via the monotone transformation \( \phi \). 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 distributions, so that \( \phi := W_2 \circ W_1^{-1} \). 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 \( \phi \) is guaranteed. We can simply choose \( \phi := G_2^{-1} \circ G_1 \). The proof of Theorem 2 is in the Appendix A.

This result has two 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., \( s = G_i(w_i) - \lim_{w_i \uparrow w} G_i(w) \), then the distribution \( G_j \) has an atom at the wage \( w_j = \phi(w_i) \) 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 a fraction 5/12 of workers from group 1 are paid $15 per hour, there must exist some productivity level corresponding to this wage. More importantly, 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.

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 \( \phi : [w_1, w_1] \to [w_2, w_2] \) satisfying \( \phi(w) \geq w \) for all \( w \in [w_1, w_1] \). In
Table 3: Inconsistency with statistical, taste-based discrimination assuming identical productivity distributions

<table>
<thead>
<tr>
<th>wage/hour</th>
<th>$10</th>
<th>$15</th>
<th>$20</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_1$</td>
<td>1/3</td>
<td>3/4</td>
<td>1</td>
<td>175/12</td>
</tr>
<tr>
<td>$G_2$</td>
<td>1/6</td>
<td>1/2</td>
<td>1</td>
<td>200/12</td>
</tr>
</tbody>
</table>

words, at all quantile $q$, workers from group 2 are paid more than workers from group 1.

4. DISCUSSION

In this section, we use our framework to revisit classic approaches to testing for discrimination. First, in Section 4.1, we consider the version of our model where the outcome is binary and we show that statistical discrimination has almost no empirical bite. Then, in Section 4.2, we illustrate the versatility of our approach by revisiting outcome tests à la Becker.

4.1. 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. Dovetailing on the insight in Heckman and Siegelman (1993); Heckman (1998), we stated that statistical discrimination has little bite in such binary settings. 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, labeled $w = 0$ 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 productivity distributions $H_1$ and $H_2$, distributions of posterior estimates $F_1$ and $F_2$ that satisfy $H_i \succsim F_i$, and a cutoff $\theta$ such that $F_i(\theta) = G_i(0)$ 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 $\theta \in [0, 1]$ 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 $G_1$ and $G_2$ are consistent with statistical discrimination if, and only if, it is not the case that either $G_1(0) = 1$ and $G_2(0) = 0$ or $G_1(0) = 0$ and $G_2(0) = 1$.

In words, any pair of binary distributions is consistent with statistical discrimination unless the outcome for one group is 0 with probability 1, while it is 1 with probability 1 for the other. 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 never observed and therefore, in practice, statistical discrimination cannot be disentangled from bias or animus in a setting with binary outcomes.

In light of our main result (Theorem 1), it might seem surprising to some readers that Theorem 3 allows two outcome distributions ordered by first-order stochastic dominance to be consistent with statistical discrimination. This difference is driven by the fact that, with binary outcomes, many different posterior estimates are grouped together and assigned a single outcome (depending on whether they are above or below the cutoff). This is explicitly ruled out in Section 2 because we require the wage function to be strictly increasing and consequently, no two workers with distinct productivity estimates are assigned the same wage.

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 $\theta$ (since $F_1(\theta) = 1$), whereas the mean of $F_2$ must be strictly greater (since $F_2(\theta) = 0$). 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 $H \gtrsim F_i$, as required. The argument is completed by setting the threshold $\theta = 1 - G_1(0)$.

4.2. 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 4.1, where only a single piece of information was observed, we–the analyst–now observe two pieces of information: the release decision along with the post-decision result, that is, whether 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 4.1. 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 \( r_i \) 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 consider the following model. Assume that \( Y_i \) is distributed with (unknown) probability \( \theta_i \in [0, 1] \) in group \( i \), with \( H_i \) the prior distribution of \( \theta_i \). 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 \( \bar{\theta} \). As in previous sections, we assume that the judge receives an unbiased signal about \( \theta_i \), 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 \( F_i(\theta) \), while the pre-trial misconduct rate conditional on release is \( \mathbb{E}_{F_i}[\theta_i|\theta_i \leq \theta] \).

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 \( \theta \) such that (i) \( \mathbb{E}_{H_1}[\theta_1] = \mathbb{E}_{H_2}[\theta_2] \), (ii) \( H_i \succ_2 F_i \), and (iii) \( r_i = F_i(\theta) \) and \( q_i = \mathbb{E}_{F_i}[\theta_i|\theta_i \leq \theta] \) 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. Bounding Mean Productivity Differences in the Absence of Bias

As we highlighted earlier, the key assumption driving our main result (Theorem 1) is that mean productivities of both groups are the same. It is worth reiterating that our view is that this is an eminently reasonable assumption around which to base a study on wage discrimination. After all, why worry about wage gaps for workers with equal qualifications if workers from the advantaged group just have a higher mean productivity (that is unobserved to the researcher) and so deserve a higher mean wage?

With that said, one could ask the following question. Suppose we assume that employers are not biased and the distinct wage distributions are the result of statistical discrimination alone. What is the smallest possible difference in mean productivities that could explain the wage differences? In this section, we show that our framework can be employed to answer this question.

We return to the main model in Section 2 that this paper focuses on. We say that wage distributions \( G_1 \) and \( G_2 \) are rationalizable with productivity difference \( d \geq 0 \) if there exist productivity distributions \( H_i \) whose means differ by at most \( d \) (that is \( |\mathbb{E}_{H_1}[\theta] - \mathbb{E}_{H_2}[\theta]| \leq d \)), distributions of posterior estimates \( F_i \) that satisfy \( H_i \succ_2 F_i \) for \( i \in \{1, 2\} \), and a continuous and strictly increasing wage function \( W \), such that these jointly induce the observed wage distributions.

Recall that we have normalized the productivities to be in \([0, 1]\). However, as we now argue, another
normalization is needed if we want to derive a meaningful bound on the difference of productivity means. To see this, suppose wage distributions $G_1$ and $G_2$ are induced by productivity distributions $H_1$ and $H_2$ with productivity difference $d$, distributions of posterior estimates $F_1$ and $F_2$, and the wage function $W$.

Consider any constant $\alpha < 1$ and construct productivity distributions and distributions of posterior estimates:

$$\tilde{H}_i(\theta) := H_i(\theta/\alpha) \quad \text{and} \quad \tilde{F}_i(\theta) := F_i(\theta/\alpha)$$

for all $\theta \in [0, \alpha]$, for all $i \in \{1, 2\}$. (Naturally, $\tilde{H}_i(\theta) = \tilde{F}_i(\theta) = 1$ for all $\theta \in (\alpha, 1]$.) Similarly, construct the wage function

$$\tilde{W}(\theta) := \begin{cases} W(\theta/\alpha) & \text{for } \theta \in [0, \alpha], \\
W(1) + (\theta - \alpha) & \text{for } \theta \in (\alpha, 1]. \end{cases}$$

It is routine to verify that these newly constructed model primitives also rationalize the wage distributions $G_1$ and $G_2$, but now the means of the productivity distributions $\tilde{H}_1$ and $\tilde{H}_2$ differ by $\alpha d$. In words, the new distributions are concentrated on the smaller interval $[0, \alpha]$ and the new wage function is scaled up appropriately so as to induce the wage distributions $G_1$ and $G_2$. So, absent another normalization, we can scale the productivity distributions along with the wage function to make the means of the productivity distributions arbitrarily close.

The discussion in the previous paragraph suggests that bounding the slope of the wage function is a natural way to normalize the model. With that in mind, we say that wage distribution is Lipschitz with constant $k$ if

$$|W(\theta) - W(\theta')| \leq k|\theta - \theta'| \quad \text{for all } \theta, \theta'.$$

This restriction implies that the slope of the wage function (wherever differentiable) satisfies $W'(\theta) \leq k$ for all $\theta$. This restriction allows us to interpret the productivity parameter in terms of dollars: a one unit increase in productivity has at most a value of $k$ dollars in terms of wages. We hasten to stress that if we impose an upper bound on the slope of the wage function, requiring the productivity distributions to be supported on subsets of $[0, 1]$ is no longer a normalization (as the above arguments demonstrate). Accordingly, we do not impose that restriction in the rest of the section.

Given a constant $k > 0$, we say that $d \geq 0$ is a lower bound on productivity differences if, whenever productivity distributions $H_1$ and $H_2$, distributions of posterior estimates $F_1$ and $F_2$, with $H_i \succsim_2 F_i$ for $i \in \{1, 2\}$, and continuous, strictly increasing, $k$-Lipschitz wage function $W$ induce wage distributions $G_1$ and $G_2$, the productivity distributions satisfy $\mathbb{E}_{H_1}[\theta] - \mathbb{E}_{H_2}[\theta] \geq d$. In words, the difference in the means of any pair of productivity distributions that induce the given wage distributions (via statistical experiments and a wage function) must be at least $d$. 

18
The next result provides a tight lower bound.

**THEOREM 5.** Given a constant $k$ and two wage distributions $G_1$ and $G_2$. A lower bound on the productivity differences is given by

$$d := \begin{cases} \frac{\vert \mathbb{E}_{G_1}[w] - \mathbb{E}_{G_2}[w] \vert}{k} & \text{if } G_1 \succsim_1 G_2 \text{ or } G_2 \succsim_1 G_1, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $G_1$ and $G_2$ are rationalizable with productivity difference $d$ and a $k$-Lipschitz wage function.

In words, the first part of this result states that the wage gap scaled by the normalizing constant $k$ is a lower bound on the mean productivity difference (assuming there is no employer bias) required to rationalize wage distributions ordered by first-order stochastic dominance. When neither distribution first-order stochastically dominates the other, this bound is 0. The proof can be found in Appendix A.

Importantly, the bound is tight in that we can always find model primitives that rationalize the wage distributions such that the mean productivity difference is exactly equal to $d$. This is straightforward to show. Suppose $G_1 \succsim_1 G_2$. Let the wage function be $W(\theta) = k\theta$ and let $H_i(\theta) = F_i(\theta) = G_i(k\theta)$ for $i \in \{1, 2\}$. Then,

$$\mathbb{E}_{H_1}[\theta] - \mathbb{E}_{H_2}[\theta] = \int_0^\infty \theta[dH_1(\theta) - dH_2(\theta)] = \frac{1}{k} \int_0^\infty w[dG_1(w) - dG_2(w)] = \frac{\mathbb{E}_{G_1}[w] - \mathbb{E}_{G_2}[w]}{k}.$$ 

When $G_1 \not\succsim_1 G_2$ and $G_2 \not\succsim_1 G_1$, the tightness of the bound follows immediately from the proof of Theorem 1. Indeed, to rationalize $G_1$ and $G_2$, we construct productivity distributions $H_1$ and $H_2$, estimate distributions $F_1$ and $F_2$, all supported on subsets of $[0, 1]$, and a piece-wise linear wage function $W$ (the function has three pieces). The function $W$ is Lipschitz, but its modulus of continuity may be higher than $k$. If so, we simply need to “stretch” the distributions and the wage functions by the appropriate factor (as we did above but with $\alpha > 1$).

To summarize, our model provides a lens via which we can interpret the wage gap (taking the normalizing constant $k = 1$) as a measure of the lowest mean productivity difference that is required to explain the wage differences if employers are not biased. But the nuance is that this measure can only be employed in the absence of first-order stochastic dominance; otherwise, we cannot rule out a zero mean productivity difference.

6. **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 conceptual contributions. In casual discussions of wage gaps, the mean wages of two groups are typically compared and differences are often interpreted (without justification) as evidence of bias. Our test is micro-founded 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 work.
A. PROOFS

We start with a preliminary remark. The distribution $G_i$ is supported on a closed subset of $[\underline{w}_i, \bar{w}_i]$, where $0 \leq \underline{w}_i = \inf\{ w \in \mathbb{R}_+ \mid G_i(w) > 0 \}$ and $\bar{w}_i = \sup\{ w \in \mathbb{R}_+ \mid G_i(w) < 1 \} < \infty$ (since the wage distributions have bounded support). Let $W_i$ be the random variable with distribution $G_i$. The random variable $f_{W_i} := 1_{\underline{w}_i}W_i$ is then supported on a closed subset of $[0, 1]$ with distribution $e_{G_i}$, where $e_{G_i}(w) = G_i(\underline{w}_i w)$. 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 $H \succcurlyeq F_i$ for both $i \in \{1, 2\}$.

**Step 1:** There exists a strictly increasing and continuous function $W : [0, 1] \rightarrow [0, 1]$ 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 $[\underline{w}, \bar{w}] \subset (0, 1)$ such that $\int_{\underline{w}}^{\bar{w}} \Delta_G(w)dw < 0$. (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
\[
\frac{1}{\gamma^+} \int_0^w \Delta_G(w)dw + \frac{1}{\gamma^-} \int_w^\bar{w} \Delta_G(w)dw + \frac{1}{\gamma^+} \int_1^{\bar{w}} \Delta_G(w)dw = 0.
\]

Define
\[
\kappa = \frac{\bar{w} - w(1 - (\gamma^-/\gamma^+))}{\gamma^-} + \frac{1 - \bar{w}}{\gamma^+} = \frac{\gamma^+(\bar{w} - w) + \gamma^-(1 - (\bar{w} - w))}{\gamma^+\gamma^-} > 0.
\]
Note the inequality follows from the fact that $\bar{w} > w$, $\bar{w} - w < 1$, $\gamma^- > 0$ and $\gamma^+ > 0$.

Using this $\kappa$, we define
\[
\theta = \frac{w}{\kappa\gamma^+} > 0,
\]
and
\[
\overline{\theta} = \frac{\bar{w} - w(1 - (\gamma^-/\gamma^+))}{\kappa\gamma^-} = \frac{\bar{w} - w}{\kappa\gamma^-} + \theta > \theta.
\]
Also note that
\[
\overline{\theta} < 1,
\]
since
\[
\kappa\gamma^- > \bar{w} - w(1 - (\gamma^-/\gamma^+)).
\]

Now consider the following piecewise linear wage function
\[
W(\theta) = \begin{cases} 
\kappa\gamma^+ \theta & \text{if } 0 \leq \theta < \theta, \\
\kappa\gamma^- \theta + w(1 - (\gamma^-/\gamma^+)) & \text{if } \theta \leq \theta \leq \overline{\theta}, \\
\kappa\gamma^+ \theta + 1 - \kappa\gamma^+ & \text{if } \overline{\theta} < \theta \leq 1.
\end{cases}
\]

Few observations are worth making. First, $W$ is continuous because
\[
\lim_{\theta \uparrow \overline{\theta}} W(\theta) = \kappa\gamma^+ \overline{\theta} = \bar{w} = \kappa\gamma^- \theta + w(1 - (\gamma^-/\gamma^+)) = W(\overline{\theta}),
\]
\[
W(\overline{\theta}) = \kappa\gamma^- \overline{\theta} + w(1 - \gamma^-/\gamma^+) = \kappa\gamma^- \frac{\bar{w} - w(1 - \gamma^-/\gamma^+)}{\kappa\gamma^-} + w(1 - \gamma^-/\gamma^+) = \bar{w}
\]
and
\[
\lim_{\theta \downarrow \overline{\theta}} W(\theta) = \kappa\gamma^+ \overline{\theta} + 1 - \kappa\gamma^+
\]

22
\[
\begin{align*}
&= \kappa \gamma^+ \frac{\bar{w} - \frac{1 - \gamma^+}{\gamma^-} \gamma^-}{\gamma^+} + 1 - \kappa \gamma^+ \\
&= \gamma^+ \frac{\bar{w} - \frac{1 - \gamma^-}{\gamma^+}}{\gamma^-} + 1 - \gamma^+ \left( \frac{\bar{w} - \frac{1 - \gamma^-}{\gamma^+}}{\gamma^-} + \frac{1 - \bar{w}}{\gamma^+} \right) = \bar{w}.
\end{align*}
\]

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
\[
\int_0^1 \Delta F(\theta) d\theta = \int_0^\theta \Delta F(\theta) d\theta + \int_\theta^1 \Delta F(\theta) d\theta \\
= \int_0^\theta \Delta G(W(\theta)) d\theta + \int_\theta^1 \Delta G(W(\theta)) d\theta \\
= \frac{1}{\kappa \gamma^+} \int_0^{\bar{w}} \Delta G(w) dw + \frac{1}{\kappa \gamma^-} \int_{\bar{w}}^w \Delta G(w) dw + \frac{1}{\kappa \gamma^+} \int_w^1 \Delta G(w) dw \\
= 0,
\]
where the second last equality follows from the change of variables from \( \theta \) 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
\[
\int_0^\theta H(x) dx \geq \max \left\{ \int_0^\theta F_1(x) dx, \int_0^\theta F_2(x) dx \right\} \text{ with equality at } \theta = 1.
\]

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

23
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 $\int_0^1 w\,dG_1(w) = \alpha \int_0^1 w\,dG_2(w)$.

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 $\varphi : [\bar{w}_1, \bar{w}] \rightarrow [\bar{w}_2, \bar{w}_2]$ such that $G_1(w) = G_2(\varphi(w))$. For future reference, note that we must have $\varphi(\bar{w}_1) = \bar{w}_2$ and $\varphi(\bar{w}_1) = \bar{w}_2$.

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

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

Notice that for all $w \in [W_1(0), W_1(1)], W_1^{-1}(w) = \frac{w - \bar{w}_1}{\bar{w}_1 - \bar{w}_1}$, hence $G_1(w) = H(W_1^{-1}(w)).$ Moreover, for all $w < W_1(0), G_1(w) = 0$ since $W_1(0) = \bar{w}_1.$ Similarly, for all $w > W_1(1), G_1(w) = 1$ since $W_1(0) = \bar{w}_1.$
Similarly, for all \( w \in [W_2(0), W_2(1)] \), \( W_2^{-1}(w) = \frac{\varphi^{-1}(w) - \overline{w}_1}{\overline{w}_1 - \overline{w}_1} \). Therefore,

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

Moreover, for all \( w < W_2(0) \), \( G_2(w) = 0 \) since \( W_2(0) = \varphi(\overline{w}_1) = \overline{w}_2 \). 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 \( \varphi \) 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, \([\overline{w}_1, \overline{w}_1] \subseteq [W_1(0), W_1(1)]\), the function \( \varphi : [\overline{w}_1, \overline{w}_1] \to [W_2(0), W_2(0)] \) 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 \( \overline{w}_1 \), \( G_1(w) = G_2(\varphi(w)) > 0 \), hence \( \lim_{w_1 \to \overline{w}_1} \varphi(w) = \varphi(\overline{w}_1) \geq \overline{w}_2 \). 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, \([\overline{w}_2, \overline{w}_2] \subseteq [W_2(0), W_2(1)]\), the function \( \varphi^{-1} : [\overline{w}_2, \overline{w}_2] \to [W_1(0), W_2(0)] \) 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, \( \varphi^{-1}(\overline{w}_2) \geq \overline{w}_1 \) and \( \varphi^{-1}(\overline{w}_2) \leq \overline{w}_1 \).

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

\[\square\]

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

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

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 \((i, j), j \neq i\).
The proof is constructive. Without loss of generality, assume that \( q_i \geq q_j \). It follows that \( r_i q_j + (1 - r_i) 1 > q_j \) 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 \( \theta = q_i + \delta \) 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 \( F_i(\theta) = r_i \).

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 \( H \succ_2 F_i \) and \( H \succ_2 F_j \). The distribution \( F_j \) is again binary and takes values \( q_j \) and \( q_i - r_j q_j + (1 - r_j) \varepsilon \) 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 \( \delta \) and \( \varepsilon \), 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 \( F_j(\theta) = r_j \). 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 \( \varepsilon \) and \( \delta \) as required.

\[\blacksquare\]

**Proof of Theorem 5.** We only need to prove the first statement of the theorem since the tightness of the bound was established in the text. So suppose the given wage distributions \( G_1 \) and \( G_2 \) are induced by model primitives \( H_1, H_2, F_1, F_2 \) and \( W \).

Since the wage function is Lipschitz continuous, it is differentiable almost everywhere. Therefore, for each group \( i \in \{1, 2\} \), we can write

\[
\mathbb{E}_{G_i}[w] = [w G_i(w)]^1_0 - \int_0^1 G_i(w) dw = 1 - \int_0^\theta W'(\theta) F_i(\theta) d\theta.
\]

Clearly, we only need to establish the bound for the case when \( G_i \succ_1 G_j \) for some \( i, j \in \{1, 2\} \). In this
case, the above equation implies that

\[
\mathbb{E}_{G_j}[w] - \mathbb{E}_{G_i}[w] = - \int_0^\theta W''(\theta)[F_j(\theta) - F_i(\theta)]d\theta \\
\geq -k \int_0^\theta [F_j(\theta) - F_i(\theta)]d\theta \\
= -k \left( \mathbb{E}_{F_j}[\theta] - \mathbb{E}_{F_i}[\theta] \right),
\]

where the inequality follows from the fact that \( W \) is \( k \)-Lipschitz and \( F_j(\theta) \geq F_i(\theta) \) (since \( G_j(\theta) \geq G_i(\theta) \)). This completes the proof. \[\square\]
REFERENCES


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


