第一页为封面页

参赛队员姓名：白小舟 Natalie Xiaozhou Bai

中学：清华附中国际部

省份：北京

国家/地区：中国

指导教师姓名：陈济冬

论文题目：Why Statistical Discrimination Can Be Inefficient: An Agency-Based Theory

第二页为创新性申明

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处，本人愿意承担一切相关责任。

参赛队员： 白小舟　　　　指导老师： 陈济冬

2020 年 9 月 15 日

第三页此处开始为论文的主体部分……

论文题目 Why Statistical Discrimination Can Be Inefficient: An Agency-Based Theory

作者 白小舟 Natalie Xiaozhou Bai

论文摘要 Suppose an agent is tasked to make a diagnosis about a certain condition of the subject. If the agent cannot observe the subject's condition directly, he will try to find and use information about the subject's condition in order to make the right diagnosis. Suppose the subjects can be divided into groups, and the different groups have different likelihood of having the condition, then using group identity helps the agent better diagnose. This would result in discrimination in the sense that agents in different groups get different diagnosis even when they look otherwise the same. Such discrimination seems efficient as it makes use of all relevant information available to the agent. This paper shows that such discrimination can be socially inefficient if the agent does not bear the full social cost of misdiagnosis. The reason is that using group identity to help diagnosis may dull the incentives for the agent to find other information about the subject's condition.

关键词 Statistical Discrimination, Agency, Efficiency

目录

论文正文：

1. Introduction

Racial discrimination is an important issue. For example, according to Quillian et al (2017), racial discrimination is still manifest in labor market. Based on the analysis of various field tests dating from 1990 to 2015, the average employment rate of white applicant is 36% higher than that of African American applicants, and 24% than that of Latino Americans, and this disparity experienced negligible fluctuation in the 25 years timespan. It has been shown that the result is valid even after the applicants' educational degree, gender, and the occupations they were applying for are accounted for. Such discrimination is not limited to employment decisions, it is also prevalent in crime investigation, college enrollment, etc. It hinders different races from sharing equal opportunities, dividing the society and widening the gap between demographic groups. The recent George Floyd case again brought the issue to the spotlight.

We can easily assume that racial discrimination is just an incidence driven by the innate hatred or dislike towards another race, but the constant repetition of the same tragedy tells us that simply attributing this event to emotional impulse is not a responsible way of viewing this matter.

There is a large literature on statistical discrimination, following the seminal work by Phelps (1972), that seems to justify the existence of discrimination by showing that discrimination is the rational response of economic agents when they face imperfect information about the characteristics of people they interact with. Given some known average differences between groups, for the decision maker to use group members' group identity as an influential factor in the evaluation of their characteristics is shown to be the result of optimization of the decision maker's objective that does not directly depend on group identity of the group members. Often times, the literature shows such statistical discrimination is socially efficient. Given such results, it becomes more difficult to eradicate discrimination.

This paper questions the result that statistical discrimination is efficient. It uses an agency based model to show that, under very reasonable conditions, statistical discrimination is inefficient. By assuming that the decision maker does not bear the full cost of making wrong inference about the characteristics of subjects they are charged to work with, the paper shows that allowing the decision maker to discriminate reduces the incentives for him to work hard to find accurate information about the characteristics of the subjects, resulting in socially inefficient outcome. The implication of this result is that, even for the purpose of pursuing efficiency, racial discrimination should not be allowed.

The organization of the paper is as follows. It reviews the relevant literature first. There, it is shown that our theory is distinct from the existing theories in important ways, and it offers news insights for us to better analyze discrimination. The paper then presents a simple model with discrete variables to illustrate our main idea and results. This is followed by the presentation and analysis of a model with continuous variables. The paper then concludes by summarizing the key ingredient of the models.

2. Literature review

When we talk about racial discrimination, we generally refer to two branches of theories. The first is taste-based theory of discrimination, first defined by Gary Becker in 1957. Taste based discrimination is derived from the agent's preference bias or dislike towards specific race or organizational culture, which results in negative effect on the welfare of minority groups. In contrast to taste-based theory of discrimination, there are statistical discrimination theories. In the model of statistical discrimination, economic agents have no ill intention against any targeted groups, however, they are utility or profit maximizers, which lead them to make decisions that may result in inequality. In this study, we will mainly be focusing on statistical discrimination.

The study of statistical discrimination can again be divided into two branches based on the different premises researchers use to derive their model. One branch is led by Phelps (1972). Phelps starts off his analysis by assuming that there is an exogenous difference between groups of workers. The cause of this innate difference is irrelevant to the study, but the disparity between group average is crucial, as we will explain in the latter paragraph. Another branch is pioneered by Arrow, K. J. (1973). In contrast to Phelps' study, Arrow derived his model by assuming that there is no exogenous difference between demographic groups. Instead, the difference is generated within the model, which is defined as endogenous difference. Comparing to Arrow's model, Phelps' model, which take the difference as given, is better related to our study, thus we will extend further on Phelps' method.

Assume there are two groups of workers whose group identities can be represented as $j = \{B, W\}$. The workers' skill, which is equivalent to the value of their marginal product when employed, is represented as q. However, the employer cannot accurately evaluate the workers' skill. The two messages the employer can receive are group identity and a noisy signal of productivity. The noisy signal, which we denote by θ, is positively related to q, but is disturbed by a zero mean error ε, which gives the equation $\theta=q+\varepsilon$. In a perfectly competitive labor market, employers are assumed to share the same information about the applicants. This setting determines that the best strategy for the employers to determine the wage is to pay the workers the number equivalent to their marginal productivity: on the one hand, if the wage is lower than the marginal productivity of the worker, there can always be other companies swooping in and take away the labor; on the other hand, the value of the wage certainly won't exceed the profit the worker is expected to generate for the company. In this scenario, given that the skill level of the workers cannot be determined with certainty, the employer will deduce the expected productivity of the workers using the noisy signal he perceives coupling with observable characteristics such as racial identity and the mean productivity of the race. In his study, Phelps shows that the expected productivity given θ equals to the weighted average of the signal and the unconditional group mean μ_j:

$$\mathrm{E}(\mathrm{q}|\theta) = \frac{\sigma_j^2}{\sigma_j^2 + \sigma_{\varepsilon j}^2}\theta + \frac{\sigma_{\varepsilon j}^2}{\sigma_j^2 + \sigma_{\varepsilon j}^2}\mu_j$$

Phelps categorizes the implication of this model into two cases. In this first case, Phelps assumes that there a difference in the two demographic groups' average productivity, but the groups signals are equally informative, which means that $\sigma_{\varepsilon B}=\sigma_{\varepsilon W}=\sigma_{\varepsilon}$, and $\sigma_B=\sigma_W=\sigma$. Given this condition, the expected productivity of the worker is partially determined by the mean productivity μ_j of the group he belongs to, and the noisier the signal is, that is, the greater the value $\sigma_{\varepsilon j}$ is, the more discretion has been placed on μ_j, and consequently inequality is yield---one employee is paid less than another employee even though the same signal is received of their productivity, because the mean productivity of the group he belongs to is lower than the other's. In the second case, the premise is that the two groups have the same mean productivity, while the signals the employers receive are differently informative. Imagine if the employer belongs to the same group with one of the two employees, then we can reasonably project that he is more likely to understand and accurately evaluate the competence of the worker from the same group that he belongs to since they share the same culture. If the employee and one of the worker both belongs to the W group, then $\sigma_{\varepsilon B}>\sigma_{\varepsilon W}$, meaning that comparing with that of the W group, the mean productivity of B group (μ_B) plays a more important role in determining the expected productivity of the worker from the group, and undermines the effect of the worker's personal performance. Consider that the average productivity of the two groups are the same, people may think that it is a relatively fair model. However, the successors of Phelps' model used more comprehensive methodologies to show that one group can still receive more mistreatment than another under this setting.

In 1977, researcher Aigner, D. and G. Cain published their article "Statistical Theories of Discrimination in the Labor Market" and improved the understanding of statistical discrimination. Their innovation is to take the employer's risk averseness into the consideration. People generally prefer certainty over uncertainty, and are willing pay less for the product or service that requires them to bear more risk in order to compensate for their loss of utility. This is the same mechanism that motivates people to pay for insurance: uninsured individuals have lower levels of utility than those who spend money on their insurances, as they are required to bear the risk themselves, instead of having the insurance company to share their risk. In the scenario provided in the last paragraph, although workers from group B are on average not disadvantaged, their signal is less informative to the employer from group W, which means that the employer perceives uncertainty. As a result, the employee is willing to pay less to workers from group B as an insurance of avoiding unworthy investments, and racial pay gap thereby emerges.

Last but not least, there is the study carried by Lundberg, S. and R. Starts in 1983. In this study, the two researchers introduce the concept of human capital investment to the model established by Phelps. Human capital investment is an effective way of elevating a worker’s marginal output, however, it is costly. Once again, based on the second case proposed by Phelps, the true marginal product of the worker from group B plays a less important role in determining the expected productivity than that of worker from group W. Consequently, with the same human capital investment, the worker from group B will receive less payback than worker W, thus have less incentive to invest his time and energy into improving his skill. This original setting will logistically result in a gap between the working ability of the two groups and continuously increase the gap as time went on, producing an endogenous difference that roots from the exogenous difference of informativeness.

Among the studies mentioned above, only the second case proposed by Phelps does not involve inequality, and all other cases results in greater disadvantage for one group over another. However, most of them concern only the damage of discrimination to a specific group instead of the total social welfare, and they denounce statistical discrimination by pointing out that it causes inequality and is ethically wrong. However, the measurement for morality is abstract, and some advocators of statistical discrimination may argue that sacrificing some extent of moral righteousness to achieve higher efficiency is justifiable. Thus immorality may not be a compelling reasoning for everyone. The conception of Lundberg and Starts’ study took the social welfare into consideration, as the reduction in human capital investment will eventually undermines the productivity of the society, but it’s still different from our case. Lundberg and Starts uses productivity as a measurement of social welfare, and they assume that all information obtainable to the agency is imperfect. However, our assumption is that the accuracy of the information can be improved through much dedication of the agent, and the unavoidable error caused by imperfect information has social implications itself. This implication could be emotional dissatisfaction, racial conflict, waste of talent and loss of security, etc. In total they can be summarized as a loss of social welfare, in other words, the social cost of misjudgment. In order to demonstrate this social cost more intuitively, we use drug test instead of the conventional employment problem as the example of our model. The implication on social welfare of misdiagnosing a drug addict is evident to most people.

In our study, we start off by assuming that there is an officer who does not observe with certainty whether the two groups of men he is dealing with have been on drug or not. To find the drug addicts, the officer can run a test. The test can be an informative indicator of the suspects’ guiltiness, however, to conduct the informative test requires high dedication. This dedication can be seen as the officer’s effort, and we assume that the all people has a natural inclination of preferring leisure over effort. There could be special cases whereas the agent in question is a workaholic and does not see the devotion of effort as a reduction in utility, but special cases usually do not affect the rationality of a general conclusion, thus this assumption is still widely employed in economic models. If the principal hiring the officer has a tolerance for discrimination, it means that the officer has the freedom to choose whether to test one group or not. Our hypothesis is that under the condition of allowing discrimination, there is a moral hazard that the agency would not choose to maximize the accuracy of his diagnosis, resulting in a higher social cost.

3. A Simple Model

Since this is an agency based theory, we start off by assuming that the principal hires an agent to diagnose subjects for a certain condition. One example of the condition is drug addiction. We denote the condition by a, $a = 0 \ or \ 1$, where $a = 1$ means the subject has the condition. We assume that there is a social cost of C associated with misdiagnosis. There are two types of misdiagnosis. Type I error is false positive, which means the agent diagnoses the subject as having the condition when the subject actually does not have the condition. Type II error is false negative, which means the agent diagnoses the subject as not having the condition when the subject actually has the condition. For the sake of simplicity, we assume that the cost of type I and type II errors are the same. Since we are only demonstrating the possibility of a certain theoretical result, this simplifying condition should be acceptable as long as it is not the only the key to a knife-edge result.

We further assume that the agent only bears part of the social cost of misdiagnosis, namely λC, where λ<1 is a positive number.

Assume that the agent can exert effort to test the subject and find out exactly what a is. To do the test, it costs the agent E. The agent can also shirk and not do the test. Then he gets no useful information. It is reasonable to assume that the principal cannot verify whether the agent has performed the test, if the agent can issue an arbitrary test result without any cost.

The subjects can be divided into two groups, and the probability of having the condition is different across the groups. If there's no such difference in reality, then there won't be such a thing as statistical discrimination. One example of group identity is ethnicity or racial identity. We denote group identity by r, r = 0 or 1. We assume the probability of r = 1 is

$$\mathrm{P}(r = 1) = \rho,$$

and the conditional probability of $a = 1$ given r is

$$\mathrm{P}(a = 1|r) = p_r,$$

where $r = 0 \ or \ 1$.

Note that the chance of having drug addiction is usually not very high, and is different across the groups, so we can reasonably assume that

$$\frac{1}{2} > p_1 > p_0 > 0$$

Given these specifications, we have

$$P(a = 1, r = 1) = \rho p_1,$$
$$P(a = 1, r = 0) = (1 - \rho)p_0,$$
$$P(a = 0, r = 1) = \rho(1 - p_1),$$
$$P(a = 0, r = 0) = (1 - \rho)(1 - p_0).$$

If no test is done, the diagnosis can only be based on the group identify. In this case, it is optimal to diagnose the subject as not having the condition because $\frac{1}{2} > p_1 > p_0 > 0$. Then the expected social cost of misdiagnosis is Cp_i depending on the group identity. We assume

***Assumption 1**: $Cp_i > E$ for any i.*

In words, it is socially optimal to test both groups of subjects.

Now we consider the agent's decision making. We compare two cases, one in which no discrimination between the two groups is allowed; either both groups are subject to the test, or none of the groups is.

Case 1: No discrimination is allowed between the groups.

If both groups are subject to the test, there is zero error and the agent's expected utility is
$$V_1 = -E.$$
If no one is subject to the test, we don't have any information except the group identity. Since no discrimination is allowed, even the group identity cannot be used for diagnosis. In this case, it is optimal to diagnose the agent of not having the condition, and then misdiagnosis happens when the agent has the condition and false negative cases are reported. Therefore, the agent's expected utility is
$$V_2 = P(r = 1)P(a = 1|r = 1)(-\lambda C) + P(r = 0)P(a = 1|r = 0)(-\lambda C)$$
$$= -\lambda C[\rho p_1 + (1 - \rho)p_0].$$
We assume

***Assumption** 2: $E < \lambda C[\rho p_1 + (1 - \rho)p_0]$.*
The social cost that the agency has to bear for misdiagnosing both groups is greater than effort the agency has to make to conduct the perfect diagnosis.

This assumption is true if C is sufficiently large relative to *E*. Under this assumption, we have
$$V_1 > V_2,$$
and it is optimal of the agent to test both groups. We summarize this result in the following proposition.

***Proposition 1**: If no discrimination is allowed between the groups, it is optimal of the agent to exert effort test both groups.*

Case 2: Discrimination is allowed between the groups.

Now, the agent can choose to test one group and not to test the other group. Consider the case where the agent only tests r=1, not r=0. For group r=0, we don't have any information except the group identity, it is optimal to diagnose the agent of not having the condition, and then misdiagnosis happens when the agent has the condition. Therefore, the agent's expected utility is

$$V_3 = P(r=1)(-E) + P(r=0)P(a=1|r=0)(-\lambda C)$$
$$= -\rho E - (1-\rho)p_0\lambda C.$$

We assume

Assumption 3*:* $E > \lambda p_0 C$.

In words, the effort that the agent has to exert to conduct the perfect testing is greater than the partial social cost that the agent has to bear for falsely diagnosing a suspect from group 0 as not having the condition.

Before going further, we need to check whether Assumption 3 and Assumption 2 are contradictory. If

$$\lambda p_0 C < \lambda C[\rho p_1 + (1-\rho)p_0],$$

then there exists in interval of *E* that satisfy both assumptions and the assumptions are not contradictory. By canceling out common factors and combining like terms, we get the result that $\lambda C p_0 < \lambda C[\rho p_1 + (1-\rho)p_0]$ is equivalent to $p_0 < p_1$, which is our original assumption. Therefore, Assumptions 2 and 3 are not contradictory. Based on reality, we can reasonably assume that the agent only bears a small fraction of the social cost of his mistake, especially when the principle can hardly tell if the agent is completing his job, thus the value of λ can be so small that Assumptions 1 ($Cp_i > E$) and 3 （$E > \lambda p_0 C$） are not contradictory either. With Assumption 3, we can show that

Proposition 2*: If discrimination is allowed between the two groups, it is optimal for the agent to discriminate.*

Since we have already shown that

$$V_1 > V_2 \ ,$$

we only need to show

$$V_3 > V_1$$

to prove Proposition 2.

$V_3 > V_1$ can be rewritten as

$$-\rho E - (1-\rho)p_0\lambda C > -E,$$

which can be reduced to $E > \lambda p_0 C$, that is, Assumption 3. Therefore, with the previous assumption, the agent would prefer to discriminate. This inclination of discrimination, as we would address on afterwards, will result in a loss of social welfare.

When the agent discriminates, he can also test group r=0 but not group r=1. If so, the expected utility of the agent is

$$V_4 = -\rho p_1 \lambda C - (1-\rho)\mathrm{E}.$$

If $V_4 \geq V_1$, then

$$-\rho p_1 \lambda C - (1-\rho)\mathrm{E} \geq -E,$$

which can be reduced to $\mathrm{E} \geq p_1 \lambda C$. By Assumption 2, $E < \lambda C[\rho p_1 + (1-\rho)p_0]$. Then,

$$\lambda C[\rho p_1 + (1-\rho)p_0] > p_1 \lambda C,$$

which is equivalent to $p_o > p_1$, contradicting our earlier assumption. Therefore, we have

$$V_4 < V_1 < V_3.$$

This implies that, if the agent discriminates, he will choose testing group r=1 only over testing group r=0 only.

The next question is whether we should allow discrimination if we want to maximize social welfare. When discrimination is not allowed, the agent tests both groups, and social welfare is

$$W_1 = -E.$$

When discrimination is allowed, the agent tests r=1 only, and social welfare is

$$W_3 = P(r=1)(-E) + P(r=0)(-p_0 C)$$
$$= -\rho \mathrm{E} - (1-\rho)p_0 C.$$

By Assumption 1, $p_0 C > E$, and therefore

$$W_3 < W_1.$$

We then have

Proposition 3*: In order to maximize social welfare, discrimination should not be allowed.*

4. A Continuous Model

In the above simple model, we formally demonstrate that, under certain conditions, the social welfare gets improved when no discrimination is allowed. This result is based on the specifications that: (1) the condition a and the diagnosis choice are both binary; and (2) agent receives precise information about the condition a whenever a costly test is done. In reality, it is hard to believe that the two assumptions always hold. The condition that is to be diagnosed can be very complex, and even a scientific test may not completely reveal the exact result.

As a response to these concerns, we now consider a variant version of the model, where both the condition and the diagnostic choice can take a continuum of values, and the condition a follows a normal distribution with mean a_0 and variance σ_0^2.

We use $\hat{a}$ to denote the diagnostic choice. The social welfare, i.e., the principal's payoff, is assumed to be

$$\mathrm{W}(\hat{a}; a) = -(\hat{a} - a)^2 \cdot C.$$

Similar as in the binary-value model, we assume that the agent only bears part of the social cost of misdiagnosis, namely $\lambda \cdot (\hat{a} - a)^2 \cdot C$, where $\lambda<1$ is a positive number. One can easily check that the assumption of the social welfare function is consistent with the binary-value case. Now assume the group identify r is correlated with the condition a in the following way

$$\mathrm{r} = a + \varepsilon_1,$$

where ε_1 is a normal distribution with mean 0 and variance σ_1^2 and is independent of condition a. The sum of two independent normal distributions is also a normal distribution. So r is a normal distribution. By the properties of normal distribution (DeGroot, 1970), we can calculate the expectation of a conditional on the group identify r

$$\mathrm{E}(a|\mathrm{r}) = \frac{\sigma_1^2 a_0 + \sigma_0^2 \mathrm{r}}{\sigma_1^2 + \sigma_0^2}.$$

As r increases, in the expectation above, a is more likely to have a higher value. This is consistent with the assumption in the discrete case that $\mathrm{P}(a = 1|r = 1) > \mathrm{P}(a = 1|r = 0)$.

Consider the additional signal s that the agent can receive. The signal is assumed to be generated according to the following process.

$$\mathrm{s} = a + \varepsilon_2,$$

where ε_2 is a normal distribution with mean 0 and variance σ_2^2 and is independent of condition a and ε_1. The sum of two independent normal distributions is also a normal distribution. So s is a normal distribution. Agent can exert effort e to reduce σ_2^2, i.e., the noise in the signal. Specifically, we assume

$$\sigma_2^2 = \frac{1}{e}.$$

When $e = +\infty$, the signal is perfectly informative about the condition a; however, when no effort is made ($e = 0$), the signal s becomes not informative at all. Similar as in the discrete model, the effect is costly for the agent. We assume that the cost is $\phi(e)$, where $\phi(e)$ is a continuously differentiable, strictly increasing and convex function with $\phi'(0) = 0$. Different from the discrete model, even after the agent exerts effort, he only gets a noisy signal about the condition a.

Similar as in the discrete model, we consider two cases, the one where no discrimination is allowed, and the one where discrimination is allowed. In the former case, the agent's diagnostic choice can only be based on the signal s, where its informativeness depends on his effort e. In the latter case, the agent's diagnostic choice can be based on both the signal s and the group identity, where informativeness of the signal also depends on his effort e.

In either case, after the agent makes an effort e, he needs to maximize his own expected payoff given a certain set of information *I*. In general, we have the following optimization problem to characterize the agent's diagnostic choice $\hat{a}$.

$$\underset{\hat{a}}{Min} \quad \lambda C \cdot E[(\hat{a} - a)^2 | I]$$

By the first order condition, we know that the optimal diagnostic choice is $\hat{a}(I) = E(a|I)$. The social welfare loss is then proportional to the conditional variance of a, i.e., $C \cdot Var(a|I)$.

In the case when no discrimination is allowed, the social welfare is $-C \cdot Var(a|s)$.

In the case when discrimination is allowed, the social welfare is $-C \cdot Var(a|s,r)$. By the properties of normal distribution (DeGroot, 1970), we have

$$Var(a|s) = \frac{\sigma_0^2 \cdot \sigma_2^2}{\sigma_0^2 + \sigma_2^2},$$

$$Var(a|s,r) = \frac{\sigma_0^2 \cdot \sigma_1^2 \cdot \sigma_2^2}{\sigma_0^2 \cdot \sigma_2^2 + \sigma_1^2 \cdot \sigma_0^2 + \sigma_2^2 \cdot \sigma_1^2}.$$

After rearrangement, we get

$$Var(a|s) = \frac{1}{e + \frac{1}{\sigma_0^2}},$$

$$Var(a|s,r) = \frac{1}{e + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_0^2}}.$$

For convenience, we define a new function

$$f(x) = -C \frac{1}{x + \frac{1}{\sigma_0^2}}.$$

This is exactly the social welfare as a function of the agent's effort. We know that the agent's optimal choice of effort e when no discrimination is allowed is captured by

$$e_N^* = \underset{e}{argMax} \quad \lambda f(e) - \phi(e).$$

When discrimination is allowed, the agent gets additional piece of information before he makes the choice of information acquisition. His net expected gain is therefore $\lambda f\left(e + \frac{1}{\sigma_1^2}\right) - \phi(e)$, where $\frac{1}{\sigma_1^2}$ reflects the informativeness of the group identity. Thus agent's optimal choice of effort when the discrimination is allowed is captured by

$$e_D^* = \underset{e}{argMax} \quad \lambda f\left(e + \frac{1}{\sigma_1^2}\right) - \phi(e).$$

Similarly, the best effort that maximizes the social welfare is pinned down by

$$e_B^* = \underset{e}{argMax} \quad f\left(e + \frac{1}{\sigma_1^2}\right) - \phi(e).$$

Notice that $f(x)$ is strictly increasing and strictly concave with $f'(0) > 0, f'(+\infty) = 0$.

Proposition 4*: When discrimination is not allowed, the agent exerts a higher effort to conduct the test than in the case when discrimination is allowed, i.e.,* $e_D^* < e_N^*$.

The result of the proposition is very intuitive. As one gets more endowments in the production, his free riding incentive arises so that he will have a less incentive to produce. Notice that both e_D^* and e_N^* can be pinned down by the first order conditions, namely we have

$$\lambda f'(e_N^*) = \phi'(e_N^*),$$

$$\lambda f'(e_D^* + \frac{1}{\sigma_1^2}) = \phi'(e_D^*).$$

We can prove Proposition 4 by contradiction. Suppose $e_D^* \geq e_N^*$. Because $f(\cdot)$ is strictly concave, we have $\lambda f'(e_D^* + \frac{1}{\sigma_1^2}) < \lambda f'(e_D^*) \leq \lambda f'(e_N^*) = \phi'(e_N^*)$. Since $\phi(e)$ is strictly convex, we have $\phi'(e_N^*) \leq \phi'(e_D^*)$. So we have $\lambda f'(e_D^* + \frac{1}{\sigma_1^2}) < \phi'(e_D^*)$, which is a contradiction. As a result, we must have $e_D^* < e_N^*$.

The following proposition shows when we can achieve the socially efficient outcome.

Proposition 5*: When* $\lambda = \frac{\phi'(e_B^*)}{f'(e_B^*)}$, *the effort chosen by the agent when no discrimination is allowed is the same as the socially efficient effort; that is,* $e_N^* = e_B^*$

First we need to check that $\frac{\phi'(e_B^*)}{f'(e_B^*)} < 1$. Notice that socially efficient effort e_B^* is captured by the first order condition

$$f'(e_B^* + \frac{1}{\sigma_1^2}) = \phi'(e_B^*).$$

Because $f(\cdot)$ is strictly concave, we have $f'(e_B^* + \frac{1}{\sigma_1^2}) < f'(e_B^*)$ so that $\frac{\phi'(e_B^*)}{f'(e_B^*)} < 1$.

Notice that $\lambda f'(e_N^*) = \phi'(e_N^*)$. By the strict concavity of $f(\cdot)$, the convexity of $\phi(\cdot)$, and the definition of λ, we conclude that e_N^* and e_B^* must be the same. Thus, we prove the proposition.

The following proposition shows that when the agent is less responsible for the social cost than is assumed in Proposition 5, not allowing discrimination can always improve social welfare.

Proposition 6*: When* λ is smaller than but close to $\frac{\phi'(e_B^*)}{f'(e_B^*)}$, *and* σ_1^2 *is sufficiently large, the outcome when no discrimination is allowed is more efficient than that when discrimination is allowed.*

Notice that the social welfare as a function of e is strictly concave. Also notice that $e_D^* < e_N^*$. We first show $e_B^* > e_N^*$ when $\lambda < \frac{\phi'(e_B^*)}{f'(e_B^*)}$. This result is obvious if we can show that $\frac{de_N^*}{d\lambda} > 0$. By taking the derivative with respect to λ in the first order condition of e_N^*, we get

$$\lambda f''(e_N^*)\frac{de_N^*}{d\lambda}+f'(e_N^*)=\frac{de_N^*}{d\lambda}\phi''(e_N^*).$$

After rearrangement, we have $\frac{de_N^*}{d\lambda}=\frac{f'(e_N^*)}{\phi''(e_N^*)-f''(e_N^*)}>0$. The social welfare when no discrimination is allowed is $f(e_N^*)-\phi(e_N^*)$ and the social welfare when discrimination is allowed is $f\left(e_D^*+\frac{1}{\sigma_1^2}\right)-\phi(e_D^*)$. From what we showed earlier,

$$f\left(e_N^*+\frac{1}{\sigma_1^2}\right)-\phi(e_N^*)>f\left(e_D^*+\frac{1}{\sigma_1^2}\right)-\phi(e_D^*),$$

and the difference between them is not going to zero as σ_1^2 goes to infinity and λ is smaller than but close to $\frac{\phi'(e_B^*)}{f'(e_B^*)}$. As σ_1^2 goes to infinity, $f(e_N^*)-\phi(e_N^*)$ is very close to $f\left(e_N^*+\frac{1}{\sigma_1^2}\right)-\phi(e_N^*)$. Therefore, $f(e_N^*)-\phi(e_N^*)>$ $f\left(e_D^*+\frac{1}{\sigma_1^2}\right)-\phi(e_D^*)$. We have thus proven the proposition.

5. Conclusions

In sum, this paper illustrates a new insight about the inefficiency of discrimination. The key assumption is the agent only bears part of the social cost of misdiagnosis. The socially optimal effort is for the agent to test everybody. If the agent is not allowed to discriminate, he is forced to choose between testing everybody or testing nobody. Facing this stark choice, it is optimal for him to test everybody even though he only bears part of the social cost of misdiagnosis. If the agent is allowed to discriminate, he will choose to test the group with higher probability of having the condition and not to test the other group, mainly because he only bears part of the social cost of misdiagnosis. Therefore, not allowing the agent to discriminate induces higher effort for both groups of subjects, resulting in higher social welfare than in the case in which discrimination is allowed.

Our result is based on a few assumptions. The first is that the social cost of reporting false negativity of any group of drug addicts exceeds the cost of the individual officer's effort; the second and the last state that the partial social cost that the agent has to bear for misdiagnosing one group is less than the effort the agency has to make to perform the right test, while the partial social cost of misdiagnosing both group is greater. Each of these assumptions is reasonable based on common sense, and together they lead us to the conclusion.

In the continuous case, given the agent's dislike for effort, the effort choice that is optimal to him is lower than the level that is socially optimal. By forbidding the agent from using group identity to make inference, we reduce the amount of information the agent starts with and therefore increase the marginal effect of gaining information from his effort, giving him more incentives to make an effort and partially offsetting the disincentive resulting from his not bearing all the social cost of misdiagnosis. Under certain conditions, the incentive effect of not allowing discrimination is so large that it enhances social efficiency.

此页开始为参考文献部分

参考文献：

Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." Industrial and Labor Relations Review 30 (2):175–87.

Arrow, Kenneth J. 1973. "The Theory of Discrimination." In Discrimination in Labor Markets, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton: Princeton University Press.

Becker, Gary S. 1958. "The Economics of Discrimination" Social Forces, Volume 37, Issue 2, December 1958, Pages 180–181

DeGroot, M. (1970), Optimal Statistical Decisions. New York: McGraw-Hill.

Fang, H. and Moro, Andrea. 2010. "Theories of Statistical Discrimination and Affirmative Action: A Survey" NBER Working Paper No. 15860

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." American Economic Review 62 (4): 659–61.

Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. 2017. "Meta-analysis of Field Experiments Shows No Change in Racial Discrimination in Hiring Over Time." PNAS October 10, 2017 114 (41) 10870-10875.

此页开始为简历部分

如果有必要，最后可以列出团队成员和指导老师的简历。

作者：白小舟 **Natalie Xiaozhou Bai**

清华附中国际部 12 年级学生

指导老师：陈济冬

2019 年起，清华大学公共管理学院副教授

2015 年-2019 年，北京师范大学经济与工商管理学院讲师

2009 年-2015 年，普林斯顿大学，2015 年获博士学位