Distribution Variability, Within-Group Competition and Statistical Discrimination

Xuejuan Su*

November 2005

Abstract

This paper develops a statistical discrimination model concerning the second order moment. It is shown that with heterogeneous individuals within a group, this form of statistical discrimination has differential impact on individuals. Specifically a belief that one group has a smaller variability may be self-fulfilling and persist without policy intervention, so two groups are identical ex ante but different ex post. The history-dependent belief offers an explanation why there may be smaller variability in observed qualifications among female, even though its underlying distribution of innate ability may be the same as male.

JEL code: J71, J24

Keywords: Statistical discrimination, Variability, Within-group competition

*Author address: Department of Economics, Finance and Legal Studies, the University of Alabama, Tuscaloosa, AL 35487-0224. Email address: xsu@cba.ua.edu.

1 Introduction

On January 14th 2005, at the NBER conference on diversifying the science and engineering workforce, the president of Harvard University, Lawrence H. Summers, advanced a hypothesis of different availability of aptitude at the high end (between male and female). He said, "It does appear that on many different human attributes - height, weight, propensity for criminality, overall IQ, mathematical ability, scientific ability - there is relatively clear evidence that whatever the difference in means - which can be debated - there is a difference in the standard deviation, and variability of a male and a female population. ... If one is talking about ... people who are three and a half, four standard deviations above the mean, ... even small differences in the standard deviation will translate into very large differences in the available pool substantially out." It caused an uproar in the media and the general public that later on February 15th he apologized.

All the political correctness aside, let us take a scientific view of such a hypothesis. It surely offers a theoretically possible explanation with regard to female under-representation in the top fraction of a given profession, even though the significance of such an explanation as compared to other alternatives, such as lack of recruiting and retaining efforts, is yet to be determined through careful and detailed empirical studies. After all, it is a statistical truism that smaller variability implies thinner tails far away from the mean, let it be left or right. This paper does not attempt to compare the relative importance of this hypothesis and its alternatives as the explanation for female under-representation at the top fraction of a profession. Instead the goal here is to construct a scenario where group variability is endogenously determined and public belief of group variability is self-fulfilling. So even if empirical studies find that female under-representation in the top fraction of a given profession is mainly correlated with its much smaller pool of available aptitude at the high end, the underlying causal relationship may still go in either direction. In short the data may not be able to help us distinguish two cases, whose policy implications are drastically different.

This observational equivalence arises only when attributes under concern are endogenously determined. When attributes are exogenously given or predetermined, public belief and individual decision has no effect on group distribution, so there is no ambiguity in inferring causality from correlation. Smaller variability and hence smaller availability of female aptitude at the high end would the cause, under-representation of female at the top fraction of a given profession would be the result. End of story. However when the attributes under concern are endogenously determined, the story changes drastically. For concreteness, from now on, we use the term qualification to refer to observed individual attribute that is most relevant to the labor market outcome, and we reserve the term ability to refer to innate ability, which is hard, if not impossible, to observe directly. Thus naturally an individual's qualification is a product of his innate ability and learning effort, with the latter depending critically on his expectation of the potential labor market outcome. In this case individuals with pessimistic expectation of their labor market outcomes tend to invest less in learning effort, so they have lower qualifications which reinforce the unfavorable outcomes. The self-fulfilling belief has been extensively studied in the existing statistical discrimination literature, whose main focus is on the difference between the mean values of two distinct groups. This is what we refer to as statistical discrimination based on the first order moment.

Relevant yet different, the focus of this paper is on the possibility of statistical discrimination based on the second order moment. The second order moment enters the center stage when there is heterogeneity within each of the two groups. When individuals differ in their innate ability, public belief of group variability may have differential impact on different individuals. More specifically, smaller variability of one group translates into thinner tails and fatter midsection, based on which rational individuals would be more pessimistic about extremely good labor market outcomes, yet more optimistic about reasonably good labor market outcomes. Since the extremely good labor market outcomes are more relevant to individuals with high innate ability, the pessimism there discourages them from investing as much learning effort. Similarly since the reasonably good labor market outcomes are more relevant to individuals with low innate ability, the optimism there encourages them to invest more learning effort. Compare these individuals to their counterparts in the other group where group variability is believed to be bigger, the qualifications of the high ability individuals are lower, and the qualifications of the low ability individuals are higher, thus the belief that there is smaller variability within this group is self-fulfilled.

To make the point as stark as possible, we start with two groups ex ante identical in the distribution of innate ability. There are three job sectors ranked in their desirability and efficiency wage rate. When there is public belief that one group exhibits smaller variability than the other, a rational individual forms expectation about his own probability of getting employed at each of the three sectors, based upon which he chooses the learning effort and hence qualification optimally. It is shown that the two groups may differ ex post in the distribution of observed qualification. When the two groups do differ ex post, it is obvious that the smaller pool of available aptitude at the high end does not cause the under-representation of one group at the top job sector. Instead both are results of the self-fulfilling public belief that there is smaller variability within the group.

The study of statistical discrimination was initiated in Arrow (1973) and Phelps (1972). Within the existing literature, there are two sources of statistical discrimination, both act through Bayesian update of belief on individual qualification, which cannot be observed perfectly. One is on the noise component of the signal, as studied in Aigner and Cain (1977) and Lundberg and Startz (1983). Bayesian updated belief on individual qualification is a weighted average of his own signal and group-wide average qualification, and the weight on his own signal is smaller when the signal is noisier. So if the noise component of the signal is bigger for one group than the other, individuals are less able to capture the benefit of investing learning effort in enhancing his own signal. Consequently the entire group has lower average qualification. Alternatively, statistical discrimination may arise when it is public belief that the group-wide average qualification is lower for one group than the other, even if the noise component of the signal is the same for both groups. This has been studied in Coate and Loury (1993). Again when one group is believed to have a lower mean, through the weighted average, individuals are less able to capture the benefit of investing learning effort in enhancing his own qualification, so that the group-wide average qualification is indeed lower, and the belief is self-fulfilled. These studies explain the difference in group mean, while the current paper offers an explanation in difference in group variability.

As common to all statistical discrimination models, some labor market friction, especially that from imperfect information, is needed to drive the results. Norman (2003) and Moro and Norman (2004) analyze the efficiency of statistical discrimination in a setup of worker/job match and mismatch. They show that depending on the trade-off between the informational gain of specialization and the loss in terms of higher investment cost,[1] statistical discrimination may generate potential efficiency gain. Without imperfect information and hence statistical discrimination, Fernandez and Gali (1999) compare the efficiency of market and tournament as the allocation mechanism in worker/job match. They

[1] The higher investment cost can be interpreted as associated with low ability individuals.

assume that the expenditure used to generate the signal in a tournament is not productive, so that with perfect credit market, tournament leads to the same efficient worker/job match as market but yields lower consumption and hence lower welfare. Su (2005) studies the within-group competition effect at the basic education stage as a tournament for the higher education opportunity. There learning effort is used to generate the noisy signal like a test score. Learning effort is productive in itself, in that it contributes to the human capital output from basic educations. Besides, learning effort is also valuable because it increases an individual's probability of winning the tournament and obtain higher education. In that case being allowed into the tournament as well as facing intense competition in the tournament increases aggregate efficiency measured as total human capital output. Following Su (2005), this paper also assumes that the costly effort in the tournament is not only productive in itself, but due to the within-group competition effect, also increases an individual's own probability of winning the tournament and obtain the desired worker/job match.

Unlike Oettinger (1996) where statistical discrimination may be eliminated through gradual learning of true individual qualification, in this paper statistical discrimination perpetuates. As belief is self-fulfilled, expectation of smaller group variability results in smaller group variability, so no additional information can be learned even when individual qualification is assumed to be perfectly observable after employment. The possibility of multiple equilibria consistent with multiple beliefs dictates that which equilibrium is actually played in an economy may be history-dependent. And without policy intervention, the history dependent statistical discrimination may perpetuate.

The remaining part of the paper is organized as follows. Section 2 describes the model. Section 3 solves the model and characterizes the equilibrium with between-group symmetry. Section 4 discusses the conditions under which equilibria without between-group symmetry may also exist. Given the system of highly non-linear equations governing the equilibrium, further analytical results are hard to obtain, so in Section 5 we produce a numerical example and solve for the equilibria, with or without between-group symmetry. Section 6 draws the conclusion.

2 The model

Consider an economy with two equal sized groups. The measure of each group is normalized to 1, so total population is 2. These two groups may be called Female and Male, with no intrinsic gender difference. More specifically, while innate ability is private information and only known to an individual himself/herself, the underlying distribution is the same within the female group as that within the male group. For expositional simplicity, it is assumed that there are only two types within each group. The high type is of measure λ and the low type is of measure $1-\lambda$, with the innate ability $a^H > a^L$.

Knowing his/her own type, an individual chooses the learning effort e to produce the qualification, which is productive in the labor market:

$$q = ae \tag{1}$$

While the two inputs, innate ability and learning effort, cannot be observed directly, the output, individual qualification, can be observed imperfectly with some random noise when the individual enters the labor market. More specifically, the noisy signal s is the sum of an individual's true qualification q and some random variable ε, with the cumulative distribution function $G(\varepsilon)$ and probability density function $g(\varepsilon) = G'(\varepsilon)$:

$$s = q + \varepsilon \tag{2}$$

It is further assumed that the random noise ε is independently and identically distributed across individuals with $E(\varepsilon|a) = 0$ and $E(\varepsilon^2|a) = \sigma_\varepsilon^2$. After the individual is employed, it is assumed that further information on his job performance perfectly reveals his productivity and hence the true qualification q.

On the demand side of the labor market, there are three distinct job sectors. Within each job sector there is a continuum of identical firms so that firms are perfectly competitive. It is assumed that the three job sectors are ranked in their job quality $w^1 > w^2 > w^3$, which we refer to as the top sector, the middle sector and the bottom sector respectively. There is limited capacity within each job sector, so that the top sector can employ a measure Ψ of the population, the middle sector a measure Φ, and the bottom sector a measure $2-\Psi-\Phi$. For simplicity it is assumed that the total capacity of all three job sectors is 2 and equal to the total population, so that there is no unemployment in the equilibrium.

With its intrinsic job quality w, each firm hires one worker, and the total output is determined by both job quality and worker qualification:

$$y = wq$$

After the employment decision is made, the worker's qualification can be perfectly observed during the working process, so that each worker is compensated by his true qualification, and the effective wage is simply w. The production technology being linear in worker qualification implies that in equilibrium a firm earns zero profit, and is indifferent in its employment decision.

In contrast to the firms' indifference with regard to workers, all workers have strict preference of the top sector over the middle sector, and the middle sector over the bottom sector. The difference in the job quality and hence effective wage across job sectors creates economy rent, and this is the source of competition among workers. Naturally we assume that a tournament is used to match workers with firms, so that ranked by their noisy signal s, the top fraction of Ψ fills the top sector, the next fraction of Φ fills the middle sector, and the remaining fraction $2 - \Psi - \Phi$ fills the bottom sector. Since the noisy signal s is all that can be observed and since the random noise ε is independently and identically across individuals, ex ante a tournament based on the signal is both fair and efficient, in that it generates positive assortative match between jobs and workers, and complementarity between the job quality and worker qualification implies that positive assortative match leads to maximum aggregate output.

In the current setup with production technology being linear in worker qualification, firms' employment decisions can be reduced to a mechanical rule easily justifiable by the zero profit. The story becomes much more complicated if, say, the production technology is concave in worker qualification so that firms use Bayesian rule to infer worker qualification from observed signal for profit maximization. It is difficult for two reasons. First, when individuals are heterogeneous in their qualification, knowing the distribution of the random noise ε and observing the distribution of the noisy signal s is not sufficient to deduce the distribution of true qualification q, unlike the case when individuals are homogeneous with a group so that the distribution of qualification q degenerates into a mass point equal to the mean of the noisy signal. Then for Bayesian update purpose, a prior belief of the distribution of qualification q is needed. If the prior belief is such that the distribution of qualification is identical across gender groups, then the rank ordering of the noisy signal in the tournament is consistent with the Bayesian rule. If on the other hand the prior belief is such that the distribution of qualification is different for the two gender groups, firms need to form belief and infer worker qualification from the noisy noise, and at the same time, workers need to form belief about firms' belief when making the decision on learning effort and hence qualification. In that case if we observe difference in the distribution of qualification ex post between the two groups, it is hard to pin down what fraction is attributed to the difference in the prior distribution and what fraction is attributed to the self-fulfilling belief. By adopting a mechanical rule in matching workers with firms, we bypass the prior distribution problem, so that any difference in the distribution of qualification is due to the self-fulfilling belief.

With the employment outcome determined by a tournament based on the noisy signal s only, all the focus of subsequent analysis is on the individual side. It is public belief that for a gender group $i \in \{F, M\}$, the signal distribution of the noisy signal is $H_i(.)$ with mean μ_i and variance $\sigma^2_{s,i}$. Since the random noise ε is uncorrected with the true qualification q, it then follows that $\sigma^2_{s,i} = \sigma^2_{q,i} + \sigma^2_{\varepsilon}$, namely a bigger variance of the true qualification is equivalent to a bigger variance of the signal within a gender group. Based on this belief, a non-discriminatory tournament imposes the same cut-off threshold $\widehat{s}^1$ and $\widehat{s}^2$ for both groups, so that the group specific capacity at each sector is given by $\Psi_i = 1 - H_i(\widehat{s}^1)$, $\Phi_i = H_i(\widehat{s}^1) - H_i(\widehat{s}^2)$ and $1 - \Psi_i - \Phi_i = H_i(\widehat{s}^2)$.

Essential to all statistical discrimination models, even though group identity is intrinsically "irrelevant", it must enter individual decision one way or another to make a difference in the equilibrium. Since firms actually don't make any decision in the tournament, the influence of the public belief can only act through individuals. Given the public belief of different variability in the distribution of signals, it is assumed that individuals believe the gender-specific capacity at each sector, namely Ψ_i, Φ_i and $1 - \Psi_i - \Phi_i$ as defined above to be an implicit quota. The implicit quota is an off-equilibrium belief in that the self fulfilling condition on belief in equilibrium implies that the implicit quota is not binding, namely the two group-specific capacity corresponds so the same cut-off threshold. However, off equilibrium the group-specific capacity may correspond to two different cut-off threshold for the two groups. As long as individuals believe that even with two different cut-off threshold for the two groups, there will not be adjustment of capacity between the groups, or equivalently they are only competing with their peers within the same group, they act accordingly and the intrinsically irrelevant group identity would have some real impact on the equilibrium.

Taking the cut-off thresholds $\widehat{s}^1$ and $\widehat{s}^2$ as given, if a high ability individual decides to exert effort e_i^H, then the perceived individual probability of getting employed at the top sector is determined as

$$\psi_i^H = \Pr(a^H e_i^H + \varepsilon \geq \widehat{s}^1) = 1 - G(\widehat{s}^1 - a^H e_i^H) \tag{3}$$

and the individual probability of getting employed at the middle sector is

$$\phi_i^H = \Pr(\widehat{s}^2 \leq a^H e_i^H + \varepsilon < \widehat{s}^1) = G(\widehat{s}^1 - a^H e_i^H) - G(\widehat{s}^2 - a^H e_i^H) \tag{4}$$

Hence the optimization problem for a high type individual is

$$\max \psi_i^H w^1 a^H e_i^H + \phi_i^H w^2 a^H e_i^H + (1 - \psi_i^H - \phi_i^H) w^3 a^H e_i^H + \theta \ln(1 - e_i^H) \tag{5}$$

Similarly for a low ability individual, the perceived individual probability of getting employed at the top sector is given by

$$\psi_i^L = \Pr(a^L e_i^L + \varepsilon \geq \widehat{s}^1) = 1 - G(\widehat{s}^1 - a^L e_i^L) \tag{6}$$

and the individual probability of getting employed at the middle sector is

$$\phi_i^H = \Pr(\widehat{s}^2 \leq a^L e_i^L + \varepsilon < \widehat{s}^1) = G(\widehat{s}^1 - a^L e_i^L) - G(\widehat{s}^2 - a^L e_i^L) \tag{7}$$

So the optimization problem faced by a low type individual is

$$\max \psi_i^L w^1 a^L e_i^L + \phi_i^L w^2 a^L e_i^L + (1 - \psi_i^L - \phi_i^L) w^3 a^L e_i^L + \theta \ln(1 - e_i^L) \tag{8}$$

From now on we focus on equilibrium with within-group symmetry, all that identical individuals within a group act the same. However we do not impose between-group symmetry, which may or may not arise in the equilibrium. Since the two gender groups are identical ex ante, any difference in the group distribution ex post can only be attributed to the public belief on group variability, and hence statistical discrimination based on the second moment.

Definition *For $i \in \{F, M\}$, an equilibrium consists of the public belief on group distribution $H_i(.)$, cut-off thresholds for the top sector and the middle sector $\{\widehat{s}^1, \widehat{s}^2\}$, individual effort levels $\{e_i^H, e_i^L\}$, and group-wide average effort levels $\{E_i^H, E_i^L\}$, such that:*

(1) Given $H_i(.)$ and $\{\widehat{s}^1, \widehat{s}^2\}$, e_i^H solves problem (5) and e_i^L problem (8);

(2) Within-group symmetry: $e_i^H = E_i^H$ and $e_i^L = E_i^L$;

(3) Self-fulfilling belief: $H_i(s) = \lambda G(s - a^H E_i^H) + (1 - \lambda) G(s - a^L E_i^L)$;

(4) Capacity constraint in the job sectors: $\Psi = 2 - H_F(\widehat{s}^1) - H_M(\widehat{s}^1)$ and $\Phi = H_F(\widehat{s}^1) - H_F(\widehat{s}^2) + H_M(\widehat{s}^1) - H_M(\widehat{s}^2)$.

3 Between-group symmetry

Since the two groups have identical distribution of innate ability ex ante, a natural benchmark case to study is when they also end up with identical distribution of qualification and hence signal ex post, i.e. when there is no statistical discrimination. Even without statistical discrimination, the within-group competition effect is critical in determining individual optimal learning effort, and similar story helps us understand the economic intuition when there is statistical discrimination in the next section.

Within-group competition effect captures the idea that an individual's own learning effort can improve his own probability of winning the tournament and getting employment at the desired job sector, even though the perceived group-wide capacity is fixed as an implicit quota. More specifically, for a high ability individual in group $i \in \{F, M\}$, the within-group competition effect for the top sector is given by

$$\frac{d\psi_i^H}{de_i^H} = g(\widehat{s}^1 - a^H e_i^H)a^H \tag{9}$$

and the within-group competition effect for the middle sector is given by

$$\frac{d\phi_i^H}{de_i^H} = (g(\widehat{s}^2 - a^H e_i^H) - g(\widehat{s}^1 - a^H e_i^H))a^H \tag{10}$$

Similarly for a low ability individual, the within-group competition effect for the top sector is given by

$$\frac{d\psi_i^L}{de_i^L} = g(\widehat{s}^1 - a^l e_i^L)a^L \tag{11}$$

and the within-group competition effect for the middle sector is given by

$$\frac{d\phi_i^L}{de_i^L} = (g(\widehat{s}^2 - a^L e_i^L) - g(\widehat{s}^1 - a^L e_i^L))a^L \tag{12}$$

It is obvious to see that the within-group competition effect is intrinsically linked to the probability density of the random noise at the cut-off threshold. The higher is the probability density, the more pronounced is the marginal effect of increasing own learning effort by infinitesimal and outperforming other individuals on the threshold, and hence the bigger competition effect. In conventional notion of a bell-shaped probability density function, if the cut-off threshold is at the far right tail, the tournament is too hard to win and people more or less give up. If the cut-off threshold is at the far left tail, the tournament is too easy to win and people also take it easy. Only when the cut-off threshold is close to the middle, is there big stake and hence intense competition. This competition effect is explicitly captured in this model.

As can be easily seen, the competition effect to get employment at the top sector is always positive (nonnegative), in that an individual can almost always improve his own probability by investing a little bit more learning effort. On the other hand, the competition effect to get employment at the middle sector may turn negative when the individual is already high above the threshold, and further increasing in learning effort makes it more likely to get employment at the top sector and out of the middle sector.

Viewed as an opportunity, it is natural to expect that the top sector is more relevant to the high ability type individuals, who may view the middle sector as a fallback option; and the middle sector is more relevant to the low ability type individuals, who can only use the bottom sector as the fallback option. For exposition simplicity, from now on it is assumed that the random noise ε is distributed on a finite support $[\underline{\varepsilon}, \overline{\varepsilon}]$, and the maximum difference in the realized values, $\overline{\varepsilon} - \underline{\varepsilon}$, is small compared to the difference between the two ability types, $a^H - a^L$. This way it can be ensured that even with the worst luck $\underline{\varepsilon}$, a high ability individual does not fall below the cut-off threshold for the middle sector $\widehat{s}^2$; and even with the best luck $\overline{\varepsilon}$, a low ability individual does not meet the cut-off threshold for the top sector $\widehat{s}^1$. This stratification of the two ability types over the three job sectors greatly simplify our exposition, but it is not a critical assumption in that main results are robust to alternative specification where the random noise can lead both ability types to all three job sectors. What is essential is that ex ante, the top sector is more relevant to the high ability type and the middle sector is more relevant to the low ability type as an opportunity.

Within a group, the stratification of the two ability types across the three job sectors - namely high ability individuals only end up in the top and the middle sector, and the low ability individuals only end up in the middle and the bottom sectors - implies that $\widehat{s}^1 > a^L E_i^L + \overline{\varepsilon}$ and $\widehat{s}^2 < a^H E_i^H + \underline{\varepsilon}$. Consequently, the competition effects for a high ability individual (9) and (10) become $\frac{d\psi_i^H}{de_i^H} = g(\widehat{s}^1 - a^H e_i^H) a^H$ and $\frac{d\phi_i^H}{de_i^H} = -g(\widehat{s}^1 - a^H e_i^H) a^H$, and the competition effects for a low ability individual (11) and (12) become $\frac{d\psi_i^L}{de_i^L} = 0$ and $\frac{d\phi_i^L}{de_i^L} = g(\widehat{s}^2 - a^L e_i^L) a^L$. From now on, we use $\widehat{\varepsilon}_i^1 = \widehat{s}^1 - a^H e_i^H$ to denote the relevant cut-off threshold faced by a high ability individual to get employment in the top sector, and $\widehat{\varepsilon}_i^2 = \widehat{s}^2 - a^L e_i^L$ the relevant cut-off threshold faced by a low ability individual to get employment in the middle sector.

Now for a high ability individual, the first order condition on e_i^H is

$$w^2 a^H + (1 - G(\widehat{\varepsilon}_i^1))(w^1 - w^2)a^H + g(\widehat{\varepsilon}_i^1)a^H(w^1 - w^2)a^H e_i^H = \frac{\theta}{1 - e_i^H} \quad (13)$$

Intuitively, the right-hand side term $\frac{\theta}{1-e_i^H}$ represents the marginal cost of learning effort, which is the foregone leisure. The left-hand side represents the marginal benefit, which can be decomposed into three distinct terms. The first term $w^2 a_i^H$ reflects how the learning effort produces worker qualification, which contributes to a high ability individual's wage income even in the worst case that he gets employed at the middle sector. The second term $(1 - G(\widehat{\varepsilon}_i^1))(w^1 - w^2)a^H$ shows how with probability $1 - G(\widehat{\varepsilon}_i^1)$, the individual gets employed at the top sector, so his learning effort contributes to his wage income by the incremental value $(w^1 - w^2)a^H$, which can be viewed as the economic rent from employment at the top sector compared to the middle sector. The third term $g(\widehat{\varepsilon}_i^1)a^H(w^1 - w^2)a^H e_i^H$ captures the within-group competition effect, where putting more learning effort increases individual probability of getting employment at the top sector by $g(\widehat{\varepsilon}_i^1)a^H$, whose incremental value is $(w^1 - w^2)a^H e_i^H$.

Similarly for a low ability individual, the first order condition on e_i^L is

$$w^3 a^L + (1 - G(\widehat{\varepsilon}_i^2))(w^2 - w^3)a^L + g(\widehat{\varepsilon}_i^2)a^L(w^2 - w^3)a^L e_i^L = \frac{\theta}{1 - e_i^L} \quad (14)$$

And the same interpretation can be attached to the three components of the marginal benefit of the learning effort. These two first order conditions for high ability and low ability types are the central equations of the analysis, and we will return to them frequently.

In an equilibrium with both within-group symmetry and between-group symmetry, the total capacity of the top sector is evenly split between the two groups. Since identical individuals behave the same within a group, it implies that the perceived individual probability equals group-wide capacity $\psi_i^H = \frac{\Psi}{2\lambda}$ and $\phi_i^L = \frac{\Psi+\Phi-2\lambda}{2(1-\lambda)}$, and the same cut-off threshold for the same type of individuals $\widehat{\varepsilon}_i^1 = G^{-1}(1 - \frac{\Psi}{2\lambda})$ and $\widehat{\varepsilon}_i^2 = G^{-1}(1 - \frac{\Psi+\Phi-2\lambda}{2(1-\lambda)})$. In this special case with between-group symmetry, an equilibrium can be determined by the solution of the system of two equations (13) and (14) with two unknowns e_i^H and e_i^L. In a more general case without between-group symmetry, an equilibrium can be determined by the solution of the system of eight equations and eight unknowns, as will be shown in the next section.

Assumption 1. $\theta < w^3 a^L$

The Inada condition on leisure guarantees that e_i^H and e_i^L are bounded below 1. Assumption 1 implies that $\theta < w^3 a^L < w^2 a^H$, so e_i^H and e_i^L as the solutions to (13) and (14) are also bounded above 0. This assumption simply says that leisure is not valued too much compared to learning and hence consumption, so that an individual finds it worthwhile to invest learning effort in improving his qualification.

Proposition 1 *An equilibrium with between-group symmetry exists. In this equilibrium there is no statistical discrimination, and the two groups are identical ex ante and ex post.*
Proof. *See the Appendix.* ■

This result is standard. A symmetric equilibrium exists where identical individuals behave alike, either within a group or between groups. Interestingly, in the next section we show that there may also exist other equilibria where identical individuals behave alike only within a group, but differently between groups. So despite the between-group symmetry ex ante, there may be between-group asymmetry ex post. When this is the case, between-group symmetry holds on a higher order between pairs of equilibria, if not within an equilibrium.

4 Between-group asymmetry

The first order conditions (13) and (14) show how individual optimal learning effort depends on: (1) the perceived individual probability; and (2) the within-group competition effect for getting employment at the desired job sector. For any individual, while the marginal cost curve for his learning effort depends only on his preference and remains fixed with regard to public belief of group distribution, the marginal benefit curve does respond to a change in public belief. For example, if it is public belief that there is smaller group variability for female than male, and if the top sector is highly competitive, this belief leads a high ability individual to expect that the group capacity at the top sector is smaller for female than that for male. On this account, the marginal benefit curve is lower for a high ability female than a high ability male. On the other hand, how the public belief affects the within-group competition depends on whether it is too hard or too easy to win the tournament. If it is too hard to win, namely the cut-off threshold is on the right tail, then a smaller group capacity for female makes it even harder for a female to win against her female peers, so the smaller competition effect further lowers the marginal benefit curve. If it is too easy to win, then a smaller group capacity for female makes it somewhat harder for a female to win, and the bigger competition effect raises the marginal benefit curve. The net effect would then depend on which shift dominates, the downward shift due to smaller group capacity or the upward shift due to bigger competition effect. The optimal learning effort is lower for a high ability female if the marginal benefit curve is lower and vice versa.

As for a low ability individual, if the middle sector is moderately competitive, then the public belief of smaller group variability for female may imply that the group capacity at the middle sector is bigger for female than that for male. On this account the marginal benefit curve is higher for a low ability female than a low ability male. Again how the public belief affects the within-group competition depends on whether it is too hard or too easy to win the tournament. If it is hard to win, a bigger group capacity for female makes it less hard, so the bigger competition effect further raises the marginal benefit curve. If it is easy to win, a bigger group capacity for female makes it even easier, and the smaller competition effect lowers the marginal benefit curve. The net effect again depends on which shift dominates, the upward shift due to bigger group capacity or the downward shift due to smaller competition effect. The optimal learning effort is higher for a low ability female if the marginal benefit curve is higher and vice versa.

Assumption 2. The probability density function $g(\varepsilon)$ is unimodal with the peak occurring at $\widetilde{\varepsilon}$. Furthermore the decreasing density part $(\widetilde{\varepsilon},\ \overline{\varepsilon}]$ is relatively big and/or the job sector capacity Ψ and Φ are relatively small.

This assumption is not necessary; however it does help reach some unambiguous results. It ensures that the change in the competition effect is in the same direct as the change in the group capacity due to different public belief, so that there is a definite shift of the marginal benefit curve. Now public belief of smaller group variability for female lowers the marginal benefit curve and leads to lower learning effort for high ability female; at the same time, it raises the marginal benefit curve and leads to higher learning effort for low ability female. Overall the high ability female ends up with lower qualification, while the low ability female ends up with higher qualification than their male counterparts, so that the public belief of smaller group variability for female is self-fulfilled.

Proposition 2 *If equilibria without between-group symmetry exist, they always exist in pairs with switched group indices.*

Proof. *See the Appendix.* ■

This proposition only characterizes the property if equilibria without between-group symmetry exist. It does not prove the existence of such equilibria. Given the system of eight equations with eight unknowns, some of which are highly non-linear, it is hard to characterize the conditions such equilibria without between-group symmetry exist analytically. In the next section numerical analysis is used to show that these equilibria do exist in some examples.

If equilibria without between-group symmetry exist, the public belief that groups differ in their signal distribution $H_i(.)$ is self-fulfilled. Interesting enough, in our specification, since firms do not use the Bayesian rule to infer worker qualification but instead use a tournament based on the noisy signal to make employment decisions, the first order moment μ_i of the group mean is only of the second order importance, while the second order moment $\sigma^2_{s,i}$ of group variability is now of the firsts order importance. That is why we refer to this type of statistical discrimination as mainly based on the second order moment.

More specifically, suppose one equilibrium is such that $\sigma^2_{s,F} < \sigma^2_{s,M}$ as discussed above, where the high ability female invests less learning effort and has lower qualification, while low ability female invests more learning effort and has higher qualification than their male counterparts. In this case it is all possible that the group mean of female is higher than, equal to or lower than that of male, namely $\mu_F \gtreqless \mu_M$. Unlike statistical discrimination based on the first order moment, where a higher group mean benefits all individuals in the group uniformly while a lower group mean hurts all individuals in the group, here group mean matters little, while the group variability generates differential impact on different individuals within the group. The high ability individuals are hurt by the belief of smaller group variability, while the low ability individuals benefit from the belief of smaller group variability.

There are two critical points here supporting equilibria without between-group symmetry. First, individual qualification cannot be directly observed, so public belief plays an important role in determining the expected payoff and hence affects individual learning effort. Statistical discrimination takes a more subtle form when there is heterogeneity within a group. Public belief on group variability has differential impact on different individuals within the group, and this self-fulfills the belief on group variability. Within-group heterogeneity brings the second order moment to the center stage, which has been largely overlooked in previous analysis.

Next, it is essential that public belief on group variability leads individuals to think the group capacity at the desired job sector as an implicit quota, even though in equilibrium the cut-off threshold on signal appears identical for both groups and the implicit quota does not appear binding. As an off-equilibrium belief, the implicit quota allows only within-group competition but forbids between-group competition, so an individual's expected payoff depends only on his relative standing within his own group but not that in the entire population. Hence different perceived group capacity implies different competition effect for the two groups, and this off equilibrium belief is essential to support an equilibrium without between-group symmetry. And since the implicit quota does not appear binding in the equilibrium, this form of statistical discrimination is hard to detect in the data.

Of course the intrinsic symmetry between the two groups dictates that equilibria without between-group symmetry always exist in pairs if at all, being mirror image of each other with switched group indices. So if there is an equilibrium where group variability is smaller for female than male, there is also a corresponding one where group variability is smaller for male than female. Which equilibrium is actually played may be history dependent. When the top sector used to be dominated by male, the self-fulfilling belief persists without policy intervention, so female will always be underrepresented at the top sector, even though the underlying distribution of innate is the same for male and female. This possibility will be shown in a numerical example in the next section.

5 Numerical example

Now we produce an numerical example to show the existence of equilibria without between-group symmetry. Let the distribution of innate ability within each group be given by $a^H = 7, a^L = 5$ and $\lambda = 0.2$. The top sector is highly competitive and has a small capacity of $\Psi = 0.1$, and the middle sector is moderately competitive and has a capacity of $\Phi = 1$. It is also assumed that the relative weight individuals put on leisure is $\theta = 5$, and the effective wage rates at the three sectors are $w^1 = 3$, $w^2 = 2$ and $w^3 = 1$.

The random noise ε is symmetrically distributed around the origin with the following PDF:

$$g(\varepsilon) = \begin{cases} 2\varepsilon^3 - 3\varepsilon^2 + 1 \ for \ \varepsilon \in [0,1] \\ -2\varepsilon^3 - 3\varepsilon^2 + 1 \ for \ \varepsilon \in [-1,0) \\ 0 \ otherwise \end{cases}$$

This PDF is bell-shaped, continuously differentiable on $(-\infty, \infty)$ and peaked at $g(0) = 1$.

Equilibrium 1. With between-group symmetry

In the equilibrium with between-group symmetry, the total capacity of a job sector is evenly split between the two groups. So within a group, a high ability individual faces the cut-off threshold $\widehat{\varepsilon}_i^1 = G^{-1}(1 - \frac{\Psi}{2\lambda}) = 0.2656$, while a low ability individual faces the cut-off threshold $\widehat{\varepsilon}_i^2 = G^{-1}(1 - \frac{\Psi+\Phi-2\lambda}{2(1-\lambda)}) = 0.0625$. Consequently, the optimal learning effort can be computed from (13) and (14), so we have $e_i^H = 0.9045$ and $e_i^L = 0.8175$. Regardless of group indices, a high ability individual has the qualification $q_i^H = 6.3314$, and a low ability individual has the qualification $q_i^L = 4.0874$, so of the distribution of qualification, the group mean is $\mu = 4.5362$, and the group variability (standard deviation) is $\sigma_q = 0.8976$. Both groups face the same cut-off thresholds on the noisy signal $\widehat{s}^1 = 6.5971$ and $\widehat{s}^2 = 4.1499$. The two groups are identical ex ante and ex post, and there is no statistical discrimination.

Equilibrium 2. high ability female underrepresented at the top sector

However, equilibria without between-group symmetry also exist. One equilibrium is such that female is underrepresented in the top sector with $\Psi_F = 0.0185 < 0.05 = \frac{\Psi}{2}$, and male is overrepresented in the top sector with $\Psi_M = \Psi - \Psi_F = 0.0815$. A high ability female hence faces the cut-off threshold $\widehat{\varepsilon}_F^1 = G^{-1}(1 - \frac{\Psi_F}{\lambda}) = 0.5039$, while a high ability male faces the cut-off threshold $\widehat{\varepsilon}_M^1 = G^{-1}(1 - \frac{\Psi_M}{\lambda}) = 0.0938$, where the smaller group capacity also implies smaller within-group competition effect for female. Overall the marginal benefit curve is lower for a high ability female than a high ability male, and she invests less learning effort than her male counterpart. From (13) it can be computed that $e_F^H = 0.8589$ and $e_M^H = 0.9176$, so that high ability female has lower qualification than high ability male, $q_F^H = 6.0125$ and $q_M^H = 6.4234$. Adding the cut-off values of the random noise, the high ability female and male face essentially identical cut-off threshold $\widehat{s}^1 = 6.5164(F) \approx 6.5172(M)$, with the difference 0.01% and well within the margin of error of the numerical approximation.

Given that high ability female underrepresented at the top sector, more of them will end up in the middle sector. So an evenly split group capacity at the middle sector would still imply low ability female underrepresented due to the trickle down effect. Taking the asymmetric outcome for the high ability female and male at the top sector as given, there are three possible outcomes for the low ability female and male at the middle sector, each constitutes an equilibrium without between-group symmetry.

Equilibrium 2.1. low ability female underrepresented at the middle sector

The first one is given by $\Phi_F = 0.2637$, so not only low ability female, but overall female is underrepresented in the middle sector. A low ability female hence faces the cut-off threshold $\widehat{\varepsilon}_F^2 = G^{-1}(1 - \frac{\Psi_F + \Phi_F - \lambda}{1-\lambda}) = 0.4824$, while a low ability male faces the cut-off threshold $\widehat{\varepsilon}_M^1 = G^{-1}(1 - \frac{\Psi_M + \Phi_M - \lambda}{1-\lambda}) = -0.2930$. From (14) it can be computed that $e_F^L = 0.6418$ and $e_M^L = 0.7973$, so that low ability female also has lower qualification than low ability male, $q_F^L = 3.2093$ and $q_M^L = 3.9864$. Adding the cut-off values of the random noise, the low ability female and male also face essentially identical cut-off threshold $\widehat{s}^2 = 3.6917(F) \approx 3.6934(M)$, with the difference 0.05% and well within the margin of error of the numerical approximation.

Equilibrium 2.2. low ability female exactly represented at the middle sector

The second is given by $\Phi_F = 0.5315$, so low ability female is exactly represented in the middle sector, while overall female is overrepresented due to the trickle down effect. A low ability female hence faces the cut-off threshold $\widehat{\varepsilon}_F^2 = G^{-1}(1 - \frac{\Psi_F + \Phi_F - \lambda}{1-\lambda}) = 0.0625$, which is the same for a low ability male. From (14) it hence follows that $e_F^L = e_M^L = 0.8175$, and $q_F^L = q_M^L = 4.0874$, and both groups face the identical cut-off threshold $\widehat{s}^2 = 4.1499$.

Equilibrium 2.3. low ability female overrepresented at the middle sector

The third one is given by $\Phi_F = 0.7993$, so not only low ability female, but overall female is overrepresented in the middle sector. This is actually the mirror image of the first one. Now a low ability female hence faces the cut-off threshold $\widehat{\varepsilon}_F^2 = G^{-1}(1 - \frac{\Psi_F + \Phi_F - \lambda}{1-\lambda}) = -0.2930$, while a low ability male faces the cut-off threshold $\widehat{\varepsilon}_M^1 = G^{-1}(1 - \frac{\Psi_M + \Phi_M - \lambda}{1-\lambda}) = 0.4824$. From (14) it can be computed that $e_F^L = 0.7973$ and $e_M^L = 0.6418$, so that low ability female has higher qualification than low ability male, $q_F^L = 3.9864$ and $q_M^L = 3.2093$. Adding the cut-off values of the random noise, the low ability female and male also face essentially identical cut-off threshold $\widehat{s}^2 = 3.6934(F) \approx 3.6917(M)$, with the difference 0.05% and well within the margin of error of the numerical approximation.

Based on the equilibria 2.1, 2.2 and 2.3, from proposition 2 it is easy to switch the group indices and find equilibria 3.1, 3.2 and 3.3, where high ability male is underrepresented at the top sector, while low ability male is under-, exactly, or overrepresented at the middle sector. Furthermore, there are equilibria where the high ability female is exactly represented at the top sector, while low ability female is either under- or overrepresented at the middle sector.

Equilibrium 3. high ability female exactly represented at the top sector

As in the equilibrium with between-group symmetry, we have $\widehat{\varepsilon}_F^1 = \widehat{\varepsilon}_M^1 = G^{-1}(1 - \frac{\Psi}{2\lambda}) = 0.2656$, $e_F^H = e_M^H = 0.9045$, $q_F^H = q_M^H = 6.3314$, and both groups face the same cut-off thresholds on the noisy signal $\widehat{s}^1 = 6.5971$.

Equilibrium 3.1. low ability female underrepresented at the middle sector

The first one is given by $\Phi_F = 0.2322$, so a low ability female hence faces the cut-off threshold $\widehat{\varepsilon}_F^2 = G^{-1}(1 - \frac{\Psi_F + \Phi_F - \lambda}{1-\lambda}) = 0.4824$, while a low ability male faces the cut-off threshold $\widehat{\varepsilon}_M^1 = G^{-1}(1 - \frac{\Psi_M + \Phi_M - \lambda}{1-\lambda}) = -0.2930$. Just like Equilibrium 2.1, we have $e_F^L = 0.6418$ and $e_M^L = 0.7973$, so that low ability female also has lower qualification than low ability male, $q_F^L = 3.2093$ and $q_M^L = 3.9864$. Adding the cut-off values of the random noise, the low ability female and male also face essentially identical cut-off threshold $\widehat{s}^2 = 3.6917(F) \approx 3.6934(M)$, with the difference 0.05% and well within the margin of error of the numerical approximation.

Equilibrium 3.2. Low ability female overrepresented at the middle sector

The third one is given by $\Phi_F = 7678$, so a low ability female hence faces the cut-off threshold $\widehat{\varepsilon}_F^2 = G^{-1}(1 - \frac{\Psi_F + \Phi_F - \lambda}{1-\lambda}) = -0.2930$, while a low ability male faces the cut-off threshold $\widehat{\varepsilon}_M^1 = G^{-1}(1 - \frac{\Psi_M + \Phi_M - \lambda}{1-\lambda}) = 0.4824$. Just like Equilibrium 2.3, we have $e_F^L = 0.7973$ and $e_M^L = 0.6418$, so that low ability female has higher qualification than low ability male, $q_F^L = 3.9864$ and $q_M^L = 3.2093$. Adding the cut-off values of the random noise, the low ability female and male also face essentially identical cut-off threshold $\widehat{s}^2 = 3.6934(F) \approx 3.6917(M)$, with the difference 0.05% and well within the margin of error of the numerical approximation.

Totally there are 9 equilibria, with high ability female being under-, exactly and overrepresented at the top sectors matched to low ability female being under-exactly and overrepresented at the middle sector. The equilibria are summarized in Table 1. The three column represent the status of high ability female at the top sector, with (-) denoting under-, (0) denoting exactly, and (+) denoting overrepresentation. Similarly the three rows represent the status of low ability female at the middle sector. We report 5 out of the 9 total equilibria, and the remaining 4 can be easily deduced by switching group indices. Of special

	F - high (-)	F - high (0)	F - high (+)
F - low (-)	$\mu_F < \mu_M, \sigma_F > \sigma_M$	$\mu_F < \mu_M, \sigma_F > \sigma_M$	
F - low (0)	$\mu_F < \mu_M, \sigma_F < \sigma_M$	$\mu_F = \mu_M, \sigma_F = \sigma_M$	
F - low (+)	$\mu_F > \mu_M, \sigma_F < \sigma_M$		

Table 1 Equilibria Summary Results

interest to us is Equilibrium 2.3, namely the third entry on the first column. This is the equilibrium where statistical discrimination leads to smaller variability among female; and despite the higher mean, female is underrepresented at the top sector while overrepresented at the middle sector. Put in other words, the phenomenon described in Larry Summer's hypothesis can arise endogenously due to public belief on group distribution of qualification, even though the two groups may be of identical group distribution of innate ability.

6 Conclusion

This paper has developed a statistical discrimination model based mainly on the second order moment. Multiple equilibria may exist differing on the public belief of group distribution. When two groups have identical distribution of innate ability ex ante, they may end up with different distribution of qualification ex post. When statistical discrimination does exist, it may have differential impact on heterogeneous individuals within a group. More specifically, when it is public belief that there is smaller variability of a group, the top fraction of that group is adverse affected and discouraged from invest learning effort in improving its qualification, while the bottom fraction benefits and invests more learning effort in improving its qualification, so the belief of smaller variability is self-fulfilled. Given the multiple equilibria, which equilibrium is actually played in the economy may be hisory dependent, statistical discrimination may perpetuate without policy interventiton.

Appendix

Proposition 1.

Proof. It is obvious that e_i^H can solved from equation (13) and e_i^L from equation (14) individually. For either of these equations, Assumption A1 implies that the left hand side (LHS) is bigger than the right hand side (RHS) when the learning effort is close to 0, while the Inada condition on leisure implies that the LHS is smaller than the RHS when the learning effort is close to 1. Since the marginal benefit curve crosses the marginal cost curve from above, the solution is a maximum and hence an equilibrium. ■

Proposition 2.

Proof. An equilibrium without between-group symmetry consists of eight unknown variables $\{e_i^H, e_i^L, \widehat{\varepsilon}_i^1, \widehat{\varepsilon}_i^2\}$ for $i \in \{F, M\}$, and a system of eight equations:

$$a^H e_F^H + \widehat{\varepsilon}_F^1 = a^H e_M^H + \widehat{\varepsilon}_M^1$$

$$a^L e_F^L + \widehat{\varepsilon}_F^2 = a^L e_M^L + \widehat{\varepsilon}_M^2$$

$$\lambda(2 - G(\widehat{\varepsilon}_F^1) - G(\widehat{\varepsilon}_M^1)) = \Psi$$

$$(1-\lambda)(2 - G(\widehat{\varepsilon}_F^2) - G(\widehat{\varepsilon}_M^2)) = \Psi + \Phi - 2\lambda$$

together with the first order conditions (13) and (14) for $i \in \{F, M\}$. Here the first two equations imply that the belief is self-fulfilled, so that even though individuals perceive the group capacity as implicit quota, the implicit quota does not appear binding in the equilibrium. And the second two equations are the capacity constraint at the top and the middle job sectors.

Since the equations are non-linear, there may be multiple solutions, one corresponding to the equilibrium with between-group symmetry, while the remaining corresponding to other equilibria without between-group symmetry. Since there is no fundamental difference between the two groups, it is easy to see that equilibria without between-group symmetry always exist in pairs, as mirror image of each other with switched group indices. ■

References

Aigner, Dennis J. and Cain, Glen C. "Statistical Theories of Discrimination in the Labor Market," Industrial and Labor Relations Review, January 1977, 30, pp. 175-87.

Arrow, Kenneth J. "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., Discrimination in Labor Markets, Princeton, NJ: Princeton Press, 1973, pp. 3-33.

Coate, Stephen and Loury, Glenn C. "Will Affirmative Action Policies Eliminate Negative Stereotypes?" American Economic Review, December 1993, 83(5), pp. 1220-40.

Fernandez, Raquel and Gali, Jordi. "To Each According to...? Markets, Tournaments, and the Matching Problem with Borrowing Constraints." Review of Economic Studies, October 1999, 66(4), pp. 799-824.

Lundberg, Shelly J. and Startz, Richard. "Private Discrimination and Social Intervention in Competitive Labor Market," American Economic Review, June 1983, 73(3), pp. 340-47.

Moro, Andrea and Norman, Peter. "A General Equilibrium Model of Statistical Discrimination," Journal of Economic Theory, 2004, 114, pp. 1-30.

Norman, Peter. "Statistical Discrimination and Efficiency," Review of Economic Studies, 2003, 70, pp. 615-27.

Oettinger, Gerald S. "Statistical Discrimination and the Early Career Evolution of the Black-White Wage Gap," Journal of Labor Economics, January 1996, 14(1), pp. 52-78.

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

Su, Xuejuan. "Education Hierarchy, Within-Group Competition and Affirmative Action," University of Alabama working paper, 2005.