Holdup and hiring discrimination with search friction

BI, Sheng and LI, Yuanyuan

University of Bielefeld, University of Paris 1

15 June 2015

Holdup and hiring discrimination with search friction*

BI Sheng[†], LI Yuanyuan[‡]

June 17, 2015

Abstract

A holdup problem on workers' skill investment arises when employers can adopt discriminatory hiring norm to extract higher than socially optimal profit. In such an economy, productivity (skills) and non-productivity oriented characteristics (discrimination) both matter when determining which worker has priority. The resulting firms' preference is an intertwined ranking order, by virtue of which the strategic interdependence in skill choices between discriminated and favored groups endogenously arises. We consider frictional markets with either posted or bargained wage. With posted wage, discrimination makes all workers worse off, firms gain. Through that payoff interdependence, we identify two effects along which one group's underinvestment may benefit all groups. With bargained wage, the discriminated (favored) group is always worse (better) off, and firms incur cost for an intermediated range of bargaining power when they discriminate. This suggests that the holdup-discrimination problem can be mitigated when search is random and wages bargained, a result in the opposite direction of Acemoglu and Shimer (1999b).

JEL Classifications: J7, J42

Key words: Discrimination; Directed Search; Pre-matching Investment

*We are very grateful for the discussions and encouragement from Professor François Langot, Professor Bernhard Eckwert, Professor Bertrand Wignoille, Professor Anna Zaharieva, Dr. Andreas Szczutkowski, Nikolai Brandt, as well as participants in BIGSEM Colloquium, BIGSEM Workshop, EDEEM Annual meeting, and GAINS seminar. Yuanyuan LI would like to thank the EDEEM program for the financial support, and BI would like to thank for the financial sponsorship from University of Paris 1 and University of Bielefeld, and Deutsch-Französishe Hoshschule.

[†]University of Paris 1, and University of Bielefeld; E-mail: brbisheng@gmail.com

[‡]University of Bielefeld, and University of Paris 1

1 Introduction

A holdup problem arises when some investment is sunked ex ante by one party, and the payoff is shared with that one party's trading partner. Since cost has no other use once sunk, that trading partner will have every incentive to squeeze the profit at the ex post stage. In an important study on such a problem in a labor market with search friction, Acemoglu and Shimer (1999b) shows that with firms' sunking capital and ex post wage bargaining, the equilibrium is always inefficient, since wages paid ex post can be so high such that the firms' ex ante incentive of investment is detrimented; while if firms are able to post wages to direct workers' search, then the holdup problem to firms' investment no longer appears; the efficiency can be achieved, because wage posting allows workers to observe offers and choose where to apply, and it induces workers to optimize their expected payoff from application by making trade-off between *every wage they observe* and the probability of obtaining it.

Within conventional wage posting framework, we emphasize a new source of inefficiency in a holdup problem where workers sunk skill investment cost: by adopting a discriminatory hiring norm, firms are able to expropriate higher than socially optimal level of profit, discouraging the investment incentives for *both the undiscriminated and discriminated groups*. This leaves room for welfare improvement if search is random and wages bargained ex post. In contrast to wage posting, we find an intermediate range of bargaining power for the values of which firms incur loss with discrimination; although discriminated group is always worse off, the favored group is always better off and "free-rides" the discriminated group on skill choice. We analyze the impact of such rent seeking behavior of firms on the structure of market segmentation, on welfare, and on the workers' skill investment incentives.

A key feature of our study is the multidimensionality of characteristics, based on which workers are ranked. On one hand, there is ranking by productivity oriented type identity: workers are either high skilled (type H) or low skilled (type L); high skilled is has priority to low skilled simply because profit is increasing in productivity. On the other hand, there is ranking by non-productivity oriented group identity: workers belong either to the favored (group a) or the discriminated group (group b). The resulting ranking schedule has the following order: $aH \succ bH \succ aL \succ bL$. Only the part $bH \succ aL$ may raise contention: the discriminated high skilled workers (bH) are ranked in priority to favored low skilled workers (aL). However, this is not due to any assumption that ranking by productivity is a stronger order, or that discrimination takes place conditional on any skill type; this results simply endogenously from any context with search friction where firms' objective is to maximize profit. Under such an "intertwined" ranking order, the skill investment decision for different groups become strategically interdependent. And we think this observation is central for the understanding of further results.

We start by considering the case where firms are not allowed to select workers according to productivity-irrelevant characteristics, and skill achievement is the only trait which can be conditioned on to segment the market. Such wage posting economy with workers' ex ante skill investment attains efficiency in the equilibria, and we show which equilibrium emerges depends on the rivalry between the measure of return to skill and the market tighness which measures the degree of market competition. The foundamental reason behind this efficiency result is that skill achievement is a quality of workers which can be legally written into the wage contracts.[1] It is a different story when other (binary) characteristics[2] which are not closely related to productivity, such as gender, race, height, origin etc. enter also into firms' preference. Under equal pay legisation, posted wages can not be conditioned explicitly on these characteristics; however, if the firms still select workers according to their discriminatory preference on these characteristics, a separating equibrium will result where separate firms post different level of wages, and workers of different groups sort themselves and apply to different wages: the market is then endogenously seggregated, and the segmentation is implicitly hierarchic along these productivity-independent characteristics. On the side of firms, they indeed have incentive to adopt such discriminatory hiring norm, since it allows them to grasp higher than the socially optimal level of operating profit. On the side of the workers, it proves that both discriminated and undiscriminated group are actually worse off: on one hand, for the former, it is because discrimination discretely reduces the labor market opportunity of these workers, who anticipate discrimination, then demand lower wages, which makes them cheaper to hire; on the other hand, for the latter, since some firms are able to hire the discriminted workers cheaply, this increases further the "market power" of firms, hence inducing them to suppress further the undiscriminated worker' expected payoff. Naturally, anticipating discrimination, all groups expect lower income from search, jeopardizing their skill investment incentives.

We then turn our attention to workers' incentives on skill investment and the resulting welfare. Because of the strategic interdependence on skill decision between discriminated and favored groups, we suggest two important means through which the discriminated group's underinvestment in skill can have impact on the payoffs of all groups: (1) the productivity destruction, and (2) the hierarchy destruction. By productivity destruction effect, we mean that firms when intending to attract discriminated group workers, now expect lower productivity hence lower operating profit, because on average the unskilled people in the discriminated group increase due to their underinvestment in response to perceived discrimination. By hierarchy destruction effect, we mean that putting workers in order of productivity-dependent quallity (here the skill level), instead of by productivity-independent quality is more pareto efficient, and could reduce the inefficiency. These two effects work jointly to decrease

[1] In competitive search, it is a common assumption firms perfectly anticipate workers' skill investment decisions, cater their demand, and post corresponding wages.

[2] In the part of discussion, we mention papers studying the more general cases where these charateristics are discretely many, or even continuous.

firms opereating profit from adopting discriminating hiring norm.

We also make comparison of our results to context where search is random and wages are determined by ex post bargaining. The strategic interdependence on payoffs has the effect of detering (bring forth) the discriminated (undiscriminated) group's skill investment incentive, in the sense that the level of workers' bargaining power required for the workers to switch from low to high skill is higher (lower). The firms' profits are piecewise monotone, because decrease of firms' bargaining power can bring gradually workers' incentive of skill investment, hence discretely improves the market skill composition and average productivity. We also find that there is an intermediate range of workers' bargaining power for values of which the firms are worse off by discriminating, due to detered skill investment of discriminated group. Then this would suggest that under random search and bargained wage, the holdup and discrimination problem can be mitigate - a result in the opposite direction of Acemoglu and Shimer (1999b). All in all, the key difference between wage posting and wage bargaining is that the ex post wage now exogenously pegs on the productivity, and firms can no longer manipulate their market power by translating their discriminatory preference into constantly lower wages.

At last, our results have also some empirical implications. In our model, there is an equilibrium where the discriminated group uses a mixed strategy in stategical skill investment: it means that, although workers are ex ante identical, discrimination has the effect of lowering their incentive of skill investment by switching *partially* to the low skilled sector to such an extent that they are indifferent from being high or low skilled. Then, neglecting these marginal workers in the low skilled sector simply leads to biased measure of wage and employment differentials in the high skilled sector, given that empirical measure of wage and employment differential on discrimination necessitates controlling for productivity characteristics, for example, skills or education. Indeed, introducing such controls, e.g. skill, seems to be based on the implicit assumption that skill is associated with one's innate ability. However, in reality, workers with similar capabilities choose their labor market skills of different levels and to different extent; analogously, workers with apparently different abilities have skill levels with similar labor market returns. Our model suggests that the rivalry between the relative return to skills, and the market tightness should be taken into account when examimining whether such an equilibrium may arise, and the expected wage instead of the acutal wages be considered when measuring discrimination.

1.1 Relation to the literature.

Job search process is an important channel through which discrimination keeps functionning in the labor market. Several papers have highlighted the impact of discrimination through job search channel to the wages gaps. To name a few, Pendakur and Woodcock (2010) shows that the existent glass ceilings for the immigrant and minority workers may be attributed by large measure to the poor access to the jobs in high-wage firms; As well, in an important article from Ritter and Taylor (2011), they show that most of the disparity in umemployment rate could not be explained by cognitive skills that emerge at an early stage, although for wage gap it could be the case. This result concerning the unemployment disparity is confirmed by the finding that this disparity is still significant even for workers of similar skill levels.

Our work is most closely related to the directed search literature[3]. In this literature, search frictions are derived endogenously through agents' sequential strategical interactions. Taking into account strategical interaction allows the search extenality to be internalized. The resulting economy remains competitive, albeit with a non-Walrasian market structure, and prices play a allocative role to achieve efficiency. To the best of our knowledege, among the discrimination literature with search friction, only two of them is built upon wage posting context. Lang, Manove, and Dickens (2005, here after LMD) show that a discriminatory hiring rule could lead to labor market segmentation, and could lead to significant wage gap within a negligible difference in productivity; however, their discriminated group turns out to have lower unemployment rate, which is in sharp contrast with evidence. Merlino (2012) aims at improving the result of LMD (2005). He considers further the pre-matching investment from the firms' side, and obtain technology dispersion and realistic unemployment gap. His results rely on the strong assumption that there is more discrimination in the high technlogy sector, and he is silent on the workers' skills levels. Our paper differs from theirs, in that our focus is to analyze how hiring discrimination could distort the structure of market segmentation, the workers' skill investment incentives, and welfare.

While the setup of wage bargaining (no information of level of wage before matching) is more prevalent, it neglects an important trade-off that the workers make to some extent in the search for jobs: the wage and the probability of obtaining it. This endogenous link between wage and employment probability is especially important, since wages convey information on whether the employers discriminate or not. Having the information on wages available before matching, workers are able to adjust accordingly their search strategy to avoid being discriminated. Workers apply to certain wage *only when* their expected income (wage times the employment probability) from this application attains certain level, and a high level of wage which attracts also the favored group discretely lowers the probability of employment for the discriminated group to such an extent that their expected payoff at these high wage firms does not meet the market payoff that they expect. This setup is supported by Lang and Lehmann (2012) and Heckman (1998), who mention that workers do not apply randomly and they actually avoid prejudiced employers to some extent, which implies between group search

[3] which is sometimes termed as wage posting game with coordination friction

extenality is taken into account by the discriminated workers. Moreover, it is well known that within group search extenality may be prevalent when wages are bargained; while in wage posting context, we are able to abstract from the entangling the effect beween discrimination and search extenality. [4] As for the empirical workers, for example, Hall and Krueger (2010) show that around one third of the current workers' wages are posted, that is not far from the number of bargained wage, which is slightly above one third.[5] They also document a negative relationship between the education level and precise information concerning the expected pay. Brenzel, Gartner and Shnabel (2013) focus on the employer's side of the study in Germany, and showed that around two thirds of the wages are posted, and the bargained wages are more likely set for those with higher education and qualification. The message is that not only is wage posting a prevalent wage determination process in the labor market, more importantly, it is also dominant in the relatively low skilled sector.[6] Within our context, employers can not post wages contingent on workers' group identity which is irrelevant to productivity, which could be understood as due to the functionning of the equal opportunity legislation.

Literature addressing discrimination problem in random search context is vaster. However, to have tractable model convenient for linking to evidence, discrimination is usually taste-based, hence to obtain realistic discrimination outcome usually may require making a few degree of compromise on assuming ex ante differences in parameters governing relevant characteristics. Rosen (1997) is an exception. In her study of hiring discrimination, there is no difference in characteristics across groups. The crucial difference is that in directed search models, workers can choose which wage (or which employer) to apply to, hence can avoid being discriminated to some extent; while in Rosen (1997), job opportunities arrive stochastically, minority workers can not choose where to apply, but can only choose reservation productivities above which they accept the job; since they know that they are more likely to meet majority workers who are always prefered, they choose to accept jobs even with low reservation wages. Although private information is the key element in Rosen (1997)'s model, search externality remains the main channel for the functionning of the discrimination mechanism, and workers do not choose where to apply. In our simple context of random search with bargained wage, the focus is on how does the ranking order of firms contribute to strategical interdependence in workers' skill investment decisions, and the comparison of agents' payoffs to the context of posted wage.

There is also the important statistical discrimination literature[7] which emphasizes the role of asymmetric information on qualities related to the productivity. One strand of this literature derives group

[4]however, by focusing on posted wage, efficiency in wage determination is guaranteed (because strategical interaction is taken into account) and we are able to focus on the effect of discrimination.

[5]Although they mention that the poste wages are usually the minimum wages.

[6]It is consistent with our knowledge that the more skilled workers, whose number is comparatively small, usually receive more attention and protections.

[7]See the survey from Hanming Fang & Andrea Moro, 2010.

inequalities endogenously even in the absence of ex ante group difference on relevant characteristics. Their mechanism is that decision makers' asymmetric beliefs on relevant characteristics of memebers for different groups could subsequently dim unfavoured agents' incentive on investment on payoff relevant technology, which in turn justifies rationalizes the firms initial beliefs. Our context is different from this literature mainly in the point that, instead of relying on the information friction which plays central role in generating the pessimistic outcome, we work through a sequential game where agents could corretly anticipate the pessimistic outcomes, hence choose to react accordingly in a rational way.

At last our work is related to the literature of ex ante investment with search frictions. Peters has a series of paper on pre-market investment. In these papers, agents' characterstics are either exogenously or endogenously continuous.[8] With continuous characteristics, the probability that any two agents have identical type is zero, so that there exists no within group competition as opposed to the case where we have the space of types is discrete, and agents of the same type compete against each other. Also, sorting is assortative. In our paper, investment choice is binary and associated with an exogenously fixed cost. Within group and between group competitions both exist. In addition, there is endogenously arised strategic payoff interdependence due to firms' ranking order, which is missing elsewhere.

The paper is organized as follows. Section 2.1 analyzes the case without discrimination. We then move to economy with discrimination: section 2.2 explains the basic setups with discrimination; in section 2.3, we start with the wage posting game given skill levels, and then examine how does discrimination alters workers' skill investment incentives. In section 3, we compare our results with those obtained in a wage bargaining context.

2 Benchmark Model

2.1 The model without discrimination

We start by a context without hiring discrimination. Consider an economy populated by two kinds of agents, the workers and the firms. The number of workers is N,[9] with the index $i \in \{1, 2, ..., N\}$, and the number of firms is M, with the index $j \in \{1, 2, ..., M\}$. Define the market tightness as $\beta \equiv \frac{N}{M}$.

[8] If these investments do not enter into the partner's utility function, then it is complete wasteful, and serve merely as signals.

[9] As noted by Lang, Manove, and Dickens (2005), the number N could be regarded as the expected number of entrants (job seekers) from the firms' perspective.

To analyze the holdup problem, we introduce a pre-investment stage in a standard wage posting game.[10] Each job seeker then makes a skill investment decision before entering into the labor market. This skill choice is assumed to be binary, such that if the worker decides to be a highly skilled worker, an investment cost E_H is paid, and otherwise E_L, with $E_H > E_L$. A highly skilled job seeker who paid E_H is capable of producing y_H; while a low skilled one could only deliver y_L. It would be useful to understand this formulation in the following way: workers who pay opportunity cost E_H enter into labor market after a longer period of training at school, hence they would expect to receive higher expected income compared to those who spend a shorter period in schooling, and enter into labor market at a much earlier stage, with a lower opportunity cost E_L.We assume that the workers' skill level is public information. And both the costs $\{E_L, E_H\}$ and productions $\{y_L, y_H\}$ are exogenous, but should satisfy some conditions which will be specified later.

Firms are ex ante identical. Having observed the distribution of skill attainment of job seekers, they cater to workers of different skill level by posting different levels of wages. If firms choose to attract a highly skilled worker, they post w_H, and the expected surplus at the ad interim stage is $y_H - w_H$, and in case a low-skilled worker is searched for, w_L is announced and the expected productivity is thus going to be $y_L - w_L$.[11] There are two remarks we would like to make. Firstly, skill level is a characteristic of workers which the wage contracts can be conditioned on; this is in sharp constrast to other qualities such as gender, race, height etc. which, under equal pay legisation, should not be conditioned on; so when firms distinguish workers according to these qualities, the wage contract becomes *"incomplete"*;[12] by this, we say that firms discriminate.

The rate of return to skills, captured by $\frac{y_H - y_L}{E_H - E_L}$, impacts workers' skill investment decisions. Depending on the magnitude of $\frac{y_H - y_L}{E_H - E_L}$, job seekers may adopt different skill formation strategies. As will be later shown, (1) if the value of $\frac{y_H - y_L}{E_H - E_L}$ is sufficiently large compared to the market tightness, all the job seekers to invest in high skill with probability 1; (2) if the value of $\frac{y_H - y_L}{E_H - E_L}$ is at an intermediate range, then it is of the job seekers' interest to adopt mixed strategy, and each firm attracts both types of job seekers while ranking the more educated in ahead of the less educated at the hiring stage; (3) when the value of $\frac{y_H - y_L}{E_H - E_L}$ strictly inferior to 1, then every job seeker will adopt pure strategy to invest in low skill. In our paper, we choose to concentrate on the case where $\frac{y_H - y_L}{E_H - E_L}$ is enough large such

[10] Our context differs from that of Merlino (2012), where he considers the pre-investment problem from the side of firms, and our emphasis is different; we also differ from Shi (2002) in that we have both homogenous firms and workers, although they differ from each other ex post.

[11] It would be useful to think the firms as adopting a skilled biased techonogy with general productivity y. A skilled worker succeeds to produce with probability p_H an output $y_H = p_H y$, and a low skilled with probability p_L gets the output $y_L = p_L y$. And $p_H > p_L$. This formulation is adopted by Shi (2002, RES).

[12] Incompleteness of contract is the source of inefficiency for the holdup probelm. See Acemoglu and Shimer (1999) for related literature.

that all the job seekers will adopt pure strategy and invest in high skills.

The timing of the wage posting game is standard. At the first stage (stage 0), workers move; their strategy consists of a decision on the level of skill from$\{L, H\}$. At the second stage (stage 1), firms move; they observe the job seekers' decision on education level, and based on this distribution of the education attainment of the job seekers, each firm chooses which type of worker to attract, and which wage to offer; firms should simultaneously announce the wage that they will commit to pay;[13] At the third stage of the game (stage 2), workers observe all the wages posted by the firms, i.e. $\mathbf{w} = \{w_1, w_2, ..., w_M\}$[14], and choose which firm to visit such that their expected income from search is maximized by taking into account the trade-off between the wage announcement he is applying for and the expected number of competing job applicants which drives down his probability of being employed; As in standard directed search literature, workers are assumed to make decisions simultaneously.[15] At the fourth stage (stage 3), each firm selects a particular worker[16] according to the following hiring rule: if no application is received ex post, then there is no production; otherwise only one will be selected randomly with equal probability among all the job applications received. The economy is hence featured with trading friction in a sequential game.

Stage 0: *Workers choose to which type of task specific trainings to devote.*

Stage 1: *The firms make expectations on the number of job seekers entering the labor market, and announce the right level of wages, taking the other firms' (best) strategies as given.*

Stage 2: *The workers observe the distribution of wages offers, and choose where to apply for the job, taking the other workers' (best) application strategies as given.*

Stage 3: *hiring selection, and payoff realized.*

As is standard in directed search literature, we will focus on subgame perfect equilibrilia. Firms adopt pure strategies by choosing wages to maximize profits, and workers adopt symmetric strategy to maximize expected payoff. By symmetric strategy, we mean that given any wage annoucenements, each worker applies to identical wage offers with the same probability. Together with the assumption that firms select one worker randomly out of the applications received, search friction naturally arises,

[13]this is equivalent to say that when firms are making decisions, they do not know the other's strategies.

[14]We can always assume, without loss of generality, that $w_1 \leq w_2 \leq ... \leq w_M$

[15]The assumption of agents' making decisions simultaneously, along with the assumption that identical agents make same moves, helps to rule out coordinations among agents, and generate trading frictions. See for example Burdett, Kenneth, Shouyong Shi and Randall Wright (2001), and the footnote 6 and 7 in

[16]Posting multiple vacancy is possible, but is not our main concern here. We refer interested readers to Shi shouyong (2002), LI and Cai (2013), Tan (2012) for further details.

because more than one applications may arrive at one firm ex post.

2.1.1 Specification of the Strategies, matching probabilities, and payoff functions.

To write the agents' payoffs, it is a routine procedure in the directed search literature to first derive the matching functions. This section provides a quick summary for the general understanding of the context.

In the labor market, it takes time for the workers to be matched with the job. According to the timing speficied above, given workers' skill levels, firms simultaneously annouce wages, knowing that the wages affect job seekers' application strategies. Having observed all the wages, job seekers choose which firm to apply to. The search friction is captured by assuming that every worker could at most send one application at each period, and each firm can only employ one worker, even if it may receive many applications. With large numbers of job seekers (supply) and firms (demand), it is natural that we put focus on symmetric mixed-strategy equilibrium, as mentioned above. The name *coordination friction* is sometimes used as a substitute for search friction, since in this game agents are not able to coordinate their decisions by applying to distinct firms.

Define a type-t job seeker i's strategy as a vector of probabilities $\Theta_t^i = \left(\theta_t^{i1}, ..., \theta_t^{iM}\right)$, where θ_t^{ij} is the probability with which the type-t worker i applies to firm j, and $t \in \{L, H\}$. It holds that $\sum_j \theta_t^{ij} = 1$ for any i and t. As in the literature, it is convenient [17] to proceed with a transformation of variable. We define q, *as expected number of applications received per firm; sometimes it is also called the expected queue length.*

Denote q^j as the queue length of firm j, and q_t^j as the queue length of the type-t workers in firms j. Then we have $q^j = q_L^j + q_H^j$, where q_L^j and q_H^j are the queue length of the workers in firm j of type-L and type-H, respectively. Since we only consider symmetric equilibria, for a given firm j, θ_t^{ij} has the same value for any job seeker i of type-t, so we denote $\theta_L^{ij} = \theta_L^j$ and $\theta_H^{ij} = \theta_H^j$ for any j. Then, q_t^j is equal to the numebr of workers of type t in firm j times their application probability[18]: $q_H^j = N_H \times \theta_H^j$, and $q_L^j = N_L \times \theta_L^j$ for any j, where N_L and N_H are the total number of workers of type L and H respectively.

Firms. A particular firm is able to match with a worker if ex post at least one worker appears, which happens with probability $1-(1-\theta_t)^{N_t}$. To see why it is the case, we notice that the probability

[17] When the number of firms and workers are large, it is no longer convenient to operate with the workers' application strategy θ_i^j, because it will tend to zero in the symmetric mixed strategy equilibrium.

[18] By definition, q is simply equal to the sum of "number of applications received" $*$ "corresponding probability".

that no job seekers send application to this firm is $(1-\theta_t)^{N_t}$, and $1-(1-\theta_t)^{N_t}$ is the probability of receiving at least one application from type-t workers. According to the aboved defined relationship $q_t = N_t\theta_t$, the probability $\left(1-(1-\theta_t)^{N_t}\right)$, goes to $(1-e^{-q_t})$ when $N_t \to \infty$. This probability is increasing in q, which means that the more the expected number of applicants, the higher the probability that the firm could fill the vacancy.

Firms' strategy is the level of wage w. Their expected payoff *from attracting a particular type of workers* is the product of the probability of meeting a worker of this type and the net profit, $(1-e^{-q_t}) \times (y_t - w_t)$, where y_t is the productivity, and $y_t = \begin{cases} y_H & \text{when } t = \text{high skill} \\ y_L & \text{when } t = \text{low skill} \end{cases}$.

As shown by Shi (2006), in case both skill types of workers appear in the market, firms will post both w_L and w_H to attract both types. Furthermore, firms will rank the high skilled groups in priority to the low skilled groups, that is to say, firms will only consider hiring the low skilled workers when they did not receive application from high skilled workers, an event which happens with probability e^{-q_H}. Then the total expected payoff *(from attracting both types of workers)* is thus $(1-e^{-q_H}) \times (y_H - w_H) + e^{-q_H}(1-e^{-q_L}) \times (y_L - w_L)$.

Job seekers. Job seekers observe all the wages $\mathbf{w}$ announced by the firms, and choose which wage to apply to. Consider a particular job seeker. Condional on visiting a particular firm, his probability of employment in that firm is $\frac{1-(1-\theta_H)^{N_H}}{N_H\theta_H}$ for a high skilled[19], and $(1-\theta_L)^{N_L} \times \frac{1-(1-\theta_L)^{N_L}}{N_L\theta_L}$ for a low skilled (see appendix for more details). And these probabilities become $\frac{1-e^{-q_H}}{q_H}$ and $e^{-q_H}\frac{1-e^{-q_L}}{q_L}$ when $N \to \infty$ and $M \to \infty$. Notice that is $\frac{1-e^{-q_t}}{q_t}$ is decreasing in q_t: the higher the expected number of applicants in this firm competiting this job with him, the lower the probability that this job seeker will be employed. Also notice that the employment probability of the low skilled workers is a product of e^{-q_H} and $\frac{1-e^{-q_L}}{q_L}$, where the former governs the between-group competition effect, and $\frac{1-e^{-q_L}}{q_L}$ governs the within group competition effect.

We remark that since q is a function of job seekers' application strategy, it depends on, or more precisely, it is induced by w. We now look more closely into their causal relationship. We should distinguish two terms: (1) each job seeker's *expected income from applying*, and (2) the *expected "market" inocme in equilibrium*, which is actually the maximal attainable level of job seeker's expected income from application.

Job seekers' expected income is a product of the wage and the probability of obtaining it, i.e., $\frac{1-e^{-q_H}}{q_H} \times w_H$ for the high skilled, and $e^{-q_H}\frac{1-e^{-q_L}}{q_L} \times w_L$ for the low skilled. *The expected "market"*

[19] $\frac{1-(1-\theta)^N}{N\theta}$ is the probability that a firm meets at least one worker divided by the expected number of applicants.

income in equilibrium, denoted U_t, could be understood as the reservation wage taken as given by all the agents in the large economy[20]. Consider a particular type-H job seeker. He is willing to send application to a firm, if and only if his expected income from sending application to the firm offering w_H, $\frac{1-e^{-q_H}}{q_H} \times w_H$, is **equal or greater than** the market wage U_H. Since U_H is defined as the maximum of expected income, we simply have the equality.[21] Thus,

$$q_H \begin{cases} > 0 & \text{if } \frac{1-e^{-q_H}}{q_H} \times w_H = U_H \\ = 0 & \text{if } \frac{1-e^{-q_H}}{q_H} \times w_H < U_H \end{cases}$$

This shows that job seekers make trade-off between the wage and the probability obtaining it. To highlight the dependence of q_t on w_t, we could rewrite the above expressions as

$$q_H \begin{cases} > 0 & \text{if } w_H > U_H \\ = 0 & \text{if } w_H \leq U_H \end{cases}$$

, because the employment probability $\frac{1-e^{-q_H}}{q_H}$ is a number between 0 and 1. It is interesting to notice that, U_H is alike the reservation wage, above which the job seekers are willing to apply to the firm.

We now show how are the firms' the workers' problem related.

The profit maximization problem for the firms. A firm maximizes expected profit, taking expected market income U_t (other firms' responses) and the functional relationship between w_t and q_t (job seekers' responses) as given.

$$\max_{w_H, w_L} \quad (1-e^{-q_H}) \times (y_H - w_H) + e^{-q_L}(1-e^{-q_H}) \times (y_L - w_L)$$

$$\text{s.to} \qquad q_H \begin{cases} > 0 & \text{if } \frac{1-e^{-q_H}}{q_H} \times w_H = U_H \\ = 0 & \text{if } \frac{1-e^{-q_H}}{q_H} \times w_H < U_H \end{cases}$$

$$q_L \begin{cases} > 0 & \text{if } e^{-q_H} \frac{1-e^{-q_L}}{q_L} \times w_L = U_L \\ = 0 & \text{if } e^{-q_H} \frac{1-e^{-q_L}}{q_L} \times w_L < U_L \end{cases}$$

A firm does not directly choose q_t, but it takes into account the functional relationship between w_t and q_t, i.e., how its wage offer may alter the job seekers' application decisions. Given U_t, firms are able

[20] When N and M tend to infinity

[21] We could could rule out the case $\frac{1-e^{-q_t}}{q_t} \times w_t > U_t$ by the following reasoning: when a firm offers w_t such that $\frac{1-e^{-q_t}}{q_t} \times w_t > U_t$, then this worker in question understands all the other workers in the economy will apply to this firm, which implies that $q_t \to \infty$; however, when $q_t \to \infty$, $\frac{1-e^{-q_t}}{q_t}$ tends to 0, so that $\frac{1-e^{-q_t}}{q_t} \times w_t > U_t$ could not hold.

to pin down q_t by determining w_t. Notice that U_t effectively comprises the other firms' responses; and since in a large economy, the number of agents tend to inifinity, a single firm's deviation does not alter the market income, it is the reason for which U_t could be taken as given. At last, it is important to remark that q_t depends on w_t continuously; as remarked by Shi (2002), in this way, a marginal change of wage w_t can only lead to a marginal modification on the expected number of applicants q_t.

2.1.2 Decentralized Market Equilibrium without discrimination.

In this section, we establish the decentralized market equilibrium, and examine its properties.

Firms' wage offers are conditioned on job seekers' skill level. So we should study the job seekers' skill decision made at first stage. Let us first introduce some notations. Denote α as the fraction of the job seekers who choose to invest in high skill, while the remaining fraction $(1-\alpha)$ represents the remaining. In a large market, α is also the probability with which a job seeker chooses to invest in high skill, by virtue of the Law of Large Number. Using α^* to denote the equilibrium fraction of job seekers with high education on the total population. There are three cases which may occur:

Possibility (1). $\alpha^* = 1$. *All job seekers use pure strategy: all will invest in high education.*
Possibility (2). $\alpha^* \in (0,1)$. *Job seekers use mixed strategy, some will invest in high education while the remaining will get low education.*
Possibility (3). $\alpha^* = 0$. *Job seekers use pure strategy: all will invest in low education.*

With Possibility (1) and Possibility (3), there exists only one skill level in the market, and since skills can be conditioned on wages, there is only one wage posted in equilibrium. However, the market with Possibility (3) features two skill levels. As in Shi (2002) and Shi (2005), firms will post two wages to attract both skill levels, while ranking the skilled in priority to the unskilled. This suggests that in equilibrium there will be segmented labor markets where different skill levels are translated into different levels of expected income.

Now when workers *strategically* choose the level of skill attainment, the magnitude of the return to skill ratio $\frac{y_H - y_L}{E_H - E_L}$ proves to be a measure of their incentive of skill investment. Specifically, we have the following lemma;

Proposition 1. (return to skills)
Given market tighness β, and return to skill $\frac{y_H - y_L}{E_H - E_L}$,

(i) when $\frac{y_H-y_L}{E_H-E_L} \geq e^{\beta}$*, the unique equilibrium is all the job seekers choose to obtain high skills, i.e.* $\alpha^* = 1$.

(ii) when $1 < \frac{y_H-y_L}{E_H-E_L} < e^{\beta}$*, the equilibrium consists of a unique value* $\alpha^* \in (0,1)$ *which satisfies* $\frac{y_H-y_L}{E_H-E_L} = e^{\alpha^*\beta}$.

(iii) when $\frac{y_H-y_L}{E_H-E_L} \leq 1$*, the unique equilibrium is* $\alpha^* = 0$.

Proof. In the appendix.

As the intuition would lead us to, when the value of return to skill $\frac{y_H-y_L}{E_H-E_L}$ is sufficiently large compared to e^{β}, which measures the tightness or intensity of compeition of the market, job seekers find it a dominant strategy to invest in high skills; There is no incentive for them to deviate, and the output is highest among all the equilibria. When the value of $\frac{y_H-y_L}{E_H-E_L}$ is moderate, there exists an equilibrium where job seekers are indifferent from being highly skilled or low skills; and all firms find it optimal to attract both the skill types; the output is lower compared to the previous equilibrium. At last, when the value of return to skill is sufficiently low, it does not provide them incentive to sunk this fixed cost against the risky job search game they are going to undergo; the equilibrium level of output turns out to be the lowest.

Now fix $\frac{y_H-y_L}{E_H-E_L}$, and define $\hat{\beta}$ such that $\frac{y_H-y_L}{E_H-E_L} = e^{\hat{\beta}}$. The incentive of skill investment decreases with the level of β in the following sense: when $0 < \beta < \hat{\beta}$, so that $\frac{y_H-y_L}{E_H-E_L} > e^{\beta}$, then workers invest with pure strategy in high skills; when $\beta > \hat{\beta}$, so that $\frac{y_H-y_L}{E_H-E_L} < e^{\beta}$, then there exists an equilibrium where workers invest with mixed strategy in high or low skills.

In the rest of the paper, we will mainly focus on the case $\frac{y_H-y_L}{E_H-E_L} \geq e^{\beta}$. So that whenever workers are discouraged to underinvest, it is due to the effect of discrimination. Now given that all workers choose to be high skilled, the firms' and workers' expected payoffs are summarized in the following proposition. The proof of it is a standard in solving directed search models.

Corollary. *When* $\frac{y_H-y_L}{E_H-E_L} \geq e^{\beta}$*, in the equilibrium, we have for the firms* $\pi_{firms} = \left(1 - e^{-q_H^*} - q_H^* e^{-q_H^*}\right) \times y_H$*, and for the workers* $w_H = \frac{q_H^* e^{-q_H^*}}{1-e^{-q_H^*}} \times y_H$.

Proof. When all the workers obtain high education, it suffices to solve the following program[22]:

$$\begin{array}{ll} \max_{w_H} & (1-e^{-q_H})\,(y_H - w_H) \\ \text{s.to} & \frac{1-e^{-q_H}}{q_H} \times w_H = U_H \end{array}$$

[22]It is interesting to notice that the competition featured in the economy is more like a monopsonistic competition instead of perfect competition. An intuition may be that because of search friction, the jobs are not perfectly substituable, although identical.

We substitute out wages w_H with the help of constraint, and maximize with respect to expected number of applicants q_H:

$$\left(1 - e^{-q_H}\right) y_H - q_H U_H$$

This expression is difference between the expected gain of the firm, and the expected cost the firm is going to pay for the job seekers who apply to this firm. Since the expected number of applicants is exactly q_H, a firm is going to pay $q_H U_H$. However, ex post there will be only workers who receive this amount U_H.[23]

Taking derivative with respect to q gives $U_H = e^{-q_H^*} y_H$. So that we obtain $\pi_{firms} = \left(1 - e^{-q_H^*} - q_H^* e^{-q_H^*}\right) \times y_H$. Substitute out U_H by $e^{-q_H^*} y_H$ in the expression for workers' expected income, we are going to obtain $w_H = \frac{q_H^* e^{-q_H^*}}{1-e^{-q_H^*}} \times y_H$, where $\frac{q_H^* e^{-q_H^*}}{1-e^{-q_H^*}}$ is actually the probability that exactly one job seeker appears conditional on the fact that at least one application is sent to the firm in question. Since in equilibrium all the workers choose to be high skilled, $q_H^* = b$. Q.E.D.

We now evaluate the efficiency of this decentralized market allocation.

2.1.3 Constrained Efficient allocations.

The objective of this section is to find the efficient allocations in the centralized market, and evaluate whether the decentralized market will achieve its efficiency. The social planner maximizes the aggregate output, subject to the same matching friction as in the decentralized equilibrium. Since *strategical interactions are internalized*, the social planner simply compares the optimal welfare given different market compositions to determine job seekers' social optimal skill levels. Specifically, in the first proposition, we find the social optimal output when all workers invest in high skills; in the second proposition, we find the social optimal output when some workers invest in high and some workers invest in low skills; in the third proposition, we compare the social optimal output of these two cases, the result of which allows us to determine the socially optimal choice of skill investment under different values of market tightness β.

***Claim 2**. **(Social Planner)** When workers all invest in high skills, the social optimum coincides the equilibrium allocation, the average aggregate income is*

$$y_H - e^{-\beta} y_H$$

[23]We refer readers to Jacquet & Tan (2012) for the possibility that for the job seekers who are not employed ex post, they are able to receive certain amount of compensations from the firm.

Proof.

The planner chooses $\{q_H\}$, to maximize the average aggregate output

$$\left(1-e^{-q_H}\right) y_H$$

subject to the feasibility constraints $q_H \leq \beta$. Denoting the Lagrangian multiplier for the constraint as λ_H, we could have the following first order condition:

With respect to q_H: $e^{-q_H^*} y_H = \lambda_H$.

As before, the job seeker's expected market income should be equivalent to the social marginal value, which is the job seekers' marginal contribution to the expected output. Q.E.D.

Claim 3*. *(Social Planner) *In case workers invest skills using mixed strategy, the social optimum coincides the equilibrium allocation. The average aggregate income is*

$$y_H - E_H + E_L - e^{-\beta} y_L$$

Proof.

Consider firstly the case where there are both high skilled and low skilled in the market. Define the priority rule $x \in [0,1]$, which is the probability of choosing group type H job applicants when both types are present. The planner chooses $\{q_H, q_L\}$, to maximize the average aggregate output

$$\left(1-e^{-q_H}\right)\left(x \times \left(1-e^{-q_L}\right)+e^{-q_L}\right) y_H + \left(1-e^{-q_L}\right)\left((1-x)\left(1-e^{-q_H}\right)+e^{-q_H}\right) y_L$$

subject to the feasibility constraints $q_H \leq \alpha\beta$, and $q_L \leq (1-\alpha)\beta$, where α is the fraction of high skilled applicants.

The expressions $(1-e^{-q_H})(x \times (1-e^{-q_L})+e^{-q_L})$ and $(1-e^{-q_L})((1-x)(1-e^{-q_H})+e^{-q_H})$ represent respectively the probability of hiring a high skilled and a low skilled applicant. We explain $(1-e^{-q_H})(x \times (1-e^{-q_L})+e^{-q_L})$, and the explanation for the low skilled is similar. $(1-e^{-q_H})$ is the probability that at least a high skilled worker will show up. Given that there is at least a high skilled worker showing up, if there is at least a lowed skilled worker showing up at the same time (which occurs with probability $(1-e^{-q_L})$), the high skilled will be hired with probability x; if no lowed skilled show up at the same time (which happens with probability e^{-q_L}), the high skilled worker will be hired for sure. So that we have $(1-e^{-q_H})(x \times (1-e^{-q_L})+e^{-q_L} \times 1)$.

Maximising with respect to x yields $(1-e^{-q_H})(1-e^{-q_L})(y_H - y_L) > 0$, so that setting $x^* = 1$ is the optimal choice. Now the planner's problem could be reduced to

$$\left(1-e^{-q_H}\right)y_H+\left(e^{-q_H}\right)\left(1-e^{-q_L}\right)y_L$$

subject to the feasibility constraints $q_H \leq \alpha\beta$, and $q_L \leq (1-\alpha)\beta$, where α is the fraction of high skilled applicants. Denoting the Lagrangian multiplier for the first constraint as λ_H, and for the second constraint as λ_L, we could have the following first order conditions:

With respect to q_H: $e^{-q_H^*}y_H - e^{-q_H^*}y_L + e^{-q_H^*-q_L}y_L = \lambda_H$

With respect to q_L: $e^{-q_H-q_L^*}y_L = \lambda_L$.

As for the interpretation, notice that since strategical interactions are taken into account, the matching externality arised from any job seekers' crowding-out of the other job seekers is internalized. As a consequence, the job seeker's expected market income should be equivalent to the social marginal value, which is the job seekers' marginal contribution to the expected output.

At the equilibrium, we have

$$\begin{aligned} e^{-q_L^*}y_L &= \lambda_L \\ e^{-q_H^*}\left(y_H-y_L\right)+\lambda_L &= \lambda_H \end{aligned}$$

Then as for the investment decision, each job seeker should be indifferent from investing in high or low skilled, so that we have

$$\lambda_H - E_H = \lambda_L - E_L$$

q_H^* is hence determined given b. The average aggregate income at the optimal level is $\left(1-e^{-q_H^*}\right)y_H+$ $e^{-q_H^*}\left(1-e^{-q_L^*}\right)y_L$, which could be succinted by virtue of the relationship $\lambda_H - E_H = \lambda_L - E_L$ to:

$$y_H - E_H + E_L - e^{-b}y_L$$

Q.E.D.

Proposition 4. ***(Social Planner)*** *When $\frac{y_H-y_L}{E_H-E_L} \geq e^{\beta}$, it is socially optimal to that workers all invest in high skills; while when $\frac{y_H-y_L}{E_H-E_L} < e^{\beta}$, it is socially optimal to that some workers invest in high, while the rest invest in low skills.*

Proof. We simply compare the average agregate output for the above two cases. Q.E.D.

Hence, when $\frac{y_H - y_L}{E_H - E_L} \geq e^{\beta}$, all the job seekers are recommended to invest in high skills due to the high rate of skill return.

2.2 The model with hiring discrimination

We now introduce discrimination. Consider an economy where workers can be partitioned into two groups, group a and group b, according to certain trait which is irrelevant to productivity directly. Gender, for example, is such one possible binary partition of labor force. The fraction of group a people is denoted as γ, and the fraction of group b people $1-\gamma$. The two group of workers are ex ante identical in all other aspects.

Discrimination modifies the matching functions of agents. We use an urn-ball example to form an analogy, in order to highlight the difference. Without discrimination, we are treating balls with different colours (colour a and colour b) identically, so that we draw balls out of the urn randomly without paying attention to their colours. With discrimination, it is as if we firstly distinguish balls with different colours, and then randomly select one among balls with the same colours.

Specifically, in order to formualize discrimination, we introduce a term x called hiring (ranking) rules specified by firms. To be precise, x could be understood as the probability that the type a workers are selected when both workers are present. The probability that a type a worker is employed by this firm is

$$F_a\left(q_a\right) = \frac{1-e^{-q_a}}{q_a} \times \left[x\left(1-e^{-q_b}\right) + e^{-q_b}\right]$$

Analogously, the probability that a type b worker is employed by this firm is

$$F_b\left(q_b\right) = \frac{1-e^{-q_b}}{q_b} \times \left[(1-x)\left(1-e^{-q_a}\right) + e^{-q_a}\right]$$

To understand these expressions, we have to notice that now when job seekers are considering their probability of being hired, they have to take into account of the impact from the competition with the other group. Take group a as an instance. Given that there are both applicants from group a and group b, the firm chooses the group a job seeker either when no group b applicants are coming, or when there is at least one applicant from b, but this job seeker from group a is preferred. The former case happens with probability e^{-q_b}, and the latter case happens with probability $(1-e^{-q_b}) \times x$. Given that the any job seeker from group a is chosen with the just obtained probability

$[x(1-e^{-q_b})+e^{-q_b}]$, a particular group a job seeker is chosen with the probability $\frac{1-e^{-q_a}}{q_a}$. These make exactly $F_a(q_a)=\frac{1-e^{-q_a}}{q_a}\times[x(1-e^{-q_b})+e^{-q_b}]$. The part for the derivation of $F_b(q_b)$ is analogous. Then the part $\frac{1-e^{-q_a}}{q_a}$ captures the within group competition, while the remaining part with x captures the between group competition.

Notice that x measures in fact the intensity of the discriminatory preference. The employment probability of group a is increasing in x, while that of group b decreases with x. When $x=1$, the employment probability for type a and type b workers become respectively,

$$F_a(q_a)=\frac{1-e^{-q_a}}{q_a}$$

$$F_b(q_a,q_b)=\frac{1-e^{-q_b}}{q_b}\times\left[e^{-q_a}\right]$$

The type b workers will only be employed when type a workers are not present. And the group a workers' employment probability is strictly higher than the case of no discrimination, and the group b workers' employment probability strictly lower. This is what we mean by hiring discrimination: different groups of people are treated differently when it is at the firms' disposal to select which worker to hire.

Another interesting example is $x=\frac{1}{2}$. The employment probability for type a and type b workers become respectively,

$$F_a(q_a,q_b)=\frac{1-e^{-q_a}}{q_a}\times\left[\frac{1}{2}\times\left(1-e^{-q_b}\right)+e^{-q_b}\right]$$

$$F_b(q_a,q_b)=\frac{1-e^{-q_b}}{q_b}\times\left[\frac{1}{2}\times\left(1-e^{-q_a}\right)+e^{-q_a}\right]$$

This means that although distinction between group a and group b is specified, each group has equal probability of being selected. In this case, it could be verified that the employment probability for both groups of workers are lower than the case without discrimination. Indeed, if there is no hiring discrimination towards any specific group, making this distinction on groups is simply redundent and inefficient.

In the rest of paper, we focus on the case $x=1$ such that group a achieves absolute priority to group b.

2.3 The case of strong discrimination: $x = 1$.

Formally, we introduce two assumptions as Merlino (2012). These assumptions help to introduce some heterogeneity which is not productivity-relevant among the labor pool.

Assumption 1: Firms are not allowed to post wages which are dependent on the group identity.

Assumption 2: Firms prefer group a job applicants in the sense that firms only hire workers from group b when group a workers are not present, i.e. $x = 1$. (By this assumption, the discrimination occurs at the hiring stage.)

Now the constraints that the job seekers' expected income from applying to any firm should be equivalent to the *expected market income* U_{aH} and U_{bH} could be written as follows repsectively.

$$
\begin{aligned}
U_{aH} &= F_{aH}\left(q_{aH}\right) \times w_H \\
&= \frac{1 - e^{-q_a}}{q_a} \times w_H
\end{aligned}
$$

$$
\begin{aligned}
U_{bH} &= F_{bH}\left(q_{aH}, q_{bH}\right) \times w_H \\
&= \frac{1 - e^{-q_{bH}}}{q_{bH}} \times \left[e^{-q_a}\right] \times w_H
\end{aligned}
$$

Without loss of generality, we start with the case where both group a and group b workers choose to be high skilled. We will proceed in section 2.3.1 to find the equilibrium of the wage posting subgame, and examine whether they have incentive to deviate from this skill investment decision.

2.3.1 Existing results revisited and reinterpreted

The above two contraints should be both binding if any firm would like to attract both of them. Lang, Manove, and Dickens (2005) (hereupon LMD) shows that there is no wage to which **both** types of job seekers apply. More precisely, no wages could maximize profit while attracting **both** types of workers (with the two constraints binding) simultaneously. The equilibrium will be separating, i.e., there will be firms posting two different levels of wages independently, with the high wage being applied by only the group a job seekers, while the low wage being applied by only the group b job seekers.

Results from LMD (2005) (separating equilibrium). *Given that all the job seekers choose to invest in high skill; In the equilibrium with hiring discrimination, some firms offer wages attracting only type a workers, and the rest offer wages attracting merely type b workers. The equilibrium conditions are*

(i) For the firms attracting type a workers:

$$Payoff_{firms,aH} = \left(1 - e^{-q_{aH}^{S}} - q_{aH}^{S} e^{-q_{aH}^{S}}\right) \times y_H$$

$$U_{aH} = e^{-q_{aH}^{S}} \times y_H$$

$$w_{aH}^{S} = \frac{q_{aH}^{S} e^{-q_{aH}^{S}}}{1 - e^{-q_{aH}^{S}}} \times y_H$$

(ii) For the firms attracting type b workers:

$$Payoff_{firms,bH} = \left(1 - e^{-q_{bH}^{S}}\right)\left(1 - e^{-q_{aH}^{S}}\right) \times y_H,$$

and

$$w_{bH}^{S} = U_{aH}$$

as well as

$$U_{bH} = \frac{1 - e^{-q_{bH}^{S}}}{q_{bH}^{S}} w_{bH}$$

where q_{bH}^{S} *and* q_{aH}^{S} *are jointly determined by*

$$Payoff_{firms,aH} = Payoff_{firms,bH}$$

(iii) $w_{aH}^{S} > w_{bH}^{S}$*, and* $q_{aH}^{S} > q_{bH}^{S}$*.*

We make some important remarks on the features of the separating equiliblrium. Firstly, the resulted equilibrlium allocations are incentive compatible. For any particular type bH job seeker, by deviating to applying for w^S_{aH}, the best they can get is $e^{-q^S_{aH}} \times w^S_{aH}$, however, this deviating payoff is strictly lower than sticking to applying to w^S_{bH} owing to the following relationship:

$$e^{-q^S_{aH}} \times w^S_{aH} = \frac{q^S_a \times e^{-q^S_a}}{1 - e^{-q^S_a}} \times U_{aH} = e^{-q^S_{bH}} U_{aH} = e^{-q^S_{bH}} w^S_{bH} < U_{bH}$$

. As for any particular type aH job seekers, by deviating to w^S_{bH}, the best they can get is $w^S_{bH} = U_{aH}$, which is as good as sticking to their destined queue length.[24]

Secondly, we do not have the reservation wage structure, where the group b job seekers apply to both the low and high wages. This is because the expected payoff from applying to the high wage is a **strictly dominated strategy** for group b: the expected income from applying to high wages is too low to match their expected market income U_{bH}. To provide more intuition on this, we also mention the following Lemma.

Results from LMD (2005). *There could exist a threshold wage $\hat{w}$ of the following properties:*

(1) when w satisfies $w \in (U_{aH}, \hat{w})$, the expected number of bH type workers is strictly greater than zero, i.e. $q_{bH} > 0$;

(2) otherwise, when $w \in \left(w^T, y_H\right)$, the expected number of bH type workers is zero $q_{bH} = 0$.

This lemma says that it's not true that any wage that attracts group a workers will also attract group b workers. There is actually a threshold $\hat{w}$, above which no group b job seekers will be attracted to the firm posting it. Notice that any group b job seeker is only considered when no group a is not applying to the firm (with probability $e^{-q_a} \times \frac{1-e^{-q_b}}{q_b}$), and high wages increase the expected number of group a job seekers to such an extent that group b job seekers' probability of finding a job decreases too fast to be compensated by the increase of wages in order to meet the market income.

Following are several noteworthy of such an equilibrium.

Results from LMD (2005)*: Under discrimination,*

(1) Both groups have lower expected income.

(2) All firms earn higher profits.

(3) The expected income of group a and group b are such that $U_{aH} > U_{bH}$.

[24] If more than one person deviates to aH firms, then the payoff from deviating is strictly lower.

Several remarks are in order. Starting with workers. Group b workers are worse off, because of firms' discriminatory hiring norm. Anticipating discrimination, group b demands lower expected payoff, which makes them cheaper to be employed. This in turn increases firms' market power in hiring group a workers. Group a workers will understand that if they demand high wages, firms will threat to switch to hiring group b workers instead. Hence group a workers demand also low wages, and are worse off too. Now more about firms. Apart from the mechanisms just described, firms are able to earn high profits because in the regime with discrimination market is segmented, which allows the firms to face less competition in each segment. As a general remark, discrimination enables firms to extract higher profit by holding up job seekers' skill investment and providing the all the job seekers lower expected income. Furthermore, since $U_{aH} > U_{bH}$, it suggests that the group b job seekers, being discriminated, are hurt to a larger extent. So that group b' incentive of skill investment is distorted further downwards. We enter more detailed discussions in the following section.

2.3.2 Analysis under our context.

In the last section, we interpreted the equilibrium of the wage posting subgame given that all workers choose to be high skilled. In this section, we study how does discrimination lead to different incentives of skill investment for these two groups respectively, and attempt to find the corresponding market equilibrium.

Before starting, we make an important observation: the skill decision for group a and group b workers is strategical, and this is a direct consequence of the coexistence of ranking through the productivity related (skill) and productivity-independent traits (discrimination). Recall that ranking by skills requires that whenever high productivity and low productivity workers appear at the same firm, the high productivity worker has the priority; While ranking by productivity-independent traits means that whenever group a and group b appear at the same firm, group a has always the priority. Athough multidimensional characteristics are involved, these two ranking schedules yield a unique market hierarchy:

$$aH \succ bH \succ aL \succ bL$$

. It reads as follows: high skilled group a workers (aH) are prefered to high skilled group b workers (bH), who are prefered to low skilled group a workers (aL), who are then prefered to low skilled group b workers (bL). Quantitatively, in terms of probability, we have

$$\frac{1-e^{-q_{aH}}}{q_{aH}} > e^{-q_{aH}} \times \frac{1-e^{-q_{bH}}}{q_{bH}} > e^{-q_{aH}-q_{bH}} \times \frac{1-e^{-q_{aL}}}{q_{aL}} > e^{-q_{aH}-q_{bH}-q_{bL}} \times \frac{1-e^{-q_{bL}}}{q_{bL}}$$

where the matching probability $\frac{1-e^{-q}}{q}$ captures the intensity of competition within the same type (within type), while the probabilitty e^{-q} captures the intensity of competition from the higher ranked type (between type). To see why this hierachy implies strategically interdependent payoffs, take group a workers as an example for explanation. Although group a is always ranked prior to group b due to discrimination, whenever they contemplate to lower skill investment, they understand that they will be ranked behind the high skilled group b workers; then the term $e^{-q_{bH}}$ which captures the competition from bH type workers simply appears in their payoffs.

Without loss of generality, we consider the case where all workers all choose to be high skilled, and see whether they have incentive deviating from this skill decision. The results from the last section allow us to make two remarks. First, since both group a and group b job seekers anticipate to be worse off in the regime with discrimination, they may both find it too costly to invest in high skills, and have incentive to deviate to low skills. Or they may both stay high skilled, as long as the reduction of expected income is not so significant. Second, since group b job seekers are worse off *to a larger extent*, their incentive of lowering skill investment is stronger compared to that of the group a job seekers. In brief, we may expect the following three situations:

(Situation 1) none of group a or group b has incentive to deviate to low skills.
(Situation 2) only group b have incentive deviating to low skilled.
(Situation 3) both group a and group b job seekers want to deviate to low skills.

Recall that we defined the ratio of number of workers to the number of firms as $\beta = \frac{N}{M}$. We have the following Lemma which suggesting that the above three situations are related to different ranges of β.

Claim 5. (Incentive of deviation) *there are two thresholds $\hat{\beta}_1$ and $\hat{\beta}_2$ defined as follows:*
(i) Holding group a's being high skilled, group b workers have incentive deviating to low skills when $\beta > \hat{\beta}_1$.
(ii) Holding group b's being high skilled, group a workers have incentive deviating to low skills when $\beta > \hat{\beta}_2$, with $\hat{\beta}_2 > \hat{\beta}_1$.

Proof in the appendix.

We have two remarks concerning this lemma. Firstly, it tells us that whenever group a workers find it better off to deviate, group b workers have incentive to deviate also; Intuitiviely, β measures the market competition, the higher the β, the expected applicants per firm are higher; so that $\hat{\beta}_1 < \hat{\beta}_2$ implies that group b workers are discouraged at milder level of market competition compared to the group a workers. This result also corresponds to our former finding that group b's expected payoff in the discrimination regime is lower than that of group a job seekers.

Secondly, this lemma also serves as a first step towards the analysis of equilibrium. For example, we know that for the parameter range $\beta \in \left(0, \hat{\beta}_1\right]$, no workers will have incentive to underinvest in skills; then the equilibrium will be essentially the same as Lang, Manove, and Dickens (2005). While for the parameter range $\beta \in \left(\hat{\beta}_1, \hat{\beta}_2\right]$, there will only be some group b workers find themselves better off by lowering skill investment, as their expected payoff from search is too low to cover the investment cost. At last, in case $\beta \in \left(\hat{\beta}_2, \hat{\beta}\right)$, the effect of discrimination towards group b is so severe that too many firms want to hire them cheaply, so that the demand of group a workers decreases sharply and they are also discouraged from investing in skills We will start with the case $\beta \in \left(\hat{\beta}_1, \hat{\beta}_2\right]$. This allows us to make a straightforward comparison between current context with endogenous and strategical skill choice and the model of Lang, Manove, and Dickens (2005) where this choice is absent.

Within the range $\beta \in \left(\hat{\beta}_1, \hat{\beta}_2\right]$, we learnt from the above lemma that only group b workers have incentive to underinvest, hence in equilibrium there is necessarily low skilled group b workers.[25] Now there are two situations to consider: either all group b workers invest in low skills (pure strategy), or some group b workers invest in low skills and some in high skills (mixed strategy). We now proceed to rule out the situation where all group b workers invest in low skills (pure strategy).

Claim 6. *It can not be an equilibrium that all group a people invest in high skills, and all group b people invest in low skills.*

Proof. In appendix.

Since we are in the region $\beta \in \left(\hat{\beta}_1, \hat{\beta}_2\right]$, then β satisfies $\beta < \hat{\beta} = \log \frac{y_H - y_L}{E_H - E_L}$, i.e., the net gain from skill investment is enough significant compared to market competition. Then holding group a's choosing high skills, this lemma says that group b workers always have incentive to invest in high skills to some extent. Then we could conclude that the only equilibrium for $\beta \in \left(\hat{\beta}_1, \hat{\beta}_2\right]$ is the case where the group b workers invest in mixed strategy in low or high skills, while group a workers invest in pure strategy in high skills. So in equilibrium there will be high skilled group a workers (group aH), high skilled group b workers (group bH), as well low skilled group b workers (group bL). We hence have the following proposition:

Proposition 7. ***(equilibria)*** *For $\beta \leq \hat{\beta}_2$, there are two possible equilibria, depending on the value*

[25] We may wonder that whether there will be also low skilled group a workers. This situation is checked by examining the equilibrium, where group a workers acutually better off.

of β.

For $\hat{\beta}_1 < \beta \leq \hat{\beta}_2$, the skill choice in equilibrium is such that all group a invest in high skills, all group b invest with mixed strategy in low or high skills. The firms' wage posting decisions are such that some firms attract only group aH workers, some firms attract both group bH and group bL workers while ranking the high skilled in priority to the low skilled.

For $\beta \leq \hat{\beta}_1$, in equilibrium all the workers invest with pure strategy in high skills. Some firms attract only group aH workers, and the rest of firms attract only group bH workers.

Proof. In appendix.

The key result of this proposition is that there is segregation: within the range $\hat{\beta}_1 < \beta \leq \hat{\beta}_2$, in equilibrium, firms either attract (high skilled) group a workers, or attract group b workers (of both skilled). We term those firms attract group aH workers as aH firms, and those firms attract group bH workers as bH firms. Then those aH firms will not be able to attract any group bL workers; in fact, the expected payoff that these aH firms can provide to group bL workers is simply lower than the expected payoff that can be offered from bH firms; this is due to the fact that group bL workers can only be hired when either an worker from group aH or a worker from group bH is not present at the firm, but the probability that a worker from group aH is not present is simply lower than the probability that a worker from group bH is present.

Now by solving the model, we are able to study how does workers' welfare differ before and after revision of skill investment choices. It is trivial that workers' welfare partially improves from revision of skill choice, but what is important is the mechanism which make it realize. We use the superscript "*" to label the equilibrium allocation of the current context after revising the skill investment decision, and the superscript "S" to label the allocation of context before the workers' revision of skill choice, which is essentially the context of LMD (2005) where possibility of endogenous skill investment is not given. Making comparison, we have:

Properties of the equilibrim.

(1) For the group a workers: $q^*_{aH} < q^S_{aH}$. We also have $w^*_{aH} > w^S_{aH}$, and $\frac{1-e^{-q^*_{aH}}}{q^*_{aH}} > \frac{1-e^{-q^S_{aH}}}{q^S_{aH}}$. The expected payoff hence increases: $U^*_{aH} > U^S_{aH}$.

(2) For the group b workers: $q^*_{bH} < q^S_{bH}$, and $q^*_{bH} + q^*_{bL} > q^S_{bH}$. We also have $w^*_{bH} > w^S_{bH}$, and $\frac{1-e^{-q^*_{bH}}}{q^*_{bH}} > \frac{1-e^{-q^S_{bH}}}{q^S_{bH}}$. The expected payoff also increases: $U^*_{bH} > U^S_{bH}$.

(3) Operating profit for the firms decreases.

Proof. In appendix.

As for the expected payoff of workers, it is not surprising that group b workers enjoy better market income, but group a workers are also better off. Hence, group b workers' underinvestment in skills not only benefits themsevles, but also benefits the group a workers: there is strategical gain for the group a workers' to remain high skilled. As for the expected queue length for both aH and bH types, both decrease: $q^*_{aH} < q^S_{aH}$ and $q^*_{bH} < q^S_{bH}$; this implies that not only the intensity of within-type competition for aH and bH types are lower, but also the between-type competition from the hierachical ranking is alleviated. We now argue that there are two effects which work jointly to bring improvement for the workers' welfare and sabotage firms' market power by employing discriminatory hiring norms. The first effect is what we term as the *productivity destruction effect*; with more group b workers lowering their skill investment, the average productivity of group b becomes lower, and the expected profit from hiring group b workers then decreases; although group b workers are still cheap to hire due to discrimination, it simply becomes more attractive for the firms to entering in the submarket for the group a workers where the average productivity is higher. The second effect is what we term as *hierarchy destruction effect*; We know that when there is skill difference, and firms are able to condition wages on skills, it is always efficient to rank the skilled group in priority to the unskilled group; now with the presence of the unskilled group b workers, firms rank these unskilled group b workers in inferiority to those skilled group b workers; this ranking within the bH firms is more efficient compared to the discriminatory hierarchic ranking by productivitiy-independant qualities **because it is based on productivity difference**, and the extent of the discrimination simply decreases, because given a skill distribution, there are less group b workers suffering from the discrimination. At last, it is simply not correct to say that the improvement of the group a workers' welfare is due to less market competition, because here the market competition is measured by β, while our above results hold taking any given value of β from the interval $\left(\hat{\beta}_1, \hat{\beta}_2\right]$.

Now we turn to the case $\beta \in \left(\hat{\beta}_2, \hat{\beta}\right]$. As mentioned before, in this region, group a will deviate holding group b's being high skilled, and group b will also deviate holding group a's being high skilled. We are able to have the following resullt:

Claim. *For $\beta \in \left(\hat{\beta}_2, \hat{\beta}\right]$, if there exists an equilibrium, at least one group uses mixed strategy in skill investment.*

Proof. In appendix.

The analysis of the possible equilibrium properties is difficult. Although we will not pursue further along, there are two pieces of important information that we could extract from this context. Firstly, this proposition maintains similar spirit as Proposition 7, where we learnt that workers' incentive in skill investment varies gradually while maintaining some continuity: it changes from pure strategy in high skill investment, to mixed strategy in either high or low skills, and lastly to pure strategy in low skill investment. Secondly, we think that both groups use mixed strategy in equilibrim may not be further welfare improving; recall that in Proposition 7, when only group b workers use mixed strategy, the overall welfare of all workers is improved because both the productivity destruction effect and hierarchy destruction effect work towards the same direction to attenuate the effect of discrimination; however, if both groups of workers adopt mixed strategy, then both effects may work towards the same direction to aggravate the effect of discrimination; on one hand, the average productivity of group a workers falls, so that it becomes again relatively attractive to consider group b workers, who can be hired cheaply; on the other hand, when both the aL type and bL type workers are present, firms discriminatory preference between group a and group b again becomes effective among these low skilled workers; hence both production destruction effect and hierarchy destruction effect are attenuated. With these arguments stated, we do have the conjecture that the equilibrium for this parameter range has the same structure as that for $\hat{\beta}_1 < \beta \leq \hat{\beta}_2$: *group a invests in high skills with pure strategy, and group b invests with mixed strategy in either high or low skills.*

2.3.3 Comparision with Wage Bargaining.

In this section, we aim to show that bargaining could mitigate the extent of discrimination, hence raise the workers' expected payoff and alleviate the inefficiency. In Acemoglu and Shimer (1999), we learn that wage bargaining pegs wages on productivity, which can either create the holdup problem, or result in inefficiently low wages levels. In this economy, the holdup problem originates from the fact that firms can provide inefficiently low wages under discriminatory hiring norms to yield higher profits; Although the posted wages can not be conditioned explicitly on workers' (productivity-independent) group identity, they perfectly reflect firms' discriminatory preference, so that group a and group b workers, by observing the wage and calculating the associated expected wage, end up choosing respectively different wage to apply to;[26] workers then are *directed* by the wages to sort themselves endogenously into segregated submarkets;[27] This segregation is inefficient, since it creates unnecessary segmentation in the markets relying on productivity-irrelevant characteristics, and causes misallocation of labor. Naturally, if we shut down the channel through which firms use wages to manipulate or influence workers' choices on applications, we may expect to alleviate the inefficiency by improving the misallocation of labor. Notice that in this section, the workers only choose the amount of skills to obtain, and not where to search.

Now consider an economy with the same discriminatory hierarchy as the previons one, but where

[26] Workers can avoid wages which leave them low expected payoff.

[27] In equillibrium, no one will deviate to the other group's wage, and it is incentive compatible.

the search is random, and the wage is determined by ex post bargaining after a job seeker meets an employer.[28] Denote the universal bargaining power for all the workers as ψ. We will focus on the case where ψ is the same for both skill levels. Within random search context, no price is observed, hence does not direct search; workers, instead of choosing where to search, wait passively until one offer to arrive. The discriminatory hiring norm is as before: on one hand, for workers of same skill level, the group a workers will be selected with probability 1 whenever both group a and group b workers are matched to the same firm; on the other hand, for workers of different skill levels, the high skilled will always be ranked in priority to the low skilled workers, i.e., we have the following firms' preference order on workers: Group $aH \succ$ Group $bH \succ$ Group $aL \succ$ Group bL. The employment probability for different types of workers is summarized as follows:

$$\begin{array}{cc}
\text{Employment probability for a high skilled group } a \text{ worker:} & \frac{1-e^{-q_{aH}}}{q_{aH}} \\
\text{Employment probability for a high skilled group } b \text{ worker:} & e^{-q_{aH}}\frac{1-e^{-q_{bH}}}{q_{bH}} \\
\text{Employment probability for a low skilled group } a \text{ worker:} & e^{-q_{aH}-q_{bH}}\frac{1-e^{-q_{aL}}}{q_{aL}} \\
\text{Employment probability for a low skilled group } b \text{ worker:} & e^{-q_{aH}-q_{bH}-q_{aL}}\frac{1-e^{-q_{bL}}}{q_{bL}}
\end{array}$$

The expected payoffs (after skill investment) for *all the possible types of workers* are calculated as follows:

$$\begin{array}{cc}
\text{Expected payoff for a high skilled group } a \text{ worker:} & \frac{1-e^{-q_{aH}}}{q_{aH}} \times \psi y_H \\
\text{Expected payoff for a high skilled group } b \text{ worker:} & e^{-q_{aH}}\frac{1-e^{-q_{bH}}}{q_{bH}} \times \psi y_H \\
\text{Expected payoff for a low skilled group } a \text{ worker:} & e^{-q_{aH}-q_{bH}}\frac{1-e^{-q_{aL}}}{q_{aL}} \times \psi y_L \\
\text{Expected payoff for a low skilled group } b \text{ worker:} & e^{-q_{aH}-q_{bH}-q_{aL}}\frac{1-e^{-q_{bL}}}{q_{bL}} \times \psi y_L
\end{array}$$

Now, we are able to specify the expected queue lengths q_{aH}, q_{bH}, q_{aL} and q_{bL} parametrically. Recall that the queue length is defined as the ratio $\frac{\text{number of workers}}{\text{number of firms that welcome these workers}}$, then we have $q_{aH} = \gamma\eta_a\beta$ for group aH workers, $q_{bH} = (1-\gamma)\,\eta_b\beta$ for group bH workers, $q_{aL} = \gamma\,(1-\eta_a)\,\beta$ for group aL workers, and $q_{bL} = (1-\gamma)\,(1-\eta_b)\,\beta$ for group bL workers, where η_a represents the fraction of high skilled group a workers, and η_b the fraction of high skilled group b workers; and we mentioned before that the fraction of group a workers is denoted by γ. The values of η_a and η_b depend on the

[28] Notice that in our context, random search is an innocuous assumption here, since the firms are ex ante actually identical, and the productivity totally depends on the skill level of workers. If the firms differ in productivity, then the workers' application decision, although can not be directed by wages directly, may be directed by other productivity-relevant traits such as firms' reputation etc..

comparison between the expected payoff from investing in high or low skills. We have

$$\eta_a \begin{cases} = 1 & \text{if } \frac{1-e^{-q_{aH}}}{q_{aH}} \times \psi y_H - E_H > e^{-q_{aH}-q_{bH}} \frac{1-e^{-q_{aL}}}{q_{aL}} \times \psi y_L - E_L \\ = 0 & \text{if } \frac{1-e^{-q_{aH}}}{q_{aH}} \times \psi y_H - E_H < e^{-q_{aH}-q_{bH}} \frac{1-e^{-q_{aL}}}{q_{aL}} \times \psi y_L - E_L \\ \in [0,1] & \text{if } \frac{1-e^{-q_{aH}}}{q_{aH}} \times \psi y_H - E_H = e^{-q_{aH}-q_{bH}} \frac{1-e^{-q_{aL}}}{q_{aL}} \times \psi y_L - E_L \end{cases}$$

and

$$\eta_b \begin{cases} = 1 & \text{if } e^{-q_{aH}} \frac{1-e^{-q_{bH}}}{q_{bH}} \times \psi y_H - E_H > e^{-q_{aH}-q_{bH}-q_{aL}} \frac{1-e^{-q_{bL}}}{q_{bL}} \times \psi y_L - E_L \\ = 0 & \text{if } e^{-q_{aH}} \frac{1-e^{-q_{bH}}}{q_{bH}} \times \psi y_H - E_H < e^{-q_{aH}-q_{bH}-q_{aL}} \frac{1-e^{-q_{bL}}}{q_{bL}} \times \psi y_L - E_L \\ \in [0,1] & \text{if } e^{-q_{aH}} \frac{1-e^{-q_{bH}}}{q_{bH}} \times \psi y_H - E_H = e^{-q_{aH}-q_{bH}-q_{aL}} \frac{1-e^{-q_{bL}}}{q_{bL}} \times \psi y_L - E_L \end{cases}$$

Remark: There are two ways of understanding the function η. Firstly, these formulations could be regarded as each individual's decision making. In this case, every individual will randomize whenever indifferent, and in case of indifference, the law of large number indicates that there will be exactly η fraction of people choosing high skilled and $(1-\eta)$ fraction of people choosing low skilled. These will endogenously determine the value of η. Secondly, we could regard this as group decision making: it is the group, instead of the individuals who randomizes, which means that the individuals in the group randomize towards the same decision. In the following will will adopt the first understanding. And, we will focus on an example where $\gamma = \frac{1}{2}$.[29] Now we start the analysis with the workers' expected payoffs. The following lemma helps explain how does workers' expected payoffs vary with respect to ψ:

Proposition 8. *There are three thresholds $\hat{\psi}_{2b}$, $\hat{\psi}_{2a}$ and $\hat{\psi}_{3b}$ such that*

(1) For values of ψ such that $\psi \in \left(0, \hat{\psi}_{2b}\right)$, all the group a and group b workers are low skilled.

(2) For values of ψ such that $\psi \in \left[\hat{\psi}_{2b}, \hat{\psi}_{2a}\right)$, there is no equilibrium except for the mixed strategy one.

(3) For values of ψ such that $\psi \in \left[\hat{\psi}_{2a}, \hat{\psi}_{3b}\right)$, all group a workers become high skilled, and all group b workers remain low skilled.

(4) For values of ψ such that $\psi \in \left[\hat{\psi}_{3b}, 1\right)$, all group a workers and group b workers become high skilled.

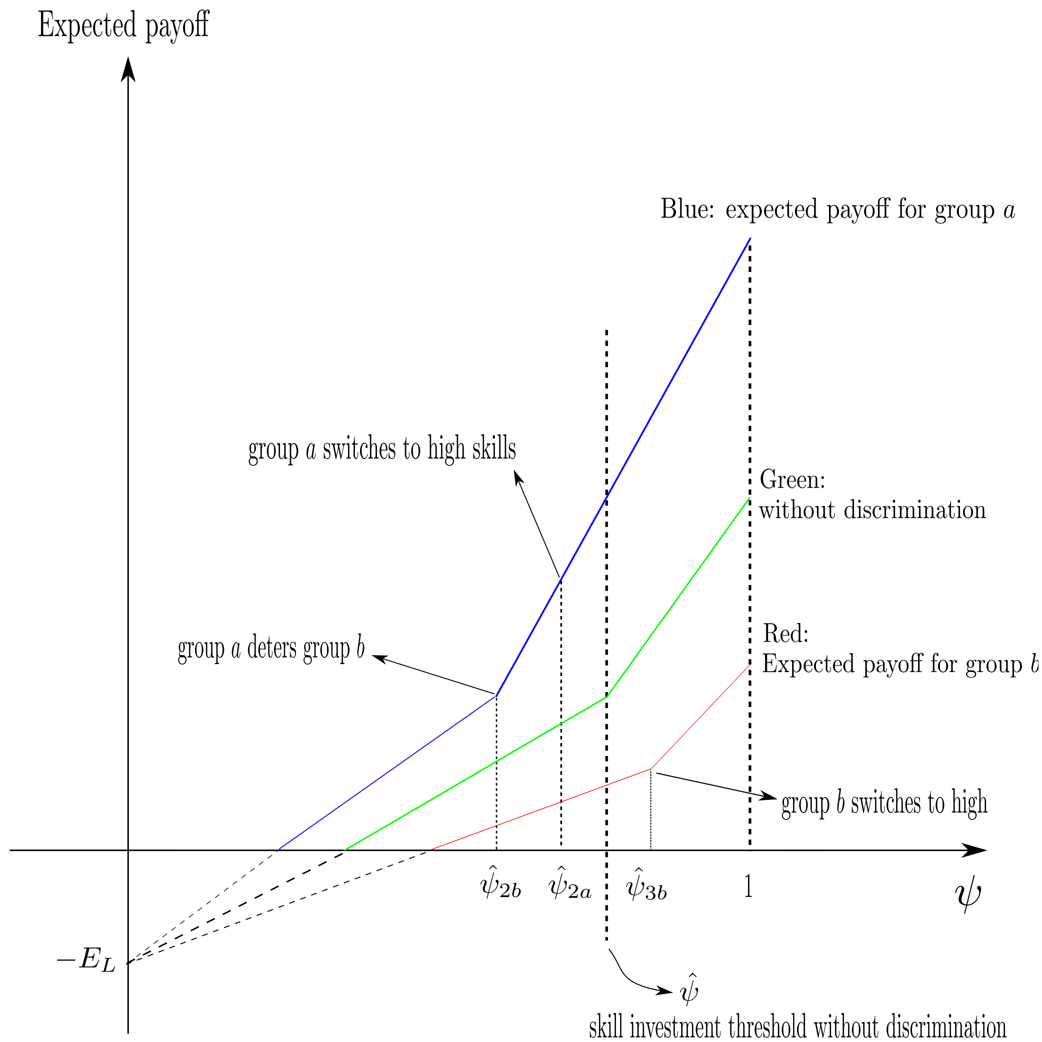
(5) $\hat{\psi}_{2a}$ and $\hat{\psi}_{3b}$ are increasing in β; and $\hat{\psi}_{2b}$ is increasing in β for small values of β when y_H/y_L is not large.

[29] To adopt the assumption that $\gamma = \frac{1}{2}$, we would like to focus on the case where both groups are of equal importance in the economy, a such example is women and men. The case where $\gamma \neq \frac{1}{2}$ introduces unequal within-group and between-group impacts on matching function; and we think that we do not have a special reason to focus on either the case of $\gamma < \frac{1}{2}$ or $\gamma > \frac{1}{2}$.

(6) Define the threshold $\hat{\psi}$ of skill investment without discrimination as $\hat{\psi} y_H \frac{1-e^{-\beta}}{\beta} - E_H = \psi y_L \frac{1-e^{-\beta}}{\beta} - E_L$; we have $\hat{\psi}_{2b} < \hat{\psi} < \hat{\psi}_{3b}$.

Proof. In the appendix.

This proposition tells us that both groups' incentive investment decision are increasing and convex functions of ψ. There are several remarks that we would like to make. Firstly, it is interesting to discover that group b workers are deterred to invest in high skills, because if they did that, group a workers will also switch to high skilled to crowd out group b, and in that case group b will be worse off. Secondly, as ψ increases to sufficiently large values, both group a workers and group b workers will switch to high skills, while recall that in the previous wage posting economy, group b workers may be discouraged from doing so; it could be also verified that the socially optimal levels of expected payoffs for both group a and group b workers are attainable for ψ sufficiently large. Thirdly, we notice that compared to the wage posting economy, the group a workers always enjoy higher than expected payoff; it is because that in the previous economy firms are able to offer group a lower expected payoff by threating to hire the group b workers cheaply; fourthly, it could be verified that the wage gap in this context is lower than that in the wage posting context for ψ small, and higher than that in the wage posting context for ψ large. The graph on workers' expected payoffs is drawn as follows.



There are two further remarks that we would like to make. Firstly, in

Now we analyze the payoff for the firms. In general, the firms' payoff is written as follows:

$$\pi = \begin{cases} \left(1 - e^{-q_{aH} - q_{bH}}\right)(1-\psi)\, y_H & \text{if all high skilled} \\ \left(1 - e^{-q_{aH} - q_{bH}}\right)(1-\psi)\, y_H + e^{-q_{aH} - q_{bH}}\left(1 - e^{-q_{aL} - q_{bL}}\right)(1-\psi)\, y_L & \text{if both high and low skilled} \\ \left(1 - e^{-q_{aL} - q_{bL}}\right)(1-\psi)\, y_L & \text{if all low skilled} \end{cases}$$

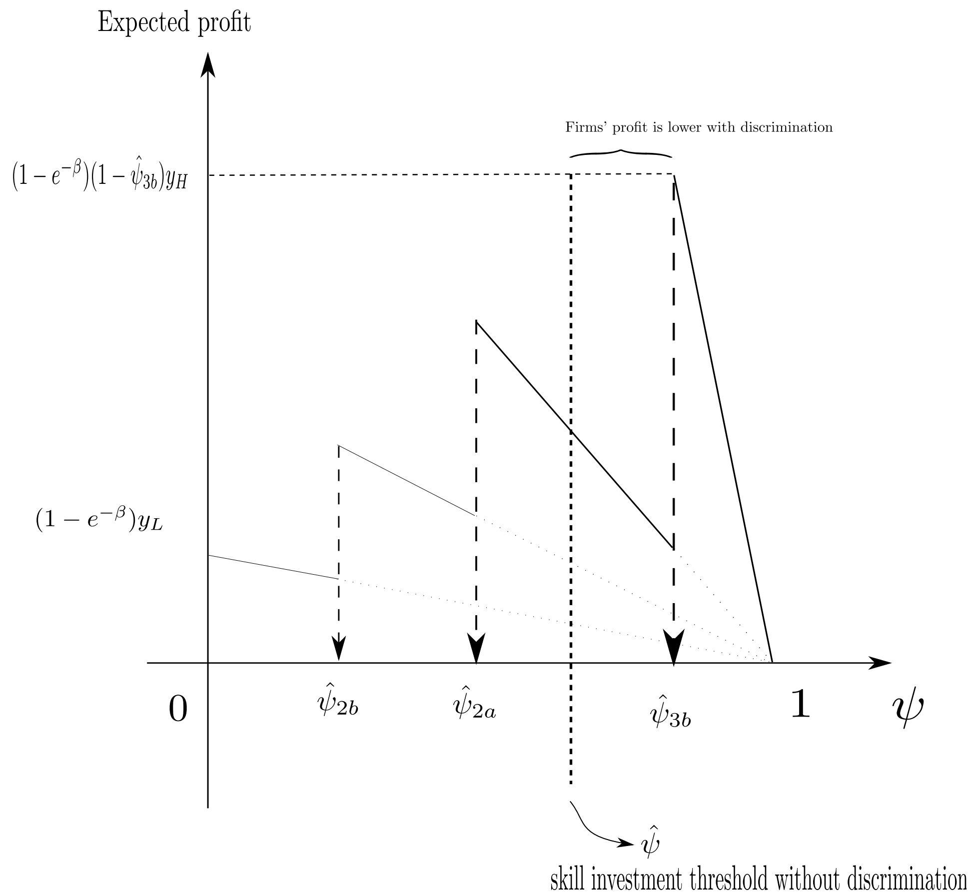
And by the above lemma, firms' payoff can be simplifed to

$$\pi = \begin{cases} \left(1-e^{-\beta}\right)(1-\psi)\, y_H & \text{for } \psi > \hat{\psi}_{3b}, \text{ both high skilled} \\ \left(1-e^{-\gamma\beta}\right)(1-\psi)\, y_H + e^{-\gamma\beta}\left(1-e^{-(1-\gamma)\beta}\right)(1-\psi)\, y_L & \text{for } \psi \in \left[\hat{\psi}_{2a}, \hat{\psi}_{3b}\right), \begin{array}{c}\text{group } a \text{ high skilled} \\ \text{group } b \text{ low skilled}\end{array} \\ \quad l_a l_b \left(1-e^{-\beta}\right)(1-\psi)\, y_L + (1-l_a)(1-l_b)(1-\psi)\, y_H & \\ +l_a(1-l_b)(1-\psi)\left[\left(1-e^{-(1-\gamma)\beta}\right) y_H + e^{-(1-\gamma)\beta}\left(1-e^{-\gamma\beta}\right) y_L\right] & \text{for } \psi \in \left[\hat{\psi}_{2b}, \hat{\psi}_{2a}\right), \text{ mixed} \\ +l_b(1-l_a)(1-\psi)\left[\left(1-e^{-\gamma\beta}\right) y_H + e^{-\gamma\beta}\left(1-e^{-(1-\gamma)\beta}\right) y_L\right] & \\ \left(1-e^{-\beta}\right)(1-\psi)\, y_L & \text{for } \psi < \hat{\psi}_{2b}, \text{ both low skilled} \end{cases}$$

where l_a and l_b are respectively the probability of investing in low skills for group a and group b workers.[30] These two values are calculated in the appendix.

Now the graph for the firms' profit is drawn straightforwardly as follows:

[30] we emphasize that they are group decisions instead of individual decisions. In case of individual decisions, these values enter into the matching function.



As ψ increases, workers' incentive in skill investment increases. The piecewise monotonicity is simply due to the fact that although ψ increases continuously, the skill composition hence the average productivity of the market improves discretely with respect to this bargaining power. We also observe that although firms can gather higher profits for the lower range of workers bargaining power $\psi < \hat{\psi}$, they encounter loss for $\psi \geq \hat{\psi}$ compared to case without discrimination. The reason is that strategical competition between different groups of workers deter the discriminated group's skill investment decision, which pulls down the market's average productivity and makes firms' expected profit dim.

It is also evident from the graph that firms, by discriminating, are better off within the range $\psi \in \left[\hat{\psi}_{2b}, \hat{\psi}\right)$, and worse off with the range $\psi \in \left[\hat{\psi}, \hat{\psi}_{3b}\right)$. Several remarks are in order. Firstly, the fact that firms incur profit loss at the bargaining power range $\psi \in \left[\hat{\psi}, \hat{\psi}_{3b}\right)$ implies that the holdup-discrimination problem may be mitigated when search is random and wages bargained. This profit loss due to discrimination happens simply because strategic interactions between two groups deter the discriminated group's skill choice to such an extent that ony sufficiently high wage (bargaining power) can induce them to participate again in high skills, and can overcome the discrete drop in employment probability due to the presence of group a workers; notice that this between-group competition effect is captured by $e^{-q_{aH}}$. Secondly, it is interesting to notice that our simple result that discrimination is costly for firms at high skilled sector (when wages are bargained) questions the plausibility of key assumption of Merlino (2012) that "there is more discrimination in the high technology sector". Although Merlino (2012) mentioned bunches of empirical evidence in support of this assumption[31], we suggest that firms are simply better off not discriminating when wages are principally bargained, since the loss in profit from discriminating in the high skilled sector may surpass the gain from discriminating in the low skilled sector. All in all, the key difference between wage posting and wage bargaining is that the ex post wage now exogenously pegs on the productivity, and firms can no longer manipulate their market power by translating their discriminatory preference into constantly lower wages.

2.4 Discussion.

The harshest critique on wage posting models with discrimination is that unrealistic (wage, employment probability) pair usually results: in the model, the favored group has a lower employment rate compared to the discriminated group; however, in reality, it is generically true that workers who previously have high wages have on average lower unemploymnet duration; this implies these workers find a job more quickly and their employment probability is higher. Unfortunately, in our above version with endogenous skill investment, we are not able to bring complete improvement to this critique compared to the orginal model. The particular result in such wage posting economy is due to the fact that the segmentation of the labor market is binary (either group a or group b). With discretely more types, there will be equilibria depending on the fraction of each groups and separating equilibrium is just one possibility, there is just more flexibility for obtaining more realistic (wage, employment probability) pairs with richer implications; however, obtaining analytical results are very difficult. We refer interested readers to Lang and Manove (2003) and Peters (2010) for more discussions in this direction.

Another important question is how do labor market institutions have impact on the distorted market allocations due to discrimination. Consider first unemployment benefit. It is well known from Acemoglu and Shimer (1999a) that unemployment benefit encourages workers to take more risks and ask for higher wage jobs. In our wage posting context with discrimination[32], higher wage is a stronger sign of discrimination; so that when such higher wage encourages group a's search, it further discourages group b's incentive of applying to it. Obliged to offer high wages, firms have more incentive

[31] See Merlino (2012) page 4 for more relevant referece.

[32] As for the case of bargained wage, introducing unemployment benefit is simply equivalent to high bargaining power of workers.

to reduce the cost, and switch to hire group b workers cheaply (i.e. to discriminate); this in turn puts more downward pressure on the group a's welfare. Unemployment benefit then has the consequence of aggravating the impact of discrimination. It is a different situation if we consider minimum wage policy. It simply has the effect of making hiring group b workers more expensive, which hence reduces the firms' incentive of discriminating.

The study of skill investment subsidy turns out to be more complicated. Here, we would like to describe some of our results up to this moment. In brief, there are two interesting cases to consider: either both groups receive same amount of subsidy (symmetric), or the discriminated group is subsidized to a larger extent (asymmetric). In the symmetric case with wage posting, introducing subsidy increases high skilled population for both group a and group b. This will have the effect of intensifying both within-type competition and between-type competition; and firms are able to offer lower wages universally because it is easier to get matched. In this sense, discrimination is aggravated. Now we turn to the asymmetric case. Although it does raise group b workers' incentive in skill investment, we find that if the magnitude of this subsidy is enough high, group a workers' investment incentives may be discouraged, so that an equilibrium where all group b workers are high skilled, and group a workers mix between high and low skills may emerge - We may observe overinvestment of skill from the discriminated group. Lang and Manove (2011) indeed finds empirical evidence that for similar earning levels, African Americans may get more education than the whites do. They propose an explanation under context of statistical discrimination: education serves as a signaling device for facilitating the employers' evaluation.[33] Under our model with asymmetic skill subsidy seems, this overinvestment result simply comes from a different mechanism based on intertwined rankings and group competitions[34]. As for a short summary for our preliminary discussion of policy implications, group specific political instrument seems to be more efficient than universal ones. And we aim to bring more rigourous treatment in the next step.

At last, we would aslo like to say something concerning what we abuse as the direction of improvement. In an independent work, Galenianos, Kircher, and Gábor Virág (2011) shows that in a wage posting framework with heterogenous productivity and firms' market power and the aggreagate output being the measure of efficiency, introducing unemployment rate can restore efficiency, but may increase wage inequality, and increase the expected unemployment rate of the economy; so that it is not necessarily a bad thing to observe an increase in wage inequality and expected unemployment rate, when we know that it is for the sake of increasing the aggregate output. In our framework, we could distinguish two regimes: with or without the skill investment decisions; The case without

[33] Search friction is missing in their context. It is in general very difficult to mechanically embed search friction into statistical discrimination models to obtain implications on matching rate.

[34] This competition comes in fact from the fact that firms are capacity constrained, and each firm can only serve one worker.

skill investment decisions can be interpreted as the first generation of immigrants or people who are "shortsighted", and will invest definitely in high skills. Consider two overlapping generations (or two states or a country may also work), if the latter generation observes that the people in the previous one suffer due to discrimination, then they simply invest less in skills. The resulting underinvestment in skills actually increases the welfare as we were able to show. Hence, it may not be necessarily a bad thing when we observe lower skill investment in an economy where discrimination is potentially important.

3 Conclusion.

In this paper, we study a holdup problem where firms can use discriminatory hiring norms to extract higher than socially optimal profits. There are three main contributions that we have done, and which are not yet in literature. Firstly, we consider an economy where workers are ranked both by productivy-dependent and productivity-independent characteristics; hiring discrimination creates extra hierarchy among workers, leading to further market segmentation; we analyze the link between the structure of market segmentation and the associated expected wage.

Secondly, we put focus on the workers' welfare under these market structures. We find that discrimination creates strategic payoff interdependence for workers. Some workers' underinvestment in skills tend to improve the overall welfare, and there are two important channels through which this realizes: (1) the productivity destruction channel, and (2) the hierarchy destruction channel. Both effects work towards the same direction to make firms' discrimination less profitable, which in turn alleviates inefficiency by partially correcting the market misallocation of labors; However, the effect on the expected aggregate output is ambiguous, because the improvement in labor allocation comes at a cost on deteriorated skill compostion which implies lower average productivity.

At last, we consider the case where wages are bargained and make a comparison. We find that the discriminated group's skill investment decision is detered while the favored group's is advanced owing to these groups' strategic payoff interdependence. Firms' profits are piecewise monotone because the skill composition hence the average productivity of the market improves discretely with respect to the bargaining power, and profit loss may be incurred with discrimination within an intermediate range of bargaining power.

References

[1] Acemoglu, Daron & Shimer, Robert (1999a): "Efficient Unemployment Insurance," Journal of Political Economy, University of Chicago Press, vol. 107(5), pages 893-928, October.

[2] Acemoglu, Daron & Shimer, Robert (1999b): "Holdups and Efficiency with Search Frictions," International Economic Review, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association, vol. 40(4), pages 827-49, November.

[3] Albrecht, James & Tan, Serene & Gautier, Pieter & Vroman, Susan (2004): "Matching with multiple applications revisited," Economics Letters, Elsevier, vol. 84(3), pages 311-314, September.Shouyong Shi, "Wage Differentials, Discrimination and Efficiency," European Economic Review 50 (2006), 849-875.

[4] Brenzel, Hanna & Gartner, Hermann & Schnabel, Claus (2014): "Wage bargaining or wage posting? Evidence from the employers' side," In: Labour Economics, Vol. 29, August, S. 41-48.

[5] Burdett, Kenneth & Shi, Shouyong & Wright, Randall (2001): "Pricing and Matching with Frictions," Journal of Political Economy, University of Chicago Press, vol. 109(5), pages 1060-1085, October.

[6] Cao, Melanie & Shi, Shouyong (2000): "Coordination, matching, and wages," Canadian Journal of Economics, Canadian Economics Association, vol. 33(4), pages 1009-1033, November.

[7] Clark, AE & Oswald, AJ (1996): "Satisfaction and comparison income," Journal of public economics 61 (3), 359-381.

[8] Clark, AE (1997): "Job satisfaction and gender: why are women so happy at work?" Labour economics 4 (4), 341-372

[9] Delacroix, Alain & Shi, Shouyong (2013): "Pricing and signaling with frictions," Journal of Economic Theory, Elsevier, vol. 148(4), pages 1301-1332.

[10] Neal, Derek A & Johnson, William R. (1996): "The Role of Premarket Factors in Black-White Wage Differences," Journal of Political Economy, University of Chicago Press, vol. 104(5), pages 869-95, October.

[11] Fang, Hanming & Moro, Andrea (2010): "Theories of Statistical Discrimination and Affirmative Action: A Survey," NBER Working Papers 15860, National Bureau of Economic Research, Inc.

[12] Galenianos, Manolis & Kircher, Philipp & Virág, Gábor (2011): "Market Power And Efficiency In A Search Model," International Economic Review, Department of Economics, University of

Pennsylvania and Osaka University Institute of Social and Economic Research Association, vol. 52(1), pages 85-103, 02.

[13] Hall, Robert E. & Krueger, Alan B. (2012): "Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search," American Economic Journal: Macroeconomics, American Economic Association, vol. 4(4), pages 56-67, October.

[14] Heckman, James J. (1998): ”Detecting Discrimination,” Journal of Economic Perspectives, 12(2): 101-16.

[15] Jacquet, Nicolas L. & Tan, Serene (2012): "Wage-vacancy contracts and coordination frictions," Journal of Economic Theory, Elsevier, vol. 147(3), pages 1064-1104.

[16] O'Neil, June (1990): "The Role of Human Capital in Earnings Differences between Black and White Men," Journal of Economic Perspectives, American Economic Association, vol. 4(4), pages 25-45, Fall.

[17] Philipp, Kircher, & Manolis, Galenianos (2009): ”Directed Search with Multiple Job Applications,” Journal of Economic Theory, 114(2), pp. 445-471.

[18] Lang, Kevin & Michael, Manove (2003): "Wage announcements with a continuum of worker type," Annales d'Economie et de Statistique, ENSAE, issue 71-72, pages 223-244.

[19] Lang, Kevin & William, T. Dickens (2005): ”Racial Discrimination in Labor Markets with Posted Wage Offers," American Economic Review, 95 (4), 1327-1340.

[20] Lang, Kevin & Michael, Manove (2011): ”Education and Labor Market Discrimination," American Economic Review, 4 (101), 1467-96.

[21] Lang, Kevin & Jee-Yeon K. Lehmann (2012): "Racial Discrimination in the Labor Market: Theory and Empirics," Journal of Economic Literature, American Economic Association, vol. 50(4), pages 959-1006, December.

[22] LI, Fei & Can, Tian (2013): ”Directed Search and Job Rotation,” Journal of Economic Theory, Volume 148 (3), 1268–1281.

[23] Merlino, Luca Paolo (2012): ”Discrimination, Technology and Unemployment,” Labour Economics, 19(4), pp. 557-567

[24] Pendakur, Krishna & Woodcock, Simon (2010): ”Glass Ceilings or Glass Doors? Wage Disparity Within and Between Firms," Journal of Business & Economic Statistics, American Statistical Association, vol. 28(1), pages 181-189.

[25] Peters, Michael (2010): "Noncontractible Heterogeneity in Directed Search," Econometrica, Econometric Society, vol. 78(4), pages 1173-1200, 07.

[26] Ritter, Joseph A. & Lowell, J. Taylor (2011): ”Racial Disparity in Unemployment," Review of Economics and Statistics, 93 (1), 30-42

[27] Rosen, Asa (1997): "An equilibrium search-matching model of discrimination," European Economic Review, Elsevier, vol. 41(8), pages 1589-1613, August.

[28] Shimer, Robert (2005): "The Assignment of Workers to Jobs in an Economy with Coordination Frictions," Journal of Political Economy, University of Chicago Press, vol. 113(5), pages 996-1025, October.

[29] Shi, Shouyong (2001): "Frictional Assignment. I. Efficiency," Journal of Economic Theory, Elsevier, vol. 98(2), pages 232-260, June.

[30] Shi, Shouyong (2002): "Product Market and the Size-Wage Differential," International Economic Review, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association, vol. 43(1), pages 21-54, February.

[31] Shi, Shouyong (2002): "A Directed Search Model of Inequality with Heterogeneous Skills and Skill-Biased Technology," Review of Economic Studies 69, 467-491.

[32] Tan, Serene (2012): ”Directed Search and Firm Size,” International Economic Review, Vol. 53(1), 95-113.

[33] Virág, Gábor (2011): "High profit equilibria in directed search models," Games and Economic Behavior, Elsevier, vol. 71(1), pages 224-234, January.

Appendix

.1 Derivation of matching probabilities.

We now derive a job seeker's matching probability and expected payoff.

Job seekers. Having observed all the wage $\mathbf{w} = \{w_1, w_2, ..., w_M\}$ announced by the firms, job seekers choose which firm (or wage) to visit (or to apply for). Consider a particular job seeker i's problem, where $i \in \{1, 2, ..., N\}$. This job seeker thinks in the following way: Suppose I visit firm j, then conditional on the fact that my application is sent to j, what is the probabitility that I could be employed? It depends upon the number of *the other job seekers* who also send their job application to the same firm competiting with me on this job in firm j. This number (of *the other job seekers*) is a random variable which has a realisation from the set $\{0, 1, ..., N-1\}$ and has a Binomial distribution. To see why it is the case, we use k to represent the realized number of competitors. If $k = 0$, which happens with probability $(1-\theta_j)^{N-1}$, then the job seeker i will be chosen by the firm with probability 1, because this job seeker is the only candidate. If $k = 1$, which happens with probability $(N-1) \times \theta_j^1 (1-\theta_j)^{(N-1)-1}$, this job seeker i will be chosen by the firm with probability $\frac{1}{2}$, because now the firm receives two applications, hence has two candidates, among whom i is one. Generalising, if $k = \hat{k}$, which happens with probability $C_{N-1}^{\hat{k}} \times \theta_j^{\hat{k}} (1-\theta_j)^{N-1-\hat{k}}$, then this job seeker i will be chosen with probability $\frac{1}{\hat{k}+1}$, because the firm j has $\hat{k}+1$ candidates at disposal.

The employment probability for the workers are $\sum_{k=0}^{N-1} C_{N-1}^{k} \theta_j^{k} (1-\theta_j)^{N-1-k} \frac{1}{k+1}$. This expression could be simplified to $\frac{1-(1-\theta_j)^N}{N\theta_j}$ [35]. Hence the job seeker's expected pay off is $\frac{1-(1-\theta_j)^N}{N\theta_j} \times w_j$, where as we have mentionned above when we are deriving the firms' expected payoff, θ_j actually depends upon $\mathbf{w}$.

.2 Proofs of propositions

Proposition 1. (return to skills)

Given market tighness β, and return to skill $\frac{y_H - y_L}{E_H - E_L}$,

(i) when $\frac{y_H - y_L}{E_H - E_L} \geq e^{\beta}$, the unique equilibrium is all the job seekers choose to obtain high skills, i.e. $\alpha^* = 1$.

[35] One way of deriving it could be seen in Melanie Cao & Shouyong Shi, 2000. "Coordination, matching, and wages". It could also be checked by change of variable, which is also represented in the Appendix.

(ii) when $1 < \frac{y_H - y_L}{E_H - E_L} < e^{\beta}$*, the equilibrium consists of a unique value* $\alpha^* \in (0,1)$ *which satisfies* $\frac{y_H - y_L}{E_H - E_L} = e^{\alpha^* \beta}$.

(iii) when $\frac{y_H - y_L}{E_H - E_L} \leq 1$*, the unique equilibrium is* $\alpha^* = 0$.

Proof. We will prove only case (i) while the proof of case (ii) and (iii) are highly similar.

Notice that $\frac{y_H - y_L}{E_H - E_L} \geq e^{\beta}$ is equivalent to $e^{-\beta} y_H - E_H \geq e^{-\beta} y_L - E_L$.

Now we would like to prove that the optimal choice is $\alpha^* = 1$, when $e^{-\beta} y_H - E_H \geq e^{-\beta} y_L - E_L$. We prove firstly that the deviation to low skills is not optimal. By this, we prove that a proportion ϵ of workers' deviating to the low skilled type job seekers is suboptimal. And it suffices to show that after deviation, the deviator can not get higher expected payoff. Before deviation, the expected payoff for these deviators is $e^{-q_H^*} y_H - E_H$, where $q_H^* = \beta$. After deviation, the expected payoff becomes $e^{-q_H^D - q_L^D} y_L - E_L$, where $q_H^D + q_L^D = \beta$. However, under the condition $e^{-\beta} y_H - E_H \geq e^{-\beta} y_L - E_L$, the expected payoff after deviation is weakly lower.

For the uniqueness. We should furthermore show that for the case of $\alpha = 0$ and $\alpha \in (0,1)$, there will be profitable deviation.

When $\alpha = 0$, the expected income from search is $e^{-\beta} y_L - E_L$. If there is a fraction ϵ deviating to high skilled, then the expected income for the deviator becomes $e^{-\epsilon\beta}(y_H - y_L) + e^{-\beta} y_L - E_H$. Then this expected payoff after deviation is greater than the the expected payoff before deviation because. $e^{-\epsilon\beta}(y_H - y_L) > e^{-\beta}(y_H - y_L) \geq E_H - E_L$. So the deviation is profitable for the deviators. As for the rest of the population $(1-\epsilon)$, their expected payoff is not affected. Hence deviating weakly increases the payoff of all the job seekers.

When $\alpha \in (0,1)$, the expected income from search is $e^{-\beta} y_L - E_L$ for the low skilled, and $e^{-\hat{\alpha}\beta}(y_H - y_L) + e^{-\beta} y_L - E_H$ for the type H job seekers, where $\hat{\alpha}$ should be pinned down by workers' indifference condition $e^{-\hat{\alpha}\beta}(y_H - y_L) = E_H - E_L$. However, this condition is incompatible for all $\alpha < 1$ with our condition $e^{-\beta} y_H - E_H \geq e^{-\beta} y_L - E_L$. So that it is impossible that job seekers are indifferent from being high or low skilled.

All in all, we have proved that when the configuration of parameters is such that $e^{-\beta} y_H - E_H > e^{-\beta} y_L - E_L$, the only equilibrium is all the job seekers choose to obtain high education, i.e. $\alpha^* = 1$.

Claim 5. (Incentive of deviation) *there are two thresholds* $\hat{\beta}_1$ *and* $\hat{\beta}_2$ *defined as follows:*

(i) Holding group a's being high skilled, group b workers have incentive deviating to low skills when $\beta > \hat{\beta}_1$.

(ii) Holding group b's being high skilled, group a workers have incentive deviating to low skills when $\beta > \hat{\beta}_2$, with $\hat{\beta}_2 > \hat{\beta}_1$.

Proof. The outline of the proof is as follows. Firstly, we notice that since group a workers are always ranked in priority to group b workers for the same skill level, then whenever group a workers have incentive to deviate, group b workers have stronger incentive to do so. Hence we start with case (iii) to find the threshold of $\hat{\beta}_2$, beyond which group a workers deviate. I then find the threshold $\hat{\beta}_1$, beyond which group b workers have incentive to deviate, and I check that $\hat{\beta}_1 < \hat{\beta}_2$ so that indeed the group b workers have stronger incentive to deviate.

Proof of (iii). Holding group b workers remaining high skilled, suppose *some group a job seeker* deviates to low skills. Then in equilibrium, there will be 3 types of workers: high skilled group a workers (aH), high skilled group b workers (bH), as well as the low skilled group b workers (aL). Then the resulting wage posting subgame has the following structure: some firms will only post wages for aH job seekers, while the remaining firms will post wages for both bH and aL job seekers. This is so, because the expected wage that these aH firms can offer to the aL job seekers is strictly lower than the expected wage that can be offered by those firms attracting bH type job seekers. Hence choosing bH firms is a dominant choice for those deviating aL workers.

Solving this wage posting subgame. At the equilibrium, the job seeker's expected market income after deviation is $e^{-q_{bH}^D - q_{aL}^D} y_L - E_L$. While before deviation, his expected market income is $e^{-q_{aH}^S} \times y_H - E_H$. Then, there is deviation if and only if

$$e^{-q_{bH}^D - q_{aL}^D} y_L - E_L > e^{-q_{aH}^S} \times y_H - E_H$$

Since we consider unilateral deviation, let q_{aL}^D tend to zero. (Also when $q_{bL}^D \to 0$, we have q_{bH}^D tends to q_{bH}^S.) We define the level of β such that the following relation holds as $\hat{\beta}_2$:

$$e^{-q_{bH}^S(\hat{\beta}_2)} y_L - E_L = e^{-q_{aH}^S(\hat{\beta}_2)} \times y_H - E_H$$

Recall that we have $\frac{1-\gamma}{1-\sigma^S}\beta = q_{bH}^S < \beta < q_{aH}^S = \frac{\gamma}{\sigma^S}\beta$. So that $e^{-q_{bH}^S} y_L - E_L$ is always above $e^{-\beta} y_L - E_L$, while $e^{-q_{aH}^S} \times y_H - E_H$ is always below $e^{-\beta} \times y_H - E_H$. Then there always exists a $\hat{\beta}_2 < \hat{\beta}$, where $\hat{\beta}$ satisfies $e^{-\hat{\beta}} y_L - E_L = e^{-\hat{\beta}} \times y_H - E_H$.[36]

As a summary, this proof tells us that holding group b's being high skilled, it is a dominant choice for group a to remain high skilled if $\beta \leq \hat{\beta}_2$, and it is a dominant choice

[36] Notice that if we do not consider unilateral deviation, but consider group deviation, the threshold is different. (in fact larger for the case of group deviation.)

for group a to go to low skilled if $\beta > \hat{\beta}_2$.

We turn to the case (ii) now.

Proof of (ii). Holding group a workers remaining high skilled, suppose *some group b job seeker* deviates to low skills. Then in equilibrium, there will be 3 types of workers: high skilled group a workers aH, high skilled group b workers bH, as well as low skilled group b worker bL. We can prove that the resulting wage posting subgame has the following structure: some firms will attract **only** aH job seekers, while the remaining firms will attract **both** bH and bL job seekers.

We solve this wage posting subgame. At the equilibrium, the deviating job seeker's expected market income is $e^{-q^D_{bH}-q^D_{bL}}y_L - E_L$. But if he does not deviate, his expected market income is $\frac{1-e^{-q^S_{bH}}}{q^S_{bH}} \times e^{-q^S_{aH}} \times y_H - E_H$. Then, there is deviation if and only if

$$e^{-q^D_{bH}-q^D_{bL}}y_L - E_L > \frac{1-e^{-q^S_{bH}}}{q^S_{bH}} \times e^{-q^S_{aH}} \times y_H - E_H$$

Since we consider unilateral deviation, $q^D_{bL} \to 0$. Then when $q^D_{bL} \to 0$, we have $q^D_{bH} \nearrow q^S_{bH}$. We define the β such that the following relation holds as $\hat{\beta}_1$.

$$e^{-q^S_{bH}(\hat{\beta}_1)}y_L - E_L = \frac{1-e^{-q^S_{bH}(\hat{\beta}_1)}}{q^S_{bH}\left(\hat{\beta}_1\right)} \times e^{-q^S_{aH}(\hat{\beta}_1)} \times y_H - E_H$$

, where $q^S_{bH} = \frac{1-\gamma}{1-\sigma^S}\beta$, and $q^S_{aH} = \frac{\gamma}{\sigma^S}\beta$. Then the same reasoning as in the proof of (iii) leads us to conclude that there always exists a $\hat{\beta}_1 < \hat{\beta}$.

As a summary, this proof tells us that holding group a's being high skilled, it is a dominant choice for group b to remain high skilled if $\beta \leq \hat{\beta}_1$, and it is a dominant choice for group b to go to low skilled if $\beta > \hat{\beta}_1$.

We now turn to the case (i).

We notice that since $e^{-q^S_{aH}} \times y_H - E_H > \frac{1-e^{-q^S_{bH}}}{q^S_{bH}} \times e^{-q^S_{aH}} \times y_H - E_H$, whenever group a job seekers have incentive to deviate, group b job seekers must have already incentive to deviate, which implies that $\hat{\beta}_2 > \hat{\beta}_1$. Or in other words, when group b workers do not have incentive to deviate, group a workers do not. This suggests that for values of $\beta < \hat{\beta}_1$, all the workers do not have incentive to more to low skills. This coincides with our former result that group b job seekers have higher incentive of lowering skills because their expected income in the discrimination regime is much lower than that of group a job seekers. Q.E.D.

Claim 6. *It can not be an equilibrium that all group a people invest in high skills, and all group b people invest in low skills.*

Proof.

We have seen above that when $\beta < \hat{\beta}_2$, all the group a job seekers will choose to invest with pure strategy to high skills. Now it suffices to check when we are in the equilibrium where group b workers all invest with pure strategy in low skills, whether group b job seekers are going to deviate to high skills.

When all group b workers invest in low skills, the resulting equilibrium is such that all firms attract both high skill group a workers, and low skill group b workers; group a workers are ranked in priority to group b workers.

Then expected income for the group b people is $e^{-\beta} y_L - E_L$. Now suppose that some group b worker choose to deviate to high skilled, then we have 3 types of people: the high skilled group b worker (type bH), the low skilled group b workers (type bL), and the high skilled group a workers (type aH).

There are two possibilities for the equilibrium after deviation:

(E1) The resulting equilibrium is such that some firms attract group aH workers, and the remaining firms attract bH and bL workers; bH workers are ranked in priority to bL workers.

(E2) The resulting equilibrium is such that some firms attract group aH workers, and some firms attract bH workers, and the remaining attract bL workers.

Whatever the supposed equilibrium after deviation is (E1) or (E2). The expected income for the deviating worker is such that $\frac{1-e^{-q_{bH}^D}}{q_{bH}^D} e^{-q_{aH}^D} y_H - E_H$. Since we consider that the deviation of a single worker, let q_{bH}^D tend to zero. Then q_{aH}^D tends to a value pinned down by $(1 - e^{-q_{aH}} - q_{aH} e^{-q_{aH}})\, y_H = (1 - e^{-q_{bL}} - q_{bL} e^{-q_{bL}})\, y_L$, which gives $q_{aH} < \beta$.

We have that the group b workers have incentive to deviate to high skills whenever

$$\frac{1 - e^{-q_{bH}^D}}{q_{bH}^D} e^{-q_{aH}^D} y_H - E_H > e^{-\beta} y_L - E_L$$

However, this is always true when $q_{bH}^D \to 0$. Since we will have $e^{-q_{aH}^D} y_H - E_H > e^{-\beta} y_H - E_H > e^{-\beta} y_L - E_L$.

***Proposition 7**. **(equilibria)** For $\beta \leq \hat{\beta}_2$, there are two possible equilibria, depending on the value of β.*

For $\hat{\beta}_1 < \beta \leq \hat{\beta}_2$, the skill choice in equilibrium is such that all group a invest in high skills, all group b invest with mixed strategy in low or high skills. The firms' wage posting decisions are such that some firms attract only group aH workers, some firms attract both group bH and group bL workers while ranking the high skilled in priority to the low skilled.

For $\beta \leq \hat{\beta}_1$, in equilibrium all the workers invest with pure strategy in high skills. Some firms attract only group aH workers, and the rest of firms attract only group bH workers.

Proof.

It is clear from Lemma 5 that when $\beta \leq \hat{\beta}_1$, both groups still stay high skilled. As for $\hat{\beta}_1 < \beta \leq \hat{\beta}_2$, we know from Lemma 6 that the only possibility is that group b workers adopt mixed strategy and invest in high or low skills. Firms' preferece of the workers in economy then is such that Group $aH \succ$ Group $bH \succ$ Group bL, i.e., firms will always rank group a workers in priority to group b workers, and high skilled group b workers are ranked in priority to low skilled group b workers.

Now to show that at equilibrium, some firms will attract aH workers, the rest will attract both bH and bL workers, we need the following two results.

Fact 1. Using the same argument from LMD (2005), it could be shown that this economy has also no pooling equilibrium, in the sense that no wage can attract both group aH and group bH workers at the same time; Then if there is an equilibrium, it must be separating, i.e., some firms post one wage to attract group aH workers, while the rest post another wage to attract bH workers, and no group has incentive to deviate.

Fact 2. The aH firms, even if they choose to attract group bL workers, group bL workers will not apply to them: since low skilled workers are always ranked behind the high skilled workers, they can only be employed when those high skilled workers are not present; however, it is straightforward to verify that, the probability that aH workers are not present (i.e. $e^{-q^S_{aH}}$) is always smaller than the probability that the bH workers are not present (i.e. $e^{-q^S_{bH}}$). So that the expected payoff the bL workers can get at aH firms is strictly lower than the expected payoff they can get at bH firms, and this leads us to conclude that the bL workers only apply to bH firms.

For the firms attracting type aH workers, they have the following problem:

$$\begin{aligned} & \max_{w_{aH}} (1 - e^{-q_{aH}}) (y_H - w_{aH}) \\ \text{s.to} \quad & \frac{1-e^{-q_{aH}}}{q_{aH}} w_{aH} = U_{aH} \end{aligned}$$

which leads to the following equilibrium conditions:

$$Payoff_{firms,aH} = \left(1 - e^{-q^*_{aH}} - q^*_{aH} e^{-q^*_{aH}}\right) \times y_H$$

$$U_{aH} = e^{-q^*_{aH}} \times y_H$$

$$w_{aH} = \frac{q^*_{aH} e^{-q^*_{aH}}}{1 - e^{-q^*_{aH}}} \times y_H$$

For the firms attracting type bH and bL workers, , they have the following problem:

$$\begin{array}{lc} & \max_{w_{bH}, w_{bL}} \left(1 - e^{-q_{bH}}\right)(y_H - w_{bH}) + e^{-q_{bH}} \left(1 - e^{-q_{bL}}\right)(y_L - w_{bL}) \\ \text{s.to} & \frac{1-e^{-q_{bH}}}{q_{bH}} w_{bH} = U_{bH} \\ & e^{-q_{bH}} \frac{1-e^{-q_{bL}}}{q_{bL}} w_{bL} = U_{bL} \\ & w_{bH} = U_{aH} \end{array}$$

where the equilibrium w^*_{bH} is determined by $w^*_{bH} = U_{aH}$ (as in LMD (2005)), and the equilibrium q^*_{bL} is determined by maximization of the above program, and q^*_{aH} is determined by the firms' payoff indifference condition (as in LMD (2005)), and we have the following equilibrium conditions

$$\text{Payoff}_{firms,bH} = \left(1 - e^{-q^*_{bH}}\right)\left(1 - e^{-q^*_{aH}}\right) \times y_H + e^{-q^*_{bH}} \times \left(1 - e^{-q^*_{bL}} - q^*_{bL} e^{-q^*_{bL}}\right) y_L,$$

and

$$w^*_{bH} = U_{aH}$$

as well as

$$U^*_{bH} = \frac{1 - e^{-q^*_{bH}}}{q^*_{bH}} w^*_{bH}$$

where q^*_{bH} and q^*_{aH} are jointly determined by

$$\text{Payoff}_{firms,aH} = \text{Payoff}_{firms,bH}$$

Notice that $w^*_{bH} = U_{aH}$ can simply obtained by incentive compatibility condition.

In the current context, the equilibrium value of q^*_{aH}, q^*_{bH} and q^*_{bL} are determined by

$$
\begin{aligned}
\frac{e^{-q^*_{aH}} q^*_{aH}}{1-e^{-q^*_{aH}}} &= e^{-q^*_{bH}} \left(1 - \frac{1-e^{-q^*_{bL}} - q^*_{bL} e^{-q^*_{bL}}}{1-e^{-q^*_{aH}}} \frac{y_L}{y_H}\right) \\
e^{-q^*_{bL}-q^*_{bH}} y_L - E_L &= \frac{1-e^{-q^*_{bH}}}{q^*_{bH}} e^{-q^*_{aH}} y_H - E_H \\
\frac{\gamma}{q^*_{aH}} + \frac{1-\gamma}{q^*_{bH}+q^*_{bL}} &= \frac{1}{\beta}
\end{aligned}
$$

It could be verified that both group a and group b workers do not have incentive to deviate to apploy for each other's wage. Q.E.D.

Properties of the equilibrim.

(1) For the group a workers: $q^*_{aH} < q^S_{aH}$*. We also have* $w^*_{aH} > w^S_{aH}$*, and* $\frac{1-e^{-q^*_{aH}}}{q^*_{aH}} > \frac{1-e^{-q^S_{aH}}}{q^S_{aH}}$*. The expected payoff hence increases:* $U^*_{aH} > U^S_{aH}$*.*

(2) For the group b workers: $q^*_{bH} < q^S_{bH}$*, and* $q^*_{bH} + q^*_{bL} > q^S_{bH}$*. We also have* $w^*_{bH} > w^S_{bH}$*, and* $\frac{1-e^{-q^*_{bH}}}{q^*_{bH}} > \frac{1-e^{-q^S_{bH}}}{q^S_{bH}}$*. The expected utility also increases:* $U^*_{bH} > U^S_{bH}$*.*

(3) Operating profit for the firms decreases.

Proof.

For firms attract aH workers, while the remaining firms attract bH workers. The expected income of group a job seekers are $e^{-q^{(1)}_{aH}} y_H - E_H$.

In case (2), some firms attract aH workers, while the remaining firms attract both bH and bL workers. The expected income of group a job seekers are $e^{-q^{(2)}_{aH}} y_H - E_H$.

In case (3), all firms attract both aH workers and bL workers. The expected income of group a job seekers are $e^{-q^{(3)}_{aH}} y_H - \left(e^{-q^{(3)}_{aH}} - e^{-q^{(3)}_{aH}-q^{(3)}_{bL}}\right) y_L - E_H$.

In LMD (2005), the equilibrium value of q^S_{aH} and q^S_{bH} is determined by

$$
\begin{aligned}
\frac{e^{-q^S_{aH}} q^S_{aH}}{1-e^{-q^S_{aH}}} &= e^{-q^S_{bH}} \\
\frac{\gamma}{q^S_{aH}} + \frac{1-\gamma}{q^S_{bH}} &= \frac{1}{\beta}
\end{aligned}
$$

In the current context, the equilibrium value of q^*_{aH}, q^*_{bH} and q^*_{bL} are determined by

$$\begin{aligned}
\frac{e^{-q^*_{aH}}q^*_{aH}}{1-e^{-q^*_{aH}}} &= e^{-q^*_{bH}}\left(1-\frac{1-e^{-q^*_{bL}}-q^*_{bL}e^{-q^*_{bL}}}{1-e^{-q^*_{aH}}}\frac{y_L}{y_H}\right)\\
e^{-q^*_{bL}-q^*_{bH}}y_L-E_L &= \frac{1-e^{-q^*_{bH}}}{q^*_{bH}}e^{-q^*_{aH}}y_H-E_H\\
\frac{\gamma}{q^*_{aH}}+\frac{1-\gamma}{q^*_{bH}+q^*_{bL}} &= \frac{1}{\beta}
\end{aligned}$$

Within the parameter range $\beta\in\left(\hat\beta_1,\hat\beta_2\right)$, we prove $q^*_{aH}<q^S_{aH}$ by contradiction. Suppose $q^S_{aH}\le q^*_{aH}$, i.e., $e^{-q^S_{aH}}y_H-E_H\ge e^{-q^*_{aH}}y_H-E_H$. Then we can deduce $q^S_{bH}\ge q^*_{bH}+q^*_{bL}$, because of $\frac{\gamma}{q^S_{aH}}+\frac{1-\gamma}{q^S_{bH}}=\frac{1}{\beta}$ and $\frac{\gamma}{q^*_{aH}}+\frac{1-\gamma}{q^*_{bH}+q^*_{bL}}=\frac{1}{\beta}$.

Now, we rewrite the equality $\frac{e^{-q^*_{aH}}q^*_{aH}}{1-e^{-q^*_{aH}}}=e^{-q^*_{bH}}\left(1-\frac{1-e^{-q^*_{bL}}-q^*_{bL}e^{-q^*_{bL}}}{1-e^{-q^*_{aH}}}\frac{y_L}{y_H}\right)$ in another way: $\frac{e^{-q^*_{aH}}q^*_{aH}}{1-e^{-q^*_{aH}}}=e^{-q^*_{bH}-q^*_{bL}}\times\left(e^{q^*_{bL}}-\frac{e^{q^*_{bL}}-1-q^*_{bL}}{1-e^{-q^*_{aH}}}\frac{y_L}{y_H}\right)$. If we compare the just obtained equality with $\frac{e^{-q^S_{aH}}q^S_{aH}}{1-e^{-q^S_{aH}}}=e^{-q^S_{bH}}$, we notice that because the lefthand side of the first equation is smaller than the left hand side of the second equation in value, and the $e^{-q^{(2)}_{bH}-q^{(2)}_{bL}}$ is greater than $e^{-q^{(1)}_{bH}}$ in the righthand side, we should have that $\left(e^{q^*_{bL}}-\frac{e^{q^*_{bL}}-1-q^*_{bL}}{1-e^{-q^*_{aH}}}\frac{y_L}{y_H}\right)<1$.

We now show that actually $\left(e^{q^*_{bL}}-\frac{e^{q^*_{bL}}-1-q^*_{bL}}{1-e^{-q^*_{aH}}}\frac{y_L}{y_H}\right)>1$. Notice that this equation can be rearranged to $\left(1-e^{-q^*_{aH}}\right)y_H>\left(1-\frac{e^{-q^*_{bL}}q^*_{bL}}{1-e^{-q^*_{bL}}}\right)y_L$. However, this is always true, because by our hypotheses, we can obtain have $q^*_{aH}>q^*_{bL}$. (In fact, $\left(1-e^{-q^*_{aH}}\right)y_H>\left(1-\frac{e^{-q^*_{bL}}q^*_{bL}}{1-e^{-q^*_{bL}}}\right)y_L$ is actually implied by $\left(1-\frac{1-e^{-q^*_{bL}}-q^*_{bL}e^{-q^*_{bL}}}{1-e^{-q^*_{aH}}}\frac{y_L}{y_H}\right)>0$, which is necessary for $\frac{e^{-q^*_{aH}}q^*_{aH}}{1-e^{-q^*_{aH}}}=e^{-q^*_{bH}}\left(1-\frac{1-e^{-q^*_{bL}}-q^*_{bL}e^{-q^*_{bL}}}{1-e^{-q^*_{aH}}}\frac{y_L}{y_H}\right)$ to have a solution.)

So we could conclude that under the parameter range $\beta\in\left(\hat\beta_1,\hat\beta_2\right)$, we have $q^S_{aH}>q^*_{aH}$ and $q^S_{bH}<q^*_{bH}+q^*_{bL}$.

We show that $U^S_{aH}<U^*_{aH}$. It is because $U^S_{aH}=e^{-q^S_{aH}}y_H<e^{-q^*_{aH}}y_H=U^*_{aH}$.

Now we show that $q^*_{bH}<q^S_{bH}$. It is obtained from these two equalities: $\frac{e^{-q^S_{aH}}q^S_{aH}}{1-e^{-q^S_{aH}}}=e^{-q^S_{bH}}$ and $\frac{e^{-q^*_{aH}}q^*_{aH}}{1-e^{-q^*_{aH}}}=e^{-q^*_{bH}}\left(1-\frac{1-e^{-q^*_{bL}}-q^*_{bL}e^{-q^*_{bL}}}{1-e^{-q^*_{aH}}}\frac{y_L}{y_H}\right)$. Firstly, we compare the lefthandside of these two equalities: because $q^S_{aH}>q^*_{aH}$, we have $\frac{e^{-q^S_{aH}}q^S_{aH}}{1-e^{-q^S_{aH}}}<\frac{e^{-q^*_{aH}}q^*_{aH}}{1-e^{-q^*_{aH}}}$. Secondly, we compare the righthand side of these two equalities: because $\left(1 - \frac{1-e^{-q_{bL}^*} - q_{bL}^* e^{-q_{bL}^*}}{1-e^{-q_{aH}^*}} \frac{y_L}{y_H}\right) < 1$ and $\frac{e^{-q_{aH}^S} q_{aH}^S}{1-e^{-q_{aH}^S}} < \frac{e^{-q_{aH}^*} q_{aH}^*}{1-e^{-q_{aH}^*}}$, we have necessarily $e^{-q_{bH}^*} > e^{-q_{bH}^S}$, which implies $q_{bH}^* < q_{bH}^S$. **So that the employment rate is higher:** $\frac{1-e^{-q_{bH}^*}}{q_{bH}^*} > \frac{1-e^{-q_{bH}^S}}{q_{bH}^S}$.

We show that $U_{bH}^* > U_{bH}^S$. It is because $U_{bH}^* = \frac{1-e^{-q_{bH}^*}}{q_{bH}^*} w_{bH}^* = \frac{1-e^{-q_{bH}^*}}{q_{bH}^*} e^{-q_{aH}^*} y_H > \frac{1-e^{-q_{bH}^S}}{q_{bH}^S} e^{-q_{aH}^S} y_H = U_{bH}^S$.

We show that firms earn lower operating profits now. It simply sufficies to compare the equilibrium payoff of the aH firms, because we have the payoff indifference condition for firms. Recall that in equilibrium, the aH firms' payoff is $\text{Payoff}_{firms,aH} = \left(1 - e^{-q_{aH}^*} - q_{aH}^* e^{-q_{aH}^*}\right) \times y_H$, which is an increasing function of q_{aH}^*, now because $q_{aH}^* < q_{aH}^S$, we could conclude that firms earn less operating profit now. Q.E.D.

Claim. *For $\beta \in \left(\hat{\beta}_2, \hat{\beta}\right]$, if there exists an equilibrium, at least one group uses mixed strategy in skill investment.*

Proof. Holding b low skilled, group a will get $e^{-\gamma\beta}(y_H - y_L) + e^{-\beta} y_L - E_H$ if high skilled, and will get $e^{-q_{aL}^S} y_L - E_L$ if low skilled where $q_{aL}^S > \beta$. Then group a's dominant choice is to get high skilled.[37] However, (a high skilled, b low skilled) is not an equilibrium by Lemma 6. So that b will either choose high skilled for sure or mix between high and low skills.

Holding a low skilled, group b will get $e^{-\gamma\beta}(y_H - y_L) + e^{-\beta} y_L - E_H$ if high skilled, and will get $\frac{1-e^{-q_{bL}^S}}{q_{bL}^S} e^{-q_{aL}^S} y_L - E_L$ if low skilled, where $q_{bL}^S < \beta$. Then group b's dominant choice is to get high skilled. We now show that (a low skilled, b high skilled) is not an equilibrium: we are able to show that if group a deviates, they will get $e^{-0\beta}(y_H - y_L) + e^{-\gamma\beta} y_L - E_H > e^{-\beta} y_L - E_L$. So that both group will have incentive to mix. Q.E.D.

Proposition 8. *There are three thresholds $\hat{\psi}_{2b}$, $\hat{\psi}_{2a}$ and $\hat{\psi}_{3b}$ such that*

(1) For values of ψ such that $\psi \in \left(0, \hat{\psi}_{2b}\right)$, all the group a and group b workers are low skilled.

(2) For values of ψ such that $\psi \in \left[\hat{\psi}_{2b}, \hat{\psi}_{2a}\right)$, there is no equilibrium except for the mixed strategy one.

(3) For values of ψ such that $\psi \in \left[\hat{\psi}_{2a}, \hat{\psi}_{3b}\right)$, all group a workers become high skilled, and all group b workers remain low skilled.

(4) For values of ψ such that $\psi \in \left[\hat{\psi}_{3b}, 1\right)$, all group a workers and group b workers become high skilled.

[37] $e^{-\gamma\beta}(y_H - y_L) + e^{-\beta} y_L - E_H > E_H - E_L + e^{-\beta} y_L - E_H > e^{-q_{aL}^S} y_L - E_L$

(5) $\hat{\psi}_{2a}$ and $\hat{\psi}_{3b}$ are increasing in β; and $\hat{\psi}_{2b}$ is increasing in β for small values of β when y_H/y_L is not large.

(6) Define the threshold $\hat{\psi}$ of skill investment without discrimination as $\hat{\psi} y_H \frac{1-e^{-\beta}}{\beta} - E_H = \psi y_L \frac{1-e^{-\beta}}{\beta} - E_L$; we have $\hat{\psi}_{2b} < \hat{\psi} < \hat{\psi}_{3b}$.

Proof.

Firstly, notice that when ψ tends to zero, both groups choose to be low skilled, because $-E_L > -E_H$.

Then, starting with both groups being low skilled. We find that as ψ increases, group a has incentive to go to high skills if $\psi y_H - E_H < \frac{1-e^{-\gamma\beta}}{\gamma\beta} \times \psi y_L - E_L$, and group b has incentive going to high skills if $\psi y_H - E_H < e^{-\gamma\beta} \frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta} \times \psi y_L - E_L$.

Now, **holding group a's decision of remaining low skilled**, define $\hat{\psi}_{1b}$ by $\hat{\psi}_{1b} y_H - E_H = e^{-\gamma\beta} \frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta} \times \hat{\psi}_{1b} y_L - E_L$. If group b indeed deviates at this point, then its expected income drops to $\psi y_H \times \frac{1-e^{-(1-\gamma)\eta_b^*\beta}}{(1-\gamma)\eta_b^*\beta} - E_H$ at this point, where η_b^* is determined by workers' indifference condition. Notice that $\psi y_H \times \frac{1-e^{-(1-\gamma)\eta_b^*\beta}}{(1-\gamma)\eta_b^*\beta} - E_H$ is strictly smaller than $\hat{\psi}_{1b} y_H - E_H$, because of the within group competition effect captured by the term $\frac{1-e^{-(1-\gamma)\eta_b^*\beta}}{(1-\gamma)\eta_b^*\beta}$. Hence, staying in low skills is a dominant choice for group b. Then given this choice of group b, group a will stay low skilled also. (for the same reasoning group a will not even considern to invest till $\psi \geq \hat{\psi}_{1a}$, holding group b's being low skilled.)

Since whenever being indifferent, group b's expected payoff is decreasing in ψ, then group b may wait until $\psi y_H \frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta} - E_H = e^{-\gamma\beta} \frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta} \times \psi y_L - E_L$, **holding group a's remaining low skilled**; By this equality, we then define another threshold $\hat{\psi}_{2b}$. Then it is a dominant choice for group b to start to think going to high skilled from this point on, **holding group a' remaining low skilled.** Now we should check that indeed for $\psi < \hat{\psi}_{2b}$, group a does not have incentive to deviate to high skills.

Holding group b's decision of remaining low skilled, define $\hat{\psi}_{1a}$ by $\hat{\psi}_{1a} y_H - E_H = \frac{1-e^{-\gamma\beta}}{\gamma\beta} \times \hat{\psi}_{1a} y_L - E_L$. If group a deviates at this point, then group a's expected income drops to $\psi y_H \times \frac{1-e^{-\gamma\eta_a^*\beta}}{\gamma\eta_a^*\beta} - E_H$, which is strictly lower than $\hat{\psi}_{1a} y_H - E_H$, because of the within group competition effect captured by the term $\frac{1-e^{-\gamma\eta_a^*\beta}}{\gamma\eta_a^*\beta}$. Hence, staying in low skills at this point is a dominant choice for group a, **holding group b's decision of remaining low skilled**. Then given this choice of group b, group a will stay low skilled at this point. When y_H is sufficiently large compared to y_L, we have indeed $\hat{\psi}_{1a} < \hat{\psi}_{2b}$: group b will indeed remain low skilled till this point.

Since whenever being indifferent, group a's expected payoff is decreasing in ψ, then group a may wait until $\hat{\psi}_{2a} y_H \frac{1-e^{-\gamma\beta}}{\gamma\beta} - E_H = \frac{1-e^{-\gamma\beta}}{\gamma\beta} \times \hat{\psi}_{2a} y_L - E_L$, **holding group b's remaining low skilled** (otherwise there is between group competition effect), and this equality defines another threshold $\hat{\psi}_{2a}$. Now it could be check that $\hat{\psi}_{2a} > \hat{\psi}_{2b}$. Hence we are able to conclude that both group a and group b will remain low skilled for $\psi < \hat{\psi}_{2b}$.

Now, we know that group b may have incentive to deviate at $\hat{\psi}_{2b}$. Then, **holding group b being high skilled**, we check at the point $\hat{\psi}_{2b}$ what is group a's best response. It could be shown that group a's best response is being low skilled.[38] Because whenever group b chooses to be high skilled, group a may randomize between being high and low skilled, which will lower group b's expected income. Then the pure strategy Nash equilibrium is such that (a: low, b: low)**. Notice that there should also exists mixed strategy equilibrium, but in that case, we should require that all the group b workers randomize towards the same direction, otherwise the fraction will enter into the workers' matching function.** Also at this point, there is no discrete drop for group a workers' expected payoff.

Now we go beyond this point for $\psi > \hat{\psi}_{2b}$. We analyze when will group a workers switch to high skills. It could be shown that whenever $\psi > \hat{\psi}_{2b}$, for $\psi < \hat{\psi}_{2a}$, where $\hat{\psi}_{2a}$ is defined by $\frac{1-e^{-\gamma\beta}}{\gamma\beta}\psi y_L - E_L = \frac{1-e^{-\gamma\beta}}{\gamma\beta}\psi y_H - E_H$, there is no pure strategy equilibrium. And there is mixed strategy equilibrium. The group a's mixed strategy choice l_a of being low skilled is defined by $e^{-\gamma\beta}\frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta}\psi y_L - E_L = l_a \times \left(\psi y_H \frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta} - E_H\right) + (1-l_a)\left(\psi y_H e^{-\gamma\beta}\frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta} - E_H\right)$, and the group b's mixed strategy of being low skilled is defined by $\frac{1-e^{-\gamma\beta}}{\gamma\beta}\psi y_H - E_H = l_b \times \left(\psi y_L \frac{1-e^{-\gamma\beta}}{\gamma\beta} - E_L\right) + (1-l_b)\left(\psi y_L e^{-(1-\gamma)\beta}\frac{1-e^{-\gamma\beta}}{\gamma\beta} - E_L\right)$. We solve $l_a = \frac{\left[e^{-\gamma\beta}\frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta}\psi y_L - E_L\right] - \left(\psi y_H e^{-\gamma\beta}\frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta} - E_H\right)}{\left(\psi y_H \frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta} - E_H\right) - \left(\psi y_H e^{-\gamma\beta}\frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta} - E_H\right)}$, and $l_b = \frac{\left(\frac{1-e^{-\gamma\beta}}{\gamma\beta}\psi y_H - E_H\right) - \left(\psi y_L e^{-(1-\gamma)\beta}\frac{1-e^{-\gamma\beta}}{\gamma\beta} - E_L\right)}{\left(\psi y_L \frac{1-e^{-\gamma\beta}}{\gamma\beta} - E_L\right) - \left(\psi y_L e^{-(1-\gamma)\beta}\frac{1-e^{-\gamma\beta}}{\gamma\beta} - E_L\right)}$. And it could be verified that $l_a < l_b$. (denominator is larger, and the numerator is smaller.) The group b's payoff is $e^{-\gamma\beta}\frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta}\psi y_L - E_L$, while group a's expected payoff is $\frac{1-e^{-\gamma\beta}}{\gamma\beta}\psi y_H - E_H$, which is greater than group b's.

Now for values of $\psi \in \left[\hat{\psi}_{2a}, \hat{\psi}_{3b}\right)$, where $\hat{\psi}_{3b}$ is defined by $e^{-\gamma\beta}\frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta}\psi y_L - E_L = e^{-\gamma\beta}\frac{1-e^{-(1-\gamma)\beta}}{(1-\gamma)\beta}\psi y_H - E_H$, group a's dominante choice is to hit high skills, and group b to remain low skills. And finally, for values of $\psi \in \left[\hat{\psi}_{3b}, 1\right)$, group a's dominante choice is to hit high skills, and group b also switch to high skills.

[38] We can not assume that they will switch to high skilled.

(5) and (6) obtain by straightforward comparison. Q.E.D.