An Essay on the Sorting Function of the Educational System
1. Introduction

In an article titled “Cheating: another year, another A-levels scandal”, a reporter of The Economist wrote the following about the A-level, Britain’s college-entrance exam:

“Last year it was grade inflation: some said that too many A’s were being awarded, meaning either that the exams were too easy or the marking was too generous. This year it's the opposite: sixth formers who were predicted stellar A-level results have found themselves with a fistful of U’s instead of the A’s they expected.”

(Economist, September 19th 2002)

How to judge whether an entrance-related exam is too easy or too hard? Or, more generally, how to evaluate whether it is too easy or too hard to complete college, or other levels of education? And why does it matter?

Of course, it matters for several reasons. In this paper, we want to single out one of the possible reasons: it matters because the toughness of an education system affects its performance as a sorting mechanism. We believe that this reason is very important for understanding the questions we are concerned with, but it has not been explored seriously by economists and is often ignored by policy makers.

In any economy there exist different employers (employers providing different jobs) and different workers, and how workers and jobs are matched has a significant impact on the production of the economy. The matching between workers and jobs is in general inefficient because workers’ ability is often not observable to the employers. Modern societies, more or less, rely on education systems to mitigate this informational problem: the education level achieved by workers reveals some information about their characters and helps the society improve the quality of matching. And, ultimately, how well this function is performed depends on how well the school system distinguishes students.

---

1 A college entrance exam which is too easy might lower some students’ learning efforts in high schools. It will also bring too many college students and make teaching less effective (due to the dilution of school resources, the potential problem of getting students with more diversified learning ability, and so on.) The labor market might also have problems “absorbing” all college graduates when getting a college diploma becomes too easy. On the contrary, if the exam is too hard and thus too few students receive college education, the society might suffer from a lower rate of human capital accumulation and lack of upward social mobility.
The distinguishing ability of school systems is not given to people by nature. It is created by human beings, costs them a lot of resources to operate, and is subject to manipulation by people with different interests and intentions. Thus, the performance of the sorting function could change dramatically when it is run by different people under different circumstances.

Some examples from Taiwan illustrate how people rely on education to sort workers and how the operation of the sorting device is changed by circumstances. A job-market magazine recently conducted a survey on tertiary labor market in Taiwan and found that more than 75% of the employers have, to some degree, “upgraded” their new employment from college graduates to higher degree holders in recent years. However, less than 40% of them attribute the reason of doing so to their increased requirements for higher skills. Other employers have done so, as the editor explained, because

“... as the narrow gate is knocked wide open² ... the college diploma, for employers, can no longer serve the function of screening workers. On the other hand, the admission rates of graduate school entrance-exams are maintained steadily around 30%, thus employers naturally shift their hiring priority to graduate students.”

(Career, October 2004)

The increases in education requirements for elementary-school teachers in Taiwan are a good example of how education systems’ sorting function is changed by circumstances. In colonial years, one only had to study for 6 years to be a teacher: students who mastered the material went right back to teach. The required education increased gradually from 9 years to 12 years, then to 14 years and, finally in the 1990s, to 16 years. Of course one cannot deny that one reason behind the increasing requirements is to increase the amount of knowledge learned by teachers. However, the material taught in elementary schools changes at a much lower rate, and there is no evidence that younger-and-more-educated teachers perform better than the older-and-less-educated teachers. We believe that the desire to maintain teachers’ ability level should be a significant factor behind the increases in requirement: when entering and completing an educational system became less and less challenging, the sorting power of the education system got weaker and weaker. To maintain the ability level of the hiring pool, one needs to gradually increase the education requirement. At last, given that getting a college diploma became so easy (and the

²The admission rate of the college entrance exam increased from 30% in the 1980s to 87% in 2004 in Taiwan.
average “ability” of diploma holders got so low), we would not be surprised to see an increase to 18 years in the near future.

In short, the information-revealing function of the education systems is a major productive activity in modern societies. Its performance has significant impacts on the efficiencies of production, and the operation of this function is very responsive to individuals’ behavior and governmental policies. The function should be operated carefully and the evaluations for any policy changes should also take into consideration the policies’ impacts on this function. Nonetheless, this function is essentially ignored by economists and policy makers.

In this paper, we want to study how the sorting function of an education system works in the real world. In other words, we try to address the following questions: What can a sorting mechanism contribute to the society? What is the cost of operating it? What are the key factors that affect its performance? How does this function affect people’s life? How is its operation effected by government policies? To answer these questions, we will delineate the benefits and costs of operating the sorting mechanism, and investigate the consequences and implications when the operation reaches the equilibrium. We will also investigate how the sorting function affects people’s life by comparing the welfare levels between the situation with and without the sorting function. Through studying comparative statics we will try to explain why the mechanism is operated differently under different circumstances. And, in the end, we want to study the relationship between the operation of this machine and government policies.

The second section of this paper links our work to the economic literature. The model is presented in section 3, and the discussions of the model are in section 4. The last section summarizes the results of this paper and proposes the directions for further research.

2. Literature Review

The goal of our research is to study theoretically how the sorting mechanism works in the real world. We are interested in related issues such as why an economy needs a sorting device, how it operates, how well it operates, and how is its performance affected by government policies. To study these issues, we construct an analytical framework consisting of the following elements.
First of all, we use a simple production function to capture the importance of a good match between workers and jobs. We introduce a production technology which is complements between workers’ and employers’ contributions by simply assuming the production function to be the product of worker’s ability and employer’s technology. Thus, a society achieves a good matching (a positively monotonic matching between workers and jobs) would produce more efficiently than a society has a random matching.

Secondly, we introduce a two-sector economy to convey another significance of a good matching. We consider a modern sector that is able to make more use of workers’ ability but potentially has informational problem and a traditional sector that makes less use of ability but is free of informational problem. As a result, to produce effectively, an economy not only has to match workers to the right jobs, but also has to allocate workers to the right sector. Combining the first two elements, we set a stage where the quality of matching between workers and jobs plays the key role.

Thirdly, we consider a very realistic and commonly observed problem in labor market: employers cannot (at least not perfectly) observe workers’ ability level. This informational problem could severely lower the quality of matching between jobs and workers, both within sector and between sectors, in an economy, and hence allows a sorting mechanism that can improve the quality of information to contribute positively to the economy.

We add the last element into our model to make sure an education system can perform a sorting function: we assume that education systems create a negative relationship between workers’ ability and their costs of taking education. Facing the benefit and cost of taking education given to them, workers select themselves into different educational levels (and thus different jobs). How this negative relationship is created determines the cost of operating the sorting machine. With this cost structure, in addition to the machine’s potential benefits created by the first three elements, we put in our model a tradeoff that allows us to study many aspects of the performance of this sorting mechanism.

None of these elements is new to economists. This literature survey below shows that all the key elements in our model had attracted a lot of attentions and efforts from economists. What is needed to be done so that the operation of the sorting machine can be properly studied is a work that pieces all the elements together. This paper is an attempt to do so.
The following connect this model to the work already done by other economists.

2.1. The human capital literature

The human capital approach, originated by Becker (1964), is still the main-stream approach to studying education in economics. The basic theoretical framework of this approach is that workers choose education levels to maximize their lifetime (net) income by balancing the cost and the benefit of taking education. Differences in workers’ characters, such as their tastes, abilities, backgrounds, economic opportunities and restrictions, bring different benefits and costs of taking education and eventually lead to different education choices. Becker’s theoretical framework triggered a huge amount of empirical work. Becker’s fellow economists devote a lot efforts to estimate the causal effect of education on income, or more precisely, to estimate the rate of return to individuals’ investment in education.

This framework shares two important elements with our model. One is the positive relationship between an individual’s ability and the education level he completes. In our model, the relationship is a valuable outcome produced by the sorting function. For human-capital economists, this relationship is a nightmare: together with the sad fact that the researchers have no data for workers’ “innate ability” and the presumption that workers’ difference in ability is well reflected in their earnings, this relationship in principle biases the estimation of the rate of return to investment in education.

The other element that we share is that the education achievement is a result of self-selection. In our model, individuals choose whether or not to invest in education in order to send a signal to the employers about their ability. In the human capital literature, the decisions are also assumed to be made by individuals. The difference is that in the human capital framework the choices might be based on more kinds of considerations. Put it in a more precise way, human capital framework does not exclude the sorting function of education: the causal relationship between high education and high income can come either from learning or from signaling. However, it seems that human capital economists do not think the difference is so important that a distinction is worth to make.³

³ Human-capital economists believed that the relationship between productivity and ability is not significant. Griliches (1977) cited a paragraph from *The Wealth of Nations*, along with his econometric
The most significant difference between our model and the human capital literature is that the theoretical framework of the human capital literature does not consider matching as a problem: workers’ productivity depends mostly (if not entirely) on their traits, not on the features of their jobs. The matching is either assumed to be done perfectly at no cost, or assumed to be useless. Thus, a sorting device, may have some private values in distributing income, has no social value.

Even though human capital economists are well aware of the relationship between ability and education and try very hard to explore the implication of this relationship, their focus is very different from ours.

2.2. Signaling literature

Another important element of our model is the problem of asymmetric information. This issue had been studied intensively in the signaling literature inspired by Spencer (1973, 1974).

The basic structure of the signaling models is very similar to our model. The labor market considered by the signaling models is one where workers have different levels of ability and the ability level is private information, i.e., cannot be observed by the employers. And since low-ability workers tend to have high cost of taking education, the level of education can serve as a signal that reveals information about workers’ ability.

In spite of this similarity, we found that the setup of the signaling models is in general inappropriate for answering questions that we are concerned with.

For example, the quality of matching is a key factor in our study, but matching is not a problem at all in the conventional signaling models.\(^4\) In our model, the amount of good jobs (high-

---

technology jobs) is limited and, thus, they are scarce resources. Therefore, it is important for the society to be able to efficiently allocate these jobs. In Spence’s model (and followed by most of the later work in this literature), the employers have no interest over workers’ ability and the employers are assumed to be able to offer jobs to accommodate all kind of workers (if it is profitable to hire them). More precisely, the signaling models do not really have an employers’ side. The employers are simply assumed as a transfer mechanism: the employers always make zero profit and thus always pay back a wage exactly equals the (expected) productivity brought by the workers. Thus better information can only affect the distribution of income among workers; it can never help the demand side of the labor market.

In short, in conventional models, signaling affects income distribution, not production. Or, we can say that the signaling models are designed to study the issues on the distribution of income, and are not so suitable for studying the production side.

Some recent work on asymmetric information, such as Gale (1992) and (2001), does have matching flavor. These papers considered heterogeneous employers, like our model, and studied the labor market outcome under asymmetric information. Our model differs from these models in that we consider a situation that workers’ types outnumber signals’ types and they discussed the opposite case. Our setup naturally generates a semi-separating equilibrium (imperfect sorting), while the setup of those papers usually aim for a separating equilibrium (perfect sorting). An imperfect sorting equilibrium allows us to study issues such as the benefit and cost of improving quality of matching, the wage differential within educational groups, and political-economy analysis for interest-conflicting within groups. A model designed to generate a perfect sorting equilibrium has no space for study on the benefit/cost of improving quality of matching, and perfect sorting means homogeneous property within each group and does not allow within-group wage differential and interest conflict.

To summarize, our model differs from the signaling models in that it considers heterogeneous employers and that it generates a semi-separating equilibrium. These variations allow us to introduce the value of a “good” matching and to study the benefit and cost of changing the quality of matching.

---

5 In other words, we study the case that workers’ ability level is a continuous variable and the levels of education that workers can choose are discrete and limited.
So far, it seems that it is our emphasis on the value of a good matching between heterogeneous employers and employees that separates our model from the conventional economics literature (the human capital models and the signaling models.) This thought leads us to check if we can connect our work to the economics research dealing with the problem of matching.

2.3.1. The matching literature

Economists had long acknowledged that two-sided matching can better portray the way two sides of the labor market are brought together than the conventional frictionless-market mechanism, at least when study some subjects. Matching process has been studied by micro- and macro-economists.

The micro approach is mainly concerned with the “results” of the matching process. The original work comes from mathematicians. Gale and Shapley (1962) were the first to study the matching process. They used marriage and college admission as examples to study matching algorithms and the natures of the matching outcomes brought by those algorithms. Roth and Sotomayor (1990) introduced the study of matching into economics. They studied the mechanism designed to match medical interns and hospitals. In the literature follows their works, economists focused on such issues as the existence of the “stable” equilibrium to the matching games, the algorithm needed to reach the stable equilibrium, the patterns of those stable matches, and the factors that have impacts on the patterns.

The macro approach is concerned more with the waiting costs that occur during the matching process. These costs serve as a source of the friction in labor market which in turn causes natural unemployment in an economy. The factors that affect this waiting cost also affect the performance of the labor market. Javanovic (1979) used a matching model to explain unemployment and other labor market features like turnovers. Petrongolo and Pissarides (2001) surveyed works in this field.

The micro approach is closer to our work because it is concerned with the “quality” of matching and its welfare results. However, most of the matching models have no informational problems.
Furthermore, the key element of our model, the in-advance sorting of workers that improves the quality of matching, is not addressed by this literature.\textsuperscript{6}

2.3.2 Matching with informational problem

One serious attempt to connect matching process to asymmetric information comes from Inderst (2001, 2002). The main goal of Inderst’s work is to extend the study of contract design mechanism, signaling and screening mechanisms, from the conventional, small-scale circumstance to a large-scale environment.

The main challenge of this extension is to allocate the dominating power of writing contracts to the two sides involved. The conventional, small-scale setup gives the power exclusively to one side of the contracting parties, either to the principal side or to the agent side, and then studies how that authorized party designs a contract to maximize the rent that can be attracted from the contract. In a large-scale environment with different employers and employees, it seems unreasonable that the power is owned exclusively by one side. In other words, Inderst tried to extend contract design mechanism to a big-scale environment where neither principals nor agents have dominant contracting power.

Inderst fixed this problem by introducing a matching process into labor market. Employer and employee are paired by a matching process, and then one side of the market is given the contracting power by random. The empowered side offers a take-it-or-leave-it contract to the other party. If the contract is accepted, the story ends. If the contract is rejected, then the matching would be dismissed and both parties would have to wait for another round of matching. Similar to the setup of matching models by macroeconomists, waiting is costly and thus it won’t be optimal for people to wait for too long to get the contracting power.

Both agents and principals have alternatives outside this labor market. And the outside chances, together with other factors such as the discount rate and the probability of allocating the contracting power, determines both the quantities and the qualities of principals and agents who

\textsuperscript{6} Similar to our model, papers in this literature contain such key words as matching and sorting. However, theses words in those papers are used to describe the relationship between a matching process and the assortment results caused by it; while in our model they are used to describe how sorting improves matching results. In short, we consider a sorting process before matching, and the matching literature studies the assorted outcomes after matching.
would participate in the market. The quantities and qualities in turn determined the tightness of the market (the frequency of getting a match) and the expected contracting power of different players. And eventually these factors determine the strategy about making and accepting/rejecting offers. The terms of the contracts and the compositions of players are determined together.

In some sense, we can say that Inderst’s model is a complicated version of sequential bargaining game. The difference is that, in Inderst’s model, players play game with different opponents in different rounds, and that the contracting power doesn’t come in turn.

The main difference between Inderst’s papers and our work is that screening/signaling mechanism is basically independent to the matching process in his research. What matching process does is nothing but bring agent and principal together, and then the empowered player tries to take advantage from contracting through screening and/or signaling. In our model, a matching itself implies the agreements and fulfillments on the terms of the contracts, and the signaling takes place before matching and can improve the quality of matching.

The other difference is that Inderst treated the demand side of the labor market in the same way with the conventional signaling models. Employers are assumed to be identical and can offer unlimited numbers of all kind of jobs. The quality of matching isn’t an issue in his papers.

Furthermore, both his work and ours try to look at contracting mechanism working in a large-scale environment and let both contracting sides somehow split the rents. Inderst lets the rents shared randomly, and our model, follows classical economics, let all employers and employees beside the marginal ones grab some rents.

2.4. The dual-economy literature

The dual economy setup used in our model can be dated back to the early study of development economics. Lewis (1954) introduced the dual-economy (a modern /industrial sector and a traditional/agricultural sector) framework to study the unlimited labor supply in the modern sector of the under-developing countries. Harris and Todaro (1970) introduced some institutional factor into the modern sector to explain the unemployment in the modern sector. Following Harris and Todaro’s idea that the market friction exists in the modern sector but not in the traditional sector, economists commonly adopted a framework about the dual-economy system: the modern sector
has higher productivity than the traditional sector but the modern sector also has more serious market-friction problems than the traditional sector.

In most research, the market friction is assumed to be the asymmetric information problems. Banerjee and Newman (1998), Udry (1994), and Townsend (1994) found that asymmetric information causes more serious problems in the credit market in the modern sector than in the traditional sector. Shapiro and Stiglitz (1984) and Burlow and Summers (1985) studied the implications of the moral hazard problem in the modern-sector labor market. Eicher and Kalaitzidakis (1997) and Eicher (1999) used the relatively more serious adverse-selection problem in the modern-sector labor market to study issues on economic growth and international economics.

The approach used by Eicher and Kalaitzidakis (1997) and Eicher (1999) is very close to our model: a dual-economy system, the modern sector has high productivity from hiring workers but the fact that workers’ ability level is not observable causes some problems, and the traditional sector has lower productivity but is free of the informational problem. However, beyond that point, our model and their work are driven toward different directions because of different focuses. Their work left the informational problem “unsolved” and moved forward to study factors that affect the severity of the informational problem and their impacts on issues like economic growth, human capital accumulation, technology adoption, and international movements of goods and factors. While our model focuses on the problem per se and tries to study how a sorting mechanism mitigates the informational problem.

In short, Eicher and Kalaitzidakis’ paper and ours share many basic structures of the informational problem, but we have different concerns. Our work focuses on mitigating the informational problem, and their papers are not concerned with solving the problem.

2.5. Selection-by-education model

A paper written by Velden (2003) shares many viewpoints and elements with our model. The main theme of Velden’s paper is that it is crucial for an economy to assign workers to the “right” jobs, but this task is hard to accomplish when employers cannot observe workers’ ability. Schools (the educational institutes) can sort and label workers (students) according to their abilities and the employers can read the label (the diploma received) to estimate workers’ ability. The sorting
and labeling are not perfect, and the accuracy depends the precision of the measuring (how well schools distinguish students) and the number of labels available (how many kinds of diplomas are offered by the educational institutions).

In short, the core argument of Velden’s paper is that the education system can help societies sorting workers, and how well the function performs depends on how societies run schools. And this is exactly the point we want to make in this paper. We also share one comment made by Velden:

“Given the fact that selecting students and sorting them into different tracks is one of the major functions of the education, it is surprising to see how little is known about the effect on labor market outcomes.”

Although we have the same goal to achieve, Velden’s paper and our models are different in several places. And thus our researches bring different results and different implications.

The quality of matching matters in both our model and Velden’s paper. However, it matters in different ways. In Velden’s paper, employers can tailor jobs for workers with different labels. Employers design a job especially for a label so that the job can be best utilized by the worker with the mean ability of that label. If workers’ ability deviates from the designed level, no matter too high or too low, the efficiency of production would be lowered dramatically. In short, the inefficiency of production caused by informational problem comes from the inaccuracy of labeling in Velden’s paper. However, in our model, the inefficiency of production comes from giving high technological jobs to low expected ability workers, or vice versa.

Another way to describe this difference is that Velden assumed that there is no limit for the employers to create jobs, while, in our model, the quantity and quality of jobs are fixed. In short, Velden’s problem is about producing the right jobs, and our problem is about distributing jobs in the right way.

Another difference between our model and Velden’s is that there is no self-selection in his paper: the education system labels students based on its one-side measure of students’ ability. In our model, the education system sets the “prices” of different diplomas and workers decide which diploma to “buy” by evaluating the associated benefits and costs. We believe our approach is more realistic and is also closer to economists’ view regarding how people’s education levels are
determined. Furthermore, Velden’s model is narrower than ours in the sense that it doesn’t consider any cost for performing the measuring/labeling function, and therefore there is no channel to studying our key concern: the tradeoff between benefit and cost of the information-revealing mechanism.

To conclude, the elements of our model have been investigated by economists in many different fields. Because of different interests and/or different focuses, their researches fail to address the questions we are concerned. And the main contribution of our paper to the literature is to assemble all the key elements so that the operation of the sorting mechanism can be properly studied.

3. The Model

The first subsection describes the setup of the model. Subsection 3.2 studies the model with no informational problem. Subsection 3.3 introduces asymmetric information and studies its consequences. And the last subsection studies the model when a sorting mechanism is available.

3.1. Setup

The Economy

The economy consists of two sectors, a modern sector where employers hire employees to manufacture product, and a traditional sector where workers are self-employed. The modern sector potentially has informational problem; and the traditional sector is free of this problem. All workers can produce more when they occupy good positions (when they are hired by good employers). And the total output of the economy is simply the total of the two sectors’ output.

The Workers

The assumption for workers has nothing significantly different from the conventional models concerned with heterogeneous workers. In our model, workers (potential employees) have different ability levels which can be summarized into a single, continuous, variable $s$. We assume $s > 0$ and $s \in [s, \bar{s}]$. Their ability levels are positively related to their self-employed income in the traditional sector. The income from self-employed then serves as their opportunity
costs (or, in other words, serves as their reservation wages) of working in the modern sector. For simplicity, we assume that the reservation wage simply equals the ability level, \(s\). Different assumptions about whether or not the ability level \(s\) is observable to employers give our model different results.

The Employers

The assumption regarding the employers singles our model out from most of the conventional models. We assume that each employer wants to hire one and only one worker.\(^7\) We also assume that employers are different in their capacity for making use of workers' ability. This difference in capacity can be thought as the result of different production technology, different positions of the job in a bureaucracy, and different levels of specialization in an organization, and so on. We use a single, continuous, variable \(\theta\) to characterize this capacity. We assume \(\theta \in [\underline{\theta}, \overline{\theta}]\), also \(0 < \underline{\theta} < 1\) and \(\overline{\theta} > 1\). The profit function for an employer with \(\theta\) is assumed to be \(\pi(\theta, s) = \theta E(s) - W\): the profit equals revenue, the product of employer's capacity and the expected value of worker’s ability, minus wage paid to the worker hired.\(^8\)

The assumption for employers is not as innocent as it looks. It is totally different from most of the conventional labor economics models. Under this assumption, a worker’s productivity depends not only on his personal features (such as human capital), but also on the position he occupies. Together with the assumption that employers have different capacity and that the number of “good jobs” is not unlimited, the matching between workers and jobs matters for the efficiency of the society and thus a sorting mechanism that could enhance the quality of matching between workers and jobs can improve the efficiency.

The distributions of employees and jobs

For simplicity, we assume that the number of potential employees equals the number of potential employees and both of them equal to one. This turns the frequency functions of their distributions into probability functions. Let the probability density function of the (employees’) ability

\(^7\) An employer who needs to hire more than one employee can be thought as different employers, as long as the employer cannot rearrange job assignments after hiring workers.

\(^8\) The setup for employers is borrowed from the model with consumers of different tastes for product with vertically differentiated quality. See Tirole J. (1990) *The Theory of Industrial Organization*. p.96.
distribution be \( g(\cdot) \) and the (employers’) capacity distribution be \( f(\cdot) \). And the corresponding probability density functions are assumed to be \( G(\cdot) \) and \( F(\cdot) \) respectively.

We assume that both \( g(\cdot) \) and \( f(\cdot) \) are continuous at the defined domains. We should have the property of \( \frac{d}{ds} E(s) > 0 \) and we assume that \( \frac{d^2}{ds^2} E(s) < 0 \) globally. (\( E(s) \equiv \int_{s}^{\infty} x g(x) dx \cdot [G(s)]^{-1} \)).

That is, when workers are hired from the bottom of the ability distribution, the average ability of the hiring pool increases as the upper bound of the hiring pool rises, and the rate of increasing decreases as it rises.

The Sorting Mechanism

We assume there is a sorting mechanism (an educational system or simply a school) that can be used to reveal information about employees’ ability in case it is not observable. The sole function of the school is to give people a hard time.\(^9\) In other words, the cost of education is the pain workers have to endure to complete the school. Like all existing signaling models, in order for this mechanism to work, we assume that the cost of education (the level of pain) is negatively correlated with workers’ ability. Namely, for any level of toughness in school, high ability workers suffer less than low ability workers. Thus high-ability workers can do something that low-ability workers cannot do, completing school, to send a message about their ability to the employers.

We use a function \( C = C(s; \phi) \) to characterize the educational cost. We assume the cost function is continuous and at least twice differentiable at \( s \). The previously made assumption that educational cost is negatively correlated with ability gives us \( C_s(\cdot) < 0 \), and we further assume \( C_{ss}(\cdot) > 0 \).\(^10\) In other words, we assume that, for any given cost structure, high ability workers endure less pain than low ability workers and this difference in the level of pain decreases as workers’ ability gets higher. The parameter \( \phi \) measures how tough the education system is, and

---

\(^9\) “Monday morning found Tom Sawyer miserable. Monday morning always found him so - because it began another week’s slow suffering in school.” … “Huckleberry came and went, at his own free will. ... he did not have to go to school or church, or call any being master or obey anybody; ... In a word, everything that goes to make life precious that boy had. ...” The Adventure of Tom Sawyer, Mark Twain.

\(^10\) \( C_s(\cdot) \equiv \frac{\partial C(\cdot)}{\partial s} \) and \( C_{ss}(\cdot) \equiv \frac{\partial^2 C(\cdot)}{\partial s^2} \).
we assume $\phi \in R$ and that $C(\cdot)$ is continuous at $\phi$.$^{11}$ The assumptions regarding $\phi$ is $C_{\phi}(\cdot) > 0$ and $C_{\phi\phi}(\cdot) < 0$. That is, a tougher educational system not only makes everyone suffer more, but also increases the difference of cost between workers of different ability. The cost function can be illustrated by <Figure 1>. There are two cost functions in <Figure 1>, with $\phi_2 > \phi_1$.

3.2. Complete Information

First of all, we consider the complete-information case: the situation that the capacity ($\theta$) of all employers and the ability ($s$) of all employees are observable to everyone. In this case, the labor market situation resembles one of the scenarios well studied by the matching literature, and we can analyze this case by simply applying some standard results of the literature.

Two important results from the matching literature are relevant to our study here. The first one is that a matching must be “stable” to be a reasonable prediction for the outcome of a matching game.$^{12}$ The second one is that for a two-sided, one-to-one matching game satisfying the following two conditions: the joint production function for any pair is supermodular$^{13}$ and the

\[
\forall z, z' \in R^k, f(z \vee z') + f(z \wedge z') \geq f(z) + f(z') \quad \text{where } z \vee z' \text{ denotes the component-wise maximum and } z \wedge z' \text{ the component-wise minimum of } z \text{ and } z'. \text{ If } f \text{ is smooth, supermodularity is equivalent to the condition}
\]

\[f(z \vee z') = \max f(z), f(z') \quad \text{and } f(z \wedge z') = \min f(z), f(z').\]

---

11 Notice that this does not mean that workers can choose among educational systems with different toughness levels. In this paper, there is only one system with a given toughness level. Different $\phi$ simply gives us possibly different toughness level for this system.

12 A matching is stable if it is not blocked by any individual or any pair of agents. (Roth and Sotomayor (1990)) They also stated that “Only stable matchings will occur.” (Although they admitted that it is probably put a bit too simply.)

13 A real-value function $f : R^k \rightarrow R$ is supermodular if and only if for any $z, z' \in R^k$, $f(z \vee z') + f(z \wedge z') \geq f(z) + f(z')$ where $z \vee z'$ denotes the component-wise maximum and $z \wedge z'$ the component-wise minimum of $z$ and $z'$. If $f$ is smooth, supermodularity is equivalent to the condition
output is transferable within any pair, the only stable outcome is the positive assortative matching: if $\alpha > \tilde{\alpha}$, then $\alpha$ ’s match $M(\alpha)$ is greater than $M(\tilde{\alpha})$, $\tilde{\alpha}$ ’s match.\footnote{Legros and Newman (2003) demonstrated and intuitively explained this result.}

The complete-information case we consider here is a two-sided, one-to-one matching game. The production function in our setup obviously satisfies supermodularity. And wage payments are used to transfer output between paired employer and employee. Thus, the only stable result of this case is the positive assortative matching: the employer with the highest $\theta$ hires the employee with the highest $s$, the employer with the second highest $\theta$ hire the employee with the second highest $s$, and so on.

In some matching games, agents may find that it is better for them not to participate in the games. Likewise, in this case of our model, workers (to be) matched with employers with $\theta < 1$ would find that it is better for them to stay self-employed than to be employed.

**Result 1:** the outcome of the complete-information case can be summarized as follows.\footnote{We sketched a proof for this result using the spirits of the matching literature. See Appendix (A) for the proof.}

1. Every employer with $\theta \geq 1$ will hire one worker.
2. Workers will be hired from the top of the ability distribution. And the workers who are not hired remain self-employed.
3. The hiring employers and the hired employees are matched in a positively monotonic manner.

We close this subsection by deriving the output level for the economy in this case. Let $M(s)$ be the matching function that assigns workers to employers. For a positive assortative matching, it must be true that $M'(s) > 0$. The output of the economy under complete information is

$$\int_{s^L}^{s^H} M(x) x g(x) dx + \int_{s^L}^{s^H} x g(x) dx$$

Where $s^{FB}$ is the last worker that gets hired. In other words, $s^{FB}$ is the employee hired by the employer with $\theta = 1$. ($s^{FB} = M^{-1}(1)$) The output level under perfect information is obviously the social optimal level, and it serves as the benchmark to study the efficiency issues of the later cases.
3.3. Incomplete Information with No Sorting Mechanism

This and the next subsections consider the case when workers’ ability level, $s$, is not observable to employers. In other words, all employees look identical to employers. This subsection studies the outcome of this asymmetric information when no information-revealing mechanism is available. In the next subsection we study the problem when a sorting mechanism is accessible.

To focus on the issue that we are most interested in, we first make an assumption about the way the labor market works.

**Assumption 1 (A1):** In equilibrium, there is only one wage in the labor market, and all employers and employees behave as price-takers.$^{16}$

Let the market wage in the modern sector be $W$.

The problem faced by the worker with the ability level $s$ is to choose the bigger one from the set $\{s, W\}$ $^{17}$: the worker will join the modern sector and be hired if $s \leq W$, and he will remain self-employed if $s > W$. Moreover, if $\hat{s} > W$, then for all $s > \hat{s}$ it must be true that $s > W$. In other words, if $\hat{s}$ chooses not to be hired, then all workers with $s > \hat{s}$ would also choose not to be hired. In short, if in the equilibrium some workers choose to be hired and others choose to remain self-employed, then the employed workers would be from the bottom of the ability distribution and the self-employed workers would be from the top of the distribution.$^{18}$ In this case, the set of employees who are willing to be hired at wage $W$ (the hiring pool) is $S(W) = \{s : s \leq W\}$. And the workers in this pool can bring an expected level of ability 

$$E(W) = \int_{\frac{1}{2}}^{W} xg(x)dx \cdot \left[\int_{\frac{1}{2}}^{W} g(x)dx\right]^{-1} = \int_{\frac{1}{2}}^{W} xg(x)dx \cdot [G(W)]^{-1}. $$

For any $[W, E(W)]$, the problem faced by the employer with $\theta$ is to choose the bigger one from the set $\{0, \theta \cdot E(W) - W\}$: the employer will hire a worker if $\theta \cdot E(W) - W \geq 0$, and he will not

---

$^{16}$ See Appendix (B) for more discussions on this assumption.
$^{17}$ Remember the self-employed income for a worker with ability $s$ is simply assumed to be $s$.
$^{18}$ This is a typical adverse-selection result.
hire one if \( \theta \cdot E(W) - W < 0 \). Furthermore, if \( \hat{\theta} \cdot E(W) - W \geq 0 \), then for all \( \theta \geq \hat{\theta} \) it must be true that \( \hat{\theta} \cdot E(W) - W \geq 0 \). In other words, if it is profitable for \( \hat{\theta} \) to hire a worker, then it must also be profitable for all employers with \( \theta \geq \hat{\theta} \) to hire one. That is, if in equilibrium some employers hire a worker and others do not, then those who hire one would be from the top of the capacity distribution and those who do not would be from the bottom of the distribution. The set of employers who are willing and able to hire worker at wage \( W \) is \( \Theta(W) = \{ \theta : \theta \cdot E(W) \geq W \} \).

**Lemma 1**: At least one employee would be hired and at least one employee would not be hired.\(^{19}\)

With Lemma 1, we are ready to characterize the equilibrium. And we called this equilibrium the adverse-selection equilibrium (or simply the AS equilibrium) hereafter.

The adverse-selection equilibrium can be characterized by a wage rate \( W^{AS} \) and a pair of \((\theta^{AS}, s^{AS})\) that satisfies

\[
\Theta(W^{AS}) = S(W^{AS}), \quad \text{or} \quad 1 - F(\theta^{AS}) = G(s^{AS})
\]

(1)

(The measure of workers demanded equals the measure of workers supplied.)

\[
\theta^{AS} \cdot \int_x^\infty xg(x)dx \cdot \left[ \int_\frac{x}{\theta^{AS}}^\infty g(x)dx \right]^{-1} = \theta^{AS} \cdot E(s^{AS}) = W^{AS}
\]

(2)

(The expected profit of the marginal employer, \( \theta^{AS} \), is zero.)

\[
W^{AS} = s^{AS}
\]

(3)

(The marginal worker, \( s^{AS} \), is indifferent between being hired and being self-employed.)

Combining (2) and (3), we can get

\[
\theta^{AS} \cdot E(s^{AS}) = W^{AS} = s^{AS}
\]

(4)

Sometimes we find it more intuitive to read (4) as

\[
\theta^{AS} = s^{AS} \sqrt{\frac{E(s)}{s}}
\]

(4')

The equilibrium level of \( \theta \) and \( s \) can be obtained by solving (1) and (4) simultaneously.

\(^{19}\) See Appendix (C) for proof.
From (1), \( \theta^{45} \) and \( s^{45} \) are negatively correlated. By intuition, a greater \( \theta^{45} \) means fewer employers and thus, to have the market cleared, there must also be fewer workers hired, which means a smaller \( s^{45} \). To see it mathematically, we totally differentiate (1),

\[
-F'(\theta^{45}) \cdot d\theta^{45} = G'(s^{45}) \cdot ds^{45},
\]

and get

\[
\begin{bmatrix}
\frac{d\theta^{45}}{ds^{45}} \\
\frac{G'(s^{45})}{-F'(\theta^{45})}
\end{bmatrix} = \begin{bmatrix}
g(s^{45}) \\
-F(\theta^{45})
\end{bmatrix} < 0.
\]

From (4), \( \theta^{45} \) and \( s^{45} \) are positively correlated. Intuitively, a higher \( \theta^{45} \) means the marginal employer has higher \( \theta \). He can afford to pay higher wage and higher wage can attract more workers, which means a higher \( s^{45} \). To see this mathematically, we totally differentiate (4'),

\[
d\theta^{45} = \frac{\hat{E}[s^{45}]}{s^{45}} \cdot ds^{45},
\]

and get

\[
\frac{d\theta^{45}}{ds^{45}} = \frac{\hat{E}[s^{45}]}{s^{45}} > 0. \quad 20
\]

Since \( \theta \) can be written as a continuous and monotonic function at \( s \) from both (1) and (4) \( 21 \), the equilibrium exists and is unique, as shown in < Figure 2 >.

![Figure 2](image)

Intuitively, the equilibrium can be derived by following a few steps:

1. Start with the highest possible wage: \( W = \bar{s} \).
2. It cannot be an equilibrium wage: all workers want to be hired but not all employers can afford to hire workers. (Namely, the measure of labor supplied (\( Q^S \)) is greater than the measure of labor demanded (\( Q^D \)).)

---

20 This comes from our assumption that the average ability of the hiring pool increases at a lower rate than the wage does as the market wage increases.

21 We have just checked the monotonic property. See appendix (D) for the continuous property.
3. To move toward the equilibrium, the wage must get lower.
4. As the market wage gets lower, more employers afford to hire workers and fewer workers want to be hired. (Q^s decreases and Q^D increases.)
5. Since the increase and the decrease are continuous, we can find a wage to clear the market.

In the adverse-selection equilibrium, the output level in the society is
\[ \int_{σ^a}^g E(s^{AS})yf(y)dy + \int_{s^a}^x xg(x)dx \]

We need one more thing to summarize and compare this outcome to the case of complete information.

**Lemma 2:** \( θ^{AS} > 1 \).

_Proof:_

In equilibrium, \( θ^{AS} \cdot E(s^{AS}) = W^{AS} = s^{AS} \)

\[ ⇒ θ^{AS} = s^{AS} \sqrt{\frac{E(s^{AS})}{E(s^{AS})}} > 1 \]

Q.E.D.

**Result 2:** In the adverse-selection equilibrium, the employers from the top of the capacity distribution hire workers from the bottom of the ability distribution. Workers in the upper part of the ability distribution remain self-employed. Compared to the complete-information case, the adverse-selection result is inefficient in three ways:

1. Macro-mismatch: High \( θ \) employers hire “wrong” workers: employees are hired from the bottom of the ability distribution.
2. Too few employers hire workers. \( (θ^{AS} < 1) \)
3. Micro-mismatch: For those do hire workers and those are hired, they are matched by random, not by the positively monotonic manner.

---

22 Remember that the last employer in the complete-information case is the one with \( θ = 1 \). From Lemma 2 we know \( θ^{AS} > 1 \), and this means that the marginal employer has higher capacity and thus fewer employers hiring in the AS equilibrium.
3.4. Incomplete Information with Sorting Mechanism Available

In this subsection we study the case when an effective sorting mechanism is available. The macro-mismatch problem in the adverse selection case results from the positive relationship between ability and reservation wage: any given market wage can only attract those workers with reservation wages (and thus ability levels) lower than it. (Panel A in <Figure 3>) For a sorting mechanism to be able to mitigate the mismatch problem, it has to somehow reverse the relationship, at least to a certain level in a certain range. The sorting function does this by adding one element \( C(s) \), the cost of education, to the original reservation wage \( s \) of working for the modern sector. Hereafter, we can think the new reservation wage, \( s + C(s) \), as the “effective reservation wage”. When the relationship between ability and effective reservation becomes negative, some certain wage levels can attract workers from the top of the ability distribution. (Panel B in <Figure 3>) Contingent on the slope of the cost function, this relationship can be either completely or partially reversed. Here, we focus on the completely-reversed case.\(^{23}\) We will present the result of the partially-reversed case informally in the end of this subsection, and leave the formal analysis of it to the appendix.

![Figure 3](image)

When the sorting mechanism turns the relationship between productivity and (effective) reservation wage into a negative one, it creates a new labor market in the modern sector where employers only hire workers who complete schools. As will be demonstrated later, the new

\(^{23}\) That is, in the text we consider the case that \( \frac{d(s + C(s))}{ds} < 0 \) for all \( s \).
market will never eliminate the original uneducated market in the adverse-selection equilibrium (the AS market). That is, there would always be two employment markets in the modern sector: one requires workers to take education and the other does not. We henceforth call the educated labor market the high market \((H)\) and the uneducated labor market the low market \((L)\).

Borrowing the spirit of \(\textbf{A1}\), we limit our analysis to the scenario that there is only one wage in each of the two markets and everyone behaves as a price taker in the equilibrium.

Let the equilibrium wage and expected ability level of workers in the high market be \(W^H\) and \(E^H\), and those in the low market be \(W^L\) and \(E^L\). \(E^H\) must be greater than \(E^L\), and this implies that \(W^H\) must be greater than \(W^L\). Otherwise no one would ever hire workers from the \(L\) market. In short, \(E^H > E^L\) and \(W^H > W^L\).

Employers have three choices: they can hire workers from the high market, hire workers from the low market, or simply hire no one. In other words, the employer with \(\theta\) chooses the largest one from the set \(\{\theta \cdot E^H - W^H, \theta \cdot E^L - W^L, 0\}\).

To illustrate the employers’ choice, we plot the three choices on a diagram (<Figure 4>). Since \(W^H > W^L\), one can see that the slope of the \((\theta \cdot E^H - W^H)\) line is greater than the slope of the \((\theta \cdot E^L - W^L)\) line (and of course both are greater than the slope of the horizontal line, 0.) If hiring workers from the \(H\) market is the best choice for the employer \(\hat{\theta}\), then hiring workers from the \(H\) market would also be the best choice for all the employers with \(\theta > \hat{\theta}\). If not hiring one is the best choice for the employer \(\tilde{\theta}\), then hiring no one would also be the best choice for all the employers with \(\theta < \tilde{\theta}\). In the case that both the \(H\) market and the \(L\) market exist, we could summarize employers’ hiring behavior as: employers from the top of the capacity distribution hire workers from the \(H\) market, employers from the bottom of the distribution choose to hire no one, and employers in between hire workers from the \(L\) market.\(^{24}\)

\(^{24}\) The conditions for both markets to exist are \(\tilde{\theta} \cdot E^H - W^H > \hat{\theta} \cdot E^L - W^L\) and \(\left(\frac{W^H}{E^H}\right) > \left(\frac{W^L}{E^L}\right)\).

The former ensures that some employers hire workers from the high market, and the latter guarantees some employers hire workers from the low market.
To sum up, in the equilibrium, those employers with $\theta \in [\theta^{ST}, \bar{\theta}]$ hire workers from the $H$ market, those with $\theta \in [\theta^{AS}, \theta^{ST}]$ hire workers from the $L$ market, and those with $\theta \in [\theta, \theta^{AS}]$ hire no one. ($\theta < \theta^{AS} < \theta^{ST} < \bar{\theta}$)

The workers also have three choices: to be (educated and) hired in the $H$ market, to work in the $L$ market, and to remain self-employed. That is, the worker with $s$ chooses the largest one from the set $\{W^H - C(s), W^L, s\}$.

Similarly, we use a simple diagram to study workers’ choices (< Figure 5 >.) Given the assumption we made previously, $\frac{d[s + C(s)]}{ds} < 0$, among these three options, the $[W^H - C(s)]$ curve has the largest slope and the $W^L$ line has the smallest one. From the workers’ side, the key difference between the $H$ market and the $L$ market is that the relationship between slope of the reservation wage curve of the high market is negative while that of the lower market is positive. Therefore, if receiving $W^L$, the wage offered by the low market, is the best choice for $\hat{s}$, then receiving $W^L$ would also be the best choice for the all workers with $s < \hat{s}$. If receiving

---

25 See footnote 23. We assume $\frac{d[s + C(s)]}{ds} < 0$ in this subsection. And the assumption implies $C_s() < -1$. The fact that $C_s() < -1$ then means the slope of the curve $W^H - C(s)$ is everywhere greater than 1, the slope of the line $s$. 

---
\[ W^H - C(s) \], the net reward from being hired in the high market, is the best choice for \( \bar{s} \), then receiving \( W^H - C(s) \) would also be the best choice for all the workers with \( s > \bar{s} \). In short, the workers from the bottom of the ability distribution will be hired by the low market, and the workers from the top of the distribution will be hired by the high market.

In short, the workers with \( s \in [s^{ST}, \bar{s}] \) are hired by the \( H \) market, the workers with \( s \in [\bar{s}, s^{AS}] \) are hired by the \( L \) market, and the workers with \( s \in [s^{AS}, s^{ST}] \) remain self-employed. \( (s < s^{AS} < s^{ST} < \bar{s}) \)

Using the patterns of employers’ and employees’ behaviors we derived above, we can try to approach the equilibrium of this model. The equilibrium can be characterized by four critical values \( (\theta^{AS}, \theta^{ST}, s^{AS}, and s^{ST}) \) that partition workers and employees into different markets and two equilibrium wages in the two labor markets \( (W^H, W^L) \).

Under the above partition, the expected ability of workers hired by the high market would be

\[
E^H = \int_{s^{AS}}^{\bar{s}} xg(x)dx \cdot [\int_{s^{AS}}^{\bar{s}} g(x)dx]^{-1} = \int_{s^{ST}}^{\bar{s}} xg(x)dx \cdot [1 - G(s^{ST})]^{-1}
\]

and that by the low market would be

\[
E^L = \int_{\frac{s^{ST}}{2}}^{s^{AS}} xg(x)dx \cdot [\int_{\frac{s^{ST}}{2}}^{s^{AS}} g(x)dx]^{-1} = \int_{\frac{s^{AS}}{2}}^{s^{ST}} xg(x)dx \cdot [G(s^{AS})]^{-1}.
\]
In the equilibrium, the following equations must hold:

\[ \theta^{s} \cdot E^{L} = W^{L} \]  
(5)
(The marginal employer \( \theta^{s} \) is indifferent between hiring and not hiring.)

\[ \theta^{ST} \cdot E^{L} - W^{L} = \theta^{ST} \cdot E^{H} - W^{H} \]  
(6)
(The marginal employer \( \theta^{ST} \) is indifferent between hiring workers from the \( H \) market and hiring workers from the \( L \) market.)

\[ W^{L} = s^{AS} \]  
(7)
(The marginal worker \( s^{AS} \) is indifferent between being hired from the \( L \) market and remaining self-employed.)

\[ W^{H} - C(s^{ST}) = s^{ST} \]  
(8)
(The marginal worker \( s^{ST} \) is indifferent between being hired from the \( H \) market and remaining self-employed.)

\[ F(\theta^{ST}) - F(\theta^{s}) = G(s^{s}) \]  
(9)
(The measure of labor supplied equals that of labor demanded in the \( L \) market.)

\[ 1 - F(\theta^{ST}) = 1 - G(s^{ST}), \text{ or } F(\theta^{ST}) = G(s^{ST}) \]  
(10)
(The measure of labor supplied equals that of labor demanded in the \( H \) market.)

From (9), \( \theta^{s} \) and \( s^{AS} \) are negatively correlated. A greater \( \theta^{s} \), given \( \theta^{ST} \), means fewer employers and thus, to have the \( L \) marker cleared, there must also be fewer workers hired, which means a smaller \( s^{AS} \). To see it mathematically, we totally differentiate (9) given \( \theta^{ST} \),

\[ -F'(\theta^{s}) \cdot d\theta^{s} = G'(s^{s}) \cdot ds^{s} \], and get

\[ \left[ \frac{d\theta^{s}}{ds^{s}} \right] = \left[ \frac{G'(s^{s})}{-F'(\theta^{s})} \right] = \left[ \frac{g(s^{s})}{-f(\theta^{s})} \right] < 0. \]

From (10), \( \theta^{ST} \) and \( s^{ST} \) are positively correlated. A greater \( \theta^{ST} \) means fewer employers and thus, to have the \( H \) marker cleared, there must also be fewer workers hired, which means a greater \( s^{ST} \). To see this mathematically, we totally differentiate (10),

\[ F'(\theta^{ST}) \cdot d\theta^{ST} = G'(s^{ST}) \cdot ds^{ST} \], and get

\[ \left[ \frac{d\theta^{ST}}{ds^{ST}} \right] = \left[ \frac{G'(s^{ST})}{F'(\theta^{ST})} \right] = \left[ \frac{g(s^{ST})}{f(\theta^{ST})} \right] > 0. \]

By combining (5) and (7) we can get
\[
\theta^{4S} \cdot E^L (= W^L) = s^{4S}, \text{ or }
\]
\[\theta^{4S} = \left[ s^{4S} / E^L \right] \tag{11'} \]
From (11), \( \theta^{4S} \) and \( s^{4S} \) are positively correlated. An employer with greater \( \theta^{4S} \) can afford to offer higher wage and thus attract more workers to the \( L \) market, namely, a greater \( s^{4S} \). To see it mathematically, we totally differentiate (11'),
\[
d\theta^{4S} = \frac{\partial}{\partial s} [s^{4S} / E^L] \cdot ds^{4S}, \text{ and get }
\]
\[
[d\theta^{4S} / ds^{4S}] = \frac{\partial}{\partial s} [s^{4S} / E^L] > 0. \tag{26}
\]
By combining (6) and (8) we can get
\[
\theta^{ST} \cdot (E^H - E^L) + W^L (= W^H) = C(s^{ST}) + s^{ST} \tag{12}
\]
From (12), \( \theta^{ST} \) and \( s^{ST} \) are negatively correlated, given \( W^L \) and \( E^L \). An employer with greater \( \theta^{ST} \) can afford to offer a higher wage and thus attract more employees to the \( H \) market, which means a smaller \( s^{ST} \). To see this mathematically, we totally differentiate (12),
\[
(E^H - E^L) \cdot d\theta^{ST} = [C_s(s^{ST}) + 1 - \theta^{ST} E^H_s] \cdot ds^{ST}, \text{ and get }
\]
\[
[d\theta^{ST} / ds^{ST}] = \frac{C_s(s^{ST}) + 1 - \theta^{ST} E^H_s}{E^H - E^L} < 0. \tag{27}
\]
We can find the equilibrium values of \( \theta^{4S} \), \( s^{4S} \), \( \theta^{ST} \), and \( s^{ST} \) by solving the following four equations (9), (10), (11), and (12) simultaneously. The solution can also be shown in < Figure 6 >. One can show that \( \theta \) is continuous at \( s \) in all of these four equations, and thus we can find a unique solution. \( \tag{28} \)

---

26 See footnote 20.

27 Again, this is because, in our study in the text, we assumed \( d[s + C(s)] / ds < 0 \).

28 For continuous property, again see Appendix (D).
We now want to show that the emergence of the $H$ market cannot eliminate the adverse selection market, namely, the $L$ market.

**Lemma 3:** The sorting mechanism cannot eliminate the low market.

Proof: this proof consists of two parts. The first part shows that the marginal employer of the $H$ market must have capacity greater than 1 ($\theta^{ST} > 1$). And the second part shows that, since $\theta^{ST} > 1$, it is always profitable for at least one employer to hire a worker from the $L$ market.

Part 1: First of all, $\theta^{ST}$ must make non-negative profit from hiring in the $H$ market: $\theta^{ST} \cdot E^H > W^H$. It must be true that $E^H < \bar{\theta}$ and $W^H > \bar{\theta}$ as long as $\bar{\theta}$ is in the high market and $\bar{\theta}$ isn’t the only worker hired. $\Rightarrow E^H < \bar{\theta} < W^H$, and thus $\theta^{ST} > 1$.  

---

29 If $\bar{\theta}$ is not in the high market, then no one will be in the high market and we go back to the AS case. If $\bar{\theta}$ is the sole employee in the high market, then $\theta^{ST}$ would be very close to $\bar{\theta}$ and would be greater than 1.
Part 2: When $\theta^{ST} > 1$, it is always profitable for the employer with $\theta = \theta^{ST} - \varepsilon > 1$ to hire workers from the bottom of the ability distribution. Since $(\theta^{ST} - \varepsilon) \cdot \bar{\xi} > \bar{\xi}$. Q.E.D.

Before we summarize the result of the sorting equilibrium, we would like to find the conditions under which the sorting equilibrium can sustain. Since the $L$ market would always exist, a sorting equilibrium sustains when the $H$ market sustains. And the $H$ market sustains as long as the highest-ability worker earns a higher net income from the $H$ market than remaining self-employed and the highest-ability employer earns a higher profit in the $H$ market than in the $L$ market in the equilibrium. Namely, we need to have $W^{H} - C(\bar{\varepsilon}) > \bar{\varepsilon}$ and $\bar{\theta} \cdot E^{H} - W^{H} > \bar{\theta} \cdot E^{L} - W^{L}$. By intuition, this requires the capacity for the best employer to be high, the educational cost for the highest-ability workers to be low, and the ability gap between workers at the top part and at the bottom part of the ability distribution to be large. From this point on, we only focus on the ST equilibria that are sustainable.

**Result 3**: Under the sorting mechanism, workers participate in three kinds of production activities. High-capacity employers hire high-ability workers in the $H$ market. Employers with lower capacity hire workers from the lowest end of the ability distribution in the $L$ market. Employers with even lower capacity are out of the labor market. Workers with ability in the middle of the distribution remain self-employed.

The net output in the society is

$$\int^{\theta^{ST}}_{0} E^{L} yf(y)dy + \int^{\theta^{ST}}_{0} xg(x)dx + \int^{\bar{\varepsilon}}_{\theta^{ST}} E^{H} yf(y)dy - \int^{\bar{\varepsilon}}_{\theta^{ST}} C(x)g(x)dx$$

$$= \int_{\frac{\theta^{ST}}{2}}^{\bar{\varepsilon}} E(\varepsilon)^{H} xg(x)dx + \int_{\frac{\theta^{ST}}{2}}^{\bar{\varepsilon}} xg(x)dx + \int_{\frac{\theta^{ST}}{2}}^{\bar{\varepsilon}} E(\varepsilon)^{H} xg(x)dx - \int_{\frac{\theta^{ST}}{2}}^{\bar{\varepsilon}} C(x)g(x)dx \tag{13}$$

When the cost function only partially inverse the opportunity cost of working for the $H$ market, the $H$ market may not attract workers with the highest ability. See < Figure 5A >. The reason we consider this case is that this case may produce a result with the highest ability stay self-employed. This variation does not change the equilibrium much, but it does bring some different and interesting results. We hereafter call the equilibrium with highest-ability workers remaining

---

30 Whether equilibrium could sustain and whether equilibrium would happen are two different questions. We would study the question about happening later on.
31 See Appendix (E) for detailed analysis.
self-employed the partially inversed sorting equilibrium (the PST equilibrium), call the equilibrium associated with globally inversed opportunity cost the globally inversed sorting equilibrium (the GST equilibrium), and we save the term “sorting equilibrium” (the ST equilibrium) for the cases that no distinction needs to be made.

4. Some Results and Discussions

In the previous section, we simply characterized the properties of the ST equilibrium. In this section, we want to discuss in more detail some results of the model. We will address questions using intuition here, and leave the formal analysis to the appendix. This section consists of three parts. The first part discusses the change in welfare from the AS equilibrium to the ST equilibrium, for both individuals’ welfare and the welfare of the economy as a whole. Subsection 4.2 focuses on the ST equilibrium and considers how the operation of the sorting function changes as the environments change. The last part concerns the pattern of self-employment that implied by our model.

4.1. Welfare change from the AS equilibrium to the ST equilibrium

The first thing we would like to investigate is how the sorting function affects people’s life. In other words, we would like to compare people’s welfare in the ST equilibrium to that in the AS equilibrium.

---

32 The partially inversed opportunity cost case will not necessarily generate the best-workers-self-employing result. We emphasize on this result because the other one is almost identical to the result of the globally inversed case.
equilibrium. Some of the comparisons are ambiguous, and those results would not be valuable unless we can delineate the conditions under which the ambiguity can be determined. We have troubles doing this precisely, though. To precisely conduct this study, we have to characterize the environment meaningfully by some parameters, solve both the AS equilibrium and the ST equilibrium, construct functions that describe the conditions we are concerned, and then conduct comparative statics to study the impacts of some changes in the environment on the welfare comparisons we are concerned with. Even when we parameterize the environment in the simplest way, the comparative statics are very complicated.33 And, learning from our informal analysis, we believe that we need to characterize the environment in a much detailed manner than the simplest way in order to get meaningful results. Because of the difficulty of conducting formal analysis, in this subsection we would discuss questions informally. We would describe those conditions in terms of the market characteristics 34(such as market sizes, capacity and ability levels of the marginal agents, average ability levels, and wages), and supplement it with our best efforts to link these characteristics back to the environment.

4.1.1. Who gains and who loses from the sorting equilibrium?

Who gains and who loses from the adoption of the sorting mechanism? This question is the key to our following analyses. On both the workers’ side and the employers’ side, we will first describe how the way they participate in the labor markets changes from the AS equilibrium to the ST equilibrium. Generally speaking, even though there are several different scenarios that bring slightly different results, the pattern of the individuals’ welfare change is very comprehensible.

First, we consider the workers’ side. In the AS equilibrium, the labor market (the AS market) hires workers with low ability, and those workers with high ability stay self-employed. Shifting to the ST equilibrium, some (if not all) of those high-ability self-employed workers would become employed in the \( H \) market. In the GST equilibrium, the worker with the highest ability would be in the \( H \) market, while in the PST equilibrium he would remain self-employed. Those low-ability workers originally hired in the AS market would either be hired in the \( L \) market or become self-employed in the GST equilibrium: those with the lowest ability level would still be hired, and those with higher ability among them would become self-employed. In case that both the AS

33 See 4.2. for the comparative-statics study on the ST equilibrium.
34 This is not valid, since those market characteristics are correlated and tend to move together when the environment changes.
market and the $H$ market are sizable, we may see some workers (those with the highest ability in the $AS$ market) move from the $AS$ market to the $H$ market.

How would individual worker’s welfare change? Details are summarized in Table 1. The apparent pattern is that high-ability workers benefit and low-ability workers suffer from shifting to the ST equilibrium. This pattern resembles the established result of the conventional signaling models from the literature. We do, however, want to compare our result to one “odd” result of the literature.

<table>
<thead>
<tr>
<th>Case #</th>
<th>Status in the AS</th>
<th>Status in the ST</th>
<th>Welfare Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Hired in the AS market</td>
<td>Hired in the $L$ market</td>
<td>Lose</td>
</tr>
<tr>
<td>II</td>
<td>Hired in the AS market</td>
<td>Self-employed</td>
<td>Lose</td>
</tr>
<tr>
<td>III</td>
<td>Hired in the AS market</td>
<td>Hired in the $H$ market</td>
<td>Lose/Gain</td>
</tr>
<tr>
<td>IV</td>
<td>Self-employed</td>
<td>Self-employed</td>
<td>No Change</td>
</tr>
<tr>
<td>V</td>
<td>Self-employed</td>
<td>Hired in the $H$ market</td>
<td>Gain</td>
</tr>
</tbody>
</table>

Note: Case I to Case V is ordered from the lowest ability workers to the highest one. Case III and Case IV do not exist at the same time. When the $AS$ marker and $H$ market are both big (and so they overlap), there would be workers shifting directly from the AS market to the H market (Case III). In case that they don’t overlap, there would be workers being self-employed in both equilibrium and thus their welfare will not change (Case IV.)

One of the most extreme results of the economic literature on this issue is that the sorting device may make everyone worse off. In other words, it may happen that the high-ability workers are forced to separate themselves from the low ability workers and it is against their benefit to do so. In our model, when the ST equilibrium sustains, there are always someone benefit from it. In other words, it cannot be the case that the information-revealing mechanism hurts everyone in the economy. Also in our model, in case that there are some workers shifting from the $AS$ market to the $H$ market, the welfare for those workers is always mixed: among them, those with higher ability are better off and those with lower ability are worse off. In other words, in our model when there are some workers suffer because of separating from the low-ability workers, there must also be others benefit from doing so.

35 Also see Appendix (F) for detailed analysis.
36 For example, see Spence (1974) and Mas-colell, et al. (1995).
On the employers’ side, the first point to notice is that there are more employers hiring workers in the ST equilibrium than in the AS equilibrium. Namely, every employer who hires workers in the AS equilibrium would still be hiring in the ST equilibrium. Some of them, those from the top of the capacity distribution, would shift to the $H$ market. And there would be some “new” employers who only hire workers in the ST equilibrium. In case that the $H$ market is smaller than the AS market, all the new employers would hire in the $L$ market. Otherwise, some of the new employers would “jump” directly to the $H$ market.

For the individual employer’s welfare, the result is mixed but the pattern is clear and meaningful. Generally speaking, the “new” employers and those with capacity level in adjacent to the new employers would certainly benefit from the sorting mechanism. Those employers at the margin of the $H$ and $L$ market in the ST equilibrium are most likely to suffer from the sorting mechanism. The intuition is that if we characterize different labor markets by their different wage/average-ability combinations ($W/E$ combinations), then the favorite $W/E$ combination for employers of different capacity tends to be different: high-capacity employers tend to prefer high-wage/high-ability combination and low-capacity ones prefer low/low combination. What the sorting function really does to the employers is to replace a medium-wage/medium-ability market (the $AS$ market) with two markets, a high-wage/high-ability market (the $H$ market) and a low-wage/low-ability market (the $L$ market). The lowest-capacity employer in the $L$ market would be the one benefit most from the creation of the $L$ market, the highest-capacity employer would be the one benefit most from the creation of the $H$ market, and the medium-capacity employers are most likely to suffer because they lose the market that close to their favorite. In short, if the $W/E$ combination in $AS$ market is fairly good, in the sense that wage is not too high and ability not too low, then the high-capacity employers in the $L$ market and the low-capacity employers in the $H$ market (those near the marginal employer) are likely to be worse off. When the term in the $AS$ market is extremely good compared to the term in the $H$ market, it is likely to have the result that every employers in the $H$ market is worse off in the ST equilibrium than in the AS equilibrium.

To summarize, the marginal employer between the $H$ and the $L$ markets in the ST equilibrium gains the least and are the most likely to lose from the sorting mechanism. The new employers

---

37 See Appendix (G) for explanation.
38 Those new employers are out of the labor market in the AS equilibrium, and get chance to earn some profit in the ST equilibrium. The marginal employer in the AS equilibrium earns zero profit before, but in the ST equilibrium he earns a positive profit. (for he must earn more than the new marginal employer)
39 This result brings some interesting point in the analysis below. For more detailed analysis, see Appendix (H).
and those with capacity slightly higher than them definitely benefit from the mechanism. Among employers in the $H$ market, those with the high capacity have higher chance to gain, and gain more if they do gain, than those with low capacity. These results cannot be compared to those of the conventional models, since most of those models assuming homogeneous employers and their profit are set to zero.

4.1.2. Under what situations is the mechanism more likely to be adopted?

To answer this question, first we need to know who makes the decision on whether or not to adopt the mechanism. In this paper, we assume an environment under which people are free to make transactions in the way they prefer. In other words, if there are some workers willing to take education and some employers willing to pay higher wage to hire educated workers, then they could try to have the mechanism work. Therefore, the necessary condition for the adoption of the sorting mechanism is that at least one worker and one employer does better in the ST equilibrium than in the AS equilibrium.\(^{40}\)

For a sustainable sorting equilibrium, from the previous analysis, we know that there are always some high-ability workers who benefit from the $H$ market. However, on the employers’ side, it could be the case that every employer in the $H$ market is worse off than they are in the AS equilibrium. Therefore, it is possible that there are some ST equilibria which are sustainable but not adopted.

The question of “Under what situations is the mechanism more likely to be adopted?” then becomes “Under what situations is it more likely to have employers in the $H$ market earn more than they do in the $AS$ market?” Our answer to this question is that it is more likely for this to happen when workers in the top of the ability distribution are very distinguishable from the workers with medium ability, when the best employers’ capacity is very high, and the cost of education for the marginal worker in the $H$ market is low.\(^{41}\)

---

\(^{40}\) This is not the sufficient condition. It is possible that they trigger the mechanism but fail to reach the ST equilibrium, and the economy heads back to the AS equilibrium. We ignore this problem to avoid doing dynamic analysis.

\(^{41}\) See Appendix (H) for more analysis. As we mention before, we are not able to define the condition precisely because the difficulties of conducting comparative statics.
4.1.3. When would the economy as a whole benefit from the sorting mechanism?

As shown before, many workers and employers change their labor-market positions from the AS equilibrium to the ST equilibrium, thus it seems a messy work to compare the welfare levels of the economy as a whole between these two equilibria. Fortunately, two things make our analysis simple. First, in the ST equilibrium the link between the (gross) output and the educational cost is “loose”, so we can analyze them separately. Secondly, since all workers work in one way or another in both equilibria, we can compare the output levels of them by simply comparing how they make use of the employers’ capacity.

First of all, we cannot be certain whether the sorting mechanism would increase or decrease the net output level of the economy: The gross output in the ST equilibrium is surely higher than that in the AS equilibrium, but the economy has to spend some amount of educational cost to get the extra output. And our model is very flexible on the relative magnitude of the benefit (the extra output) and the cost. To see this point more clearly, we first consider the benefit and the cost separately.

The gross output level in the ST equilibrium is always greater than that in the AS equilibrium because of two reasons. First, there are more jobs used in the ST equilibrium. Secondly, those jobs are occupied by better workers. In other words, we can say that the one-level sorting mechanism mitigates two of the problems caused by the asymmetric information: the macro-mismatch problem and the problem that too-few jobs occupied. To mitigate the third, micro-mismatching, problem, we need to consider a sorting mechanism with more than one level.

The magnitude of this increase in gross output is large when the capacity of employers drops slowly when $\theta$ decreases from the highest level, when the cost of education increases slowly and the ability level decreases slowly when $s$ decreases from the highest level, when the capacity of the new employers are high, when the difference in ability between the workers at the upper margin of the $AS$ market and the $L$ market is large, and so on.  

---

42 This should actually be regarded as a big flaw of this paper.
43 To mitigate the third, micro-mismatching problem, we need to consider a sorting mechanism with more than one level.
44 See Appendix (I) for formal analysis.
On the cost side, the cost of education can be shown by the shaded areas in all three panels in <Figure 5A>. And we can see the problem clearly from the diagram. The output level in the ST equilibrium depends entirely on the way workers and employers participate in the labor markets. And we can see from <Figure 5A> that workers and employers in all three panels participate in the labor market in exactly the same way and thus the output levels in the three panels should also be the same. However, the educational costs are very different in the three panels: the cost in panel (C) is much greater than that in panel (B). Namely, the same output level can be associated with many possible amounts of educational cost. And this is the reason we claimed that the link between the benefit and the cost is loose. This loose link makes our analysis easier but also makes our study less meaningful.

To us, the educational-cost structure in panel (B) is likely to come from an educational system heavily using aptitude tests to determine the admission/grades/graduation of students. The authorities only allow a certain ratio of high-ability people to continue in schools and “kick” others out. Thus it will not cost those able students much to complete schools, and those with low ability face very high cost: a threshold too high to pass. The educational-cost structure in panel (C) is likely to come from an educational system relies on achievement tests to evaluate students’ success. All students are required to learn some material of the same quantity and quality and to pass tests based on their learning from the material. For the same standard of passing/failing the

---

45 One does not have to study for the aptitude tests, and studying hard for the tests will not improve one’s grade much, either.
tests, high-ability workers suffer less than low-ability workers. If the distinguishing power of the tests is low and the return-to-ability in learning given by the material is low, then the most able student would still need to spend a lot of efforts in school. In short, an educational system relies on the aptitude tests tends to have a lower cost of education, and one depends on achievement tests tends to have a higher cost.

We can now combine our study on benefit and cost. In our model, a certain amount of benefit generated by the sorting mechanism could be associated with many different levels of educational cost. Thus, in general we cannot determine whether the sorting function increases or decreases the net output level of the economy. The more meaningful question will then be “under what circumstance would the economy be more likely to benefit from the sorting mechanism?” And the answer is clearly “under the conditions that bring us larger benefit with lower cost.” The condition for the benefit (the extra output) to be large is characterized before. And the educational systems rely more on the aptitude tests tend to have lower cost of education.

4.1.4. Combining individuals’ welfare and economy’s welfare

So far, we have two results from our welfare analysis: the first one is that the sorting mechanism could either increase or decrease the net output of the economy as a whole, and the second one is that the sorting function might either be adopted or not adopted. What is the relationship between these two results?

These two results base on similar but not identical conditions. Thus, any mix of these two results: good/adopted, good/non-adopted, bad/adopted, and bad/non-adopted, is theoretically possible to happen. Generally speaking, since the conditions for the mechanism to be adopted and the conditions for the mechanism to increase net output are derived from similar environment, we believe that the good/adopted and the bad/non-adopted cases are the common situations. And since those two set of conditions are not exactly identical, we cannot rule out the other two cases: the case that the economy as a whole suffers from the sorting mechanism but it is adopted, and the case that the economy as a whole benefits from the sorting mechanism but it is not adopted.

46 If we consider the human-capital-accumulation function of the educational system, the achievement tests would induce more efforts and might be a better way then the aptitude tests.
47 The term “good” means the net output increases and “bad” means the net output decreases.
48 Actually they might be considered as more interesting cases.
The first case is actually the way most people look at the sorting function of education. The situation that this case is likely to happen, to put it intuitively, is that the creation of the $H$ market brings only a small increase in output (and thus the economy as a whole suffers), while at the same time the capacity level of the top employers is extraordinarily high so they can earn higher profit than they do in the $AS$ market. Even though the conditions for this case to happen are not as small as the next one, there is no reason to believe that this would be the most possible case to happen. And thus the conventional view that the sorting function of the educational system can only hurt the economy and that people should try to rid this function from the educational system is not universally valid.

The second case is a surprising result. It says that it is possible to have the situation that the sorting function of the educational system can help the economy but people fail to adopt the function. And it implies that sometimes policy makers can improve the efficiency of the economy through inducing or improving the sorting function. The necessary condition for this case to happen is that no employer in the $H$ market does better than in the $AS$ market. Intuitively, this case is likely to happen when the benefit generated by the sorting function is largely absorbed by the employees and no employer has particularly high capacity. However, as shown in Appendix (H), the chance for these conditions to be satisfied is small and we do not expect this case to happen often.

To summarize, in our model, the interest of individuals and that of the economy as a whole do not have such a major contradiction. Thus we believe the most commonly claimed results that “the sorting function of education hurts the economy as a whole but it is adopted because a minority of people benefit from it” is not the most-likely-to-happen scenario predicted by our model. In other words, there is no reason for us to believe that market failure is the most prevailing case. Our model does allow the possibility of market failure, but it also allows it to go in either direction.

4.1.5. Over-education or under-education?

From the traditional view, the problem of over-education seems to be a natural product of the sorting function. We would like to apply our model to address this problem. Roughly speaking, our model predicts the possibility of both over-education and under-education. One easiest way to see this is to go back to the previous analysis. When the bad/adopted case happens, we have the
over-education problem of under-education; and when the good/non-adopted case happens, we have the under-education problem.

More interesting, and hopefully more convincing, way to illustrate this point is to focus on the good/adopted case. In our model, asking whether we have over-education problem is identical to asking whether the $H$ market is bigger than the optimal size. In the GST equilibrium, one can show that the $H$ market is always greater than the optimal level. In other words, we always have over-education problem in the GST equilibrium. In the case of PST equilibrium, both over-education and under-education are possible.49

Intuitively, this result is just an application of the classical analysis on activities conveying externality: people ignore the impacts of their behavior on other people and thus tend to produce or consume too much or too little compared to the optimal level. Imagine that the ST equilibrium is reached by the creating and expanding of the $H$ market. People move into the $H$ market change both their own welfare and other people’s welfare (the externality). And when the sorting equilibrium is reached, all the private benefits are exhausted and only the externality remains. In the GST equilibrium, the expanding of the $H$ market always has a negative externality on other people: the expansion lower average ability in both the $H$ market and the $L$ market. And thus in the GST equilibrium the $H$ market is always greater than the optimal level and we have the over-education problem. While in the PST equilibrium, the $H$ market expands toward both directions and it is possible to have a positive externality in the equilibrium. And if the externality in the $H$ market is positive and this positive externality outweighs the negative externality in the $L$ market, the $H$ market is smaller than the optimal and the under-education problem appears.

In short, our model predicts that both over-education and under-education problems is possible. For the educational system takes in the highest-ability worker (the GST equilibrium), there is always an over-education problem. For an educational system that “skims” a part of the ability distribution (the PST equilibrium), the under-education problem is likely to happen.

49 See Appendix (J) for formal analysis.
4.2. Comparative Statics Analysis

Without changing the basic results of our model, we can rewrite our model slightly to study how the sorting results change as the environment changes.\(^{50}\)

First, we add a new parameter, \(\alpha\), to the profit function and it becomes

\[
\pi(\theta,s) = \alpha \theta E(s) - W .
\]

The new parameter \(\alpha\) can represent some real-world factors. For the most straightforward explanation, a bigger \(\alpha\) means more production from any pair of \((\theta,s)\). Therefore an increase in \(\alpha\) can be considered as a technological progress in the modern sector (compare to the traditional sector.) Furthermore, even within the modern sector, the technological progress is not neutral. We can see easily that a high \((\theta,s)\) match can benefit more from a high \(\alpha\) than a low \((\theta,s)\) match. Namely, a higher \(\alpha\) does not only make the modern sector more productive compared to the traditional sector, but also, within the modern sector, it raises the value of a good matching.

Another parameter we want to discuss is \(\phi\). A higher \(\phi\) means completing education becomes harder for everyone. In the real world, some phenomena can be considered as relevant measures for this parameter. For example, the ratio of students repeating certain grades, the college admission rates, and the ratio students graduate successfully, and so on. There are many education systems operating jointly to perform the sorting function. And there are many factors having impacts on \(\phi\), such as government’s educational policies, the strategic behaviors of players in the education industry, and the designs of educational institutions. Therefore, our discussions below is so far more like a purely theoretical investigation, and there is a long way to go for us to come up with propositions regarding \(\phi\) that we can test rigorously.

First of all, we consider the impacts of the parameters on the critical values.

Under an insignificant assumption, we derive the following results.\(^{51}\)

\(^{50}\) See Appendix (K) for formal analysis.
\[
\frac{d\theta^sT}{d\alpha} = \theta^sT < 0, \quad \frac{d\theta^aS}{d\alpha} = \theta^aS < 0, \quad \frac{ds^ST}{d\alpha} = s^ST < 0, \quad \frac{ds^AS}{d\alpha} = s^AS < 0, \quad \frac{d\theta^ST}{d\phi} = \theta^ST > 0, \quad \frac{d\theta^aS}{d\phi} = \theta^aS > 0, \quad \frac{ds^ST}{d\phi} = s^ST > 0, \quad \frac{ds^AS}{d\phi} = s^AS > 0
\]

We then translate the results for the critical values into the ones for some of the market characters.

First of all, we consider the size of the labor markets. Let \(\Psi^H\), \(\Psi^L\), and \(\Psi^{SE}\) be the size of the \(H\) market, the \(L\) market, and the self-employed sector respectively.

We get \(\frac{\partial\Psi^H}{\partial\alpha} > 0, \frac{\partial\Psi^L}{\partial\alpha} \leq 0, \frac{\partial\Psi^{SE}}{\partial\alpha} < 0\); and \(\frac{\partial\Psi^H}{\partial\phi} < 0, \frac{\partial\Psi^L}{\partial\phi} > 0, \frac{\partial\Psi^{SE}}{\partial\phi} > 0\).

The predicted impacts of \(\alpha\) on the size of the labor markets are very intuitive. A higher \(\alpha\) does two things: it gives the modern sector an edge over the traditional sector, and it gives the high-capacity/high-ability matching an advantage against the low/low matching. The self-employed sector is only affected by the first edge and the impact is negative, so it would certainly shrinks. The \(H\) market would expand because both forces push it into the same direction. The size of the \(L\) market is undetermined because the two edges operate in the opposite directions: the \(L\) market is a part of the modern market (so its size should increase) but it contains worse matching compared to the \(H\) market (so its size should increase).

The predicted impacts of \(\phi\) on the size of the labor markets can also be interpreted intuitively. A higher \(\phi\) makes it more costly for the \(H\) market to operate and thus the \(H\) market will shrink. And the workers released by the \(H\) market will be shared by the \(L\) market and the self-employed sector.

Secondly, we look at how the equilibrium wages change as \(\alpha\) and \(\phi\) change.

We get \(\frac{\partial W^H}{\partial\alpha} > 0, \frac{\partial W^L}{\partial\alpha} \leq 0\); and \(\frac{\partial W^H}{\partial\phi} < 0, \frac{\partial W^L}{\partial\phi} > 0\).

\[\text{51} \quad \text{The assumption we need is that} \quad (1 - \alpha \cdot \theta^ST \cdot E^L) \quad \text{is not significantly negative. We know with certain that} \quad (1 - \alpha \cdot \theta^aS \cdot E^L) > 0, \quad \text{but we can not determine the sign of} \quad (1 - \alpha \cdot \theta^ST \cdot E^L). \quad \text{The results we present hold as long as the term is not very negative.}\]

41
As $\alpha$ rises, the $H$ market expands. With $\phi$ stays unchanged, the wage in the $H$ market must rise in order to attract workers with higher effective reservation wages. Since the $L$ market can either expand or shrink, wage in the $L$ market (the reservation wage for the marginal employees) can either go up or down.

As $\phi$ rises, the $H$ market shrinks. $W^H$ can either go up or down, depending on the relative size of the increase in $\phi$ and the decrease in the effective reservation wage. And since the $L$ market expands, $W^L$ must be higher than before to attract more employees.

It is interesting to compare our results to some established results derived in literature. There are mainly two contending documents about this kind of results: the wage-competition model and the job-competition model.

The wage competition model predicts $W^H$ to fall when the supply of the $H$ market employees increases. At the same time, it must be true that the supply for the $L$ market decreases and thus $W^L$ rises. In the end, the wage gap ($\Delta W = W^H - W^L$) would decrease. The job competition model predicts that when the supply of $H$ market employees increases, $W^H$ falls. That means that some $L$ market jobs are taken away, and thus $W^L$ will not only fall but also fall by a bigger proportion. In the end, the wage gap will actually rise. Our model predicts that when the supply of $H$ market employees increases (caused by a fall in $\phi$), $W^H$ can either go up or down. The wage in the $L$ market jobs, $W^L$, will fall. And the wage gap may go up or down but will most likely go up.

The empirical results from the literature are mixed. We believe that our model is a better theoretical framework to study this issue in the sense that it gives a more flexible prediction and thus can accommodate diverse results.

4.3. The Self-Employed Workers

Workers of what level ability have a higher tendency to be self-employed? Since self-employment is an important source of entrepreneurs and high-ability entrepreneurs are one of the keys to the economic growth, we believe it is important to know in what situation high-ability
people are more likely to be self-employed. Economics literature has not addressed this issue very successfully.

According to the GST equilibrium, the self-employed workers are the “average” people: their reservation wage is too high to be hired by the low market, and their ability is not high enough to edge into the high market. This seems to be a result that suits the real world well. To be self-employed, one needs to have a certain ability to manage a variety of business. Workers with very low ability are better off being hired by someone else in the $L$ market and making their living simply by taking clear directions and doing some regular, uncomplicated tasks. On the other hand, to participate in the $H$ market, one needs to obtain education to show that his ability level is high enough to satisfy the employers’ demand. And it is too expensive for these average people to do so.

The PST equilibrium provides a different point. According to the equilibrium, there would be another group of self-employed workers: some very talented workers would choose to run their own business. The intuition is that it is not their best interest to be grouped with people with lower ability. And this seems to be a good explanation for the phenomenon that some very successful business men like Bill Gates chose not to finish college education.

Our model offers a theoretical framework to investigate the impact of the sorting function of the educational system on the decision of the highest-ability workers to be self-employed. When the sorting function performs well and the highest-ability workers can separate themselves well from the workers with lower ability, it is more likely that they would take education and join the educated labor market. When the sorting function pools them with a large group of lower ability workers, we expect to see more highest-ability workers choose to be self-employed. We would like to explore this issue further both theoretically and empirically in the future.

5. Conclusion

In this paper, we investigate theoretically the operation of the sorting function performed by the educational systems. The key feature of our model is that we emphasize the importance of a good matching between workers and jobs. A perfect match is in principle not possible because of the informational problems in labor market. Educational systems that can reveal some information about workers’ inner characters are thus able to perform a sorting function and help economies
improve the quality of matching. We believe this is an important issue but it has not received the attention it deserves from economists and policy makers.

We find that the asymmetric information creates two kinds of problems for matching: mismatching occurs between workers and jobs within sector (micro-mismatching) and allocating workers to the “wrong” sectors (macro-mismatching). The one-level sorting device we study in this paper can mitigate the macro-mismatching by reversing the relationship between workers’ ability and their opportunity cost of working for the educated labor market (the $H$ market). Alleviation the micro-mismatching problem is beyond the scope of this paper, and we would have to introduce a multi-level educational system to do that.

The alleviation of the macro-mismatching is not free: the economy has to bear some cost of education. In general, we cannot tell if the sorting function would increase the net output of the economy. Also, the adoption of the sorting function changes individuals’ life. High-ability workers tend to benefit and low-ability workers tend to suffer from using schools for sorting purpose. For employers, the new employers are the most likely to benefit from the sorting mechanism, the highest-capacity employers the second, and the marginal employers between the $H$ and $L$ market are the most likely to suffer from the mechanism.

In some situations the sorting function helps the economy; in other situation it hurts the economy. Sometimes the sorting function would be adopted; sometimes it would not. All four combinations of those two cases are likely to happen. And there is no reason to believe that the bad-but-adopted case would be the most prevailing one. Our model derives the over-education problem from the adoption of the sorting mechanism, but also allows the possibility of under-education.

To conclude, our model declines the conventional viewpoints such as using the educational system as a sorting function certainly hurts the economy, the contradiction between private and public interest in adopting this mechanism is widespread and market failure is prevailing, and using education for sorting definitely brings over-education. We believe more studies on this topic are necessary for us to understand this function and to conduct right policies to deal with it.

There are some interesting researches we plan to explore in the future through extending this model.
First of all, to fully understand the operation of the sorting mechanism, we would have to study a multi-level sorting mechanism. More levels can take workers apart in more detail and thus can alleviate the micro-matching problem. Instead of calling this study an extension of this paper, we think it is more natural to think this study as the second part of this paper.

Secondly, we want to include the human-capital-accumulation function (the learning function) of the educational system into our study. Doing this would not only make our study more realistic, but would also allow us to study the operation of the sorting function from many different angles. The sorting and learning functions of the educational systems are substitutes in some aspects and complements in others. To find the optimal way to operate an educational system, one must consider these two functions and their correlation altogether.

Thirdly, in the real world, the educational systems are operated by many people with different goals and toughness of the education systems is determined by many factors. One particularly interesting angle of studying this issue is to treat the education system as an industry where many education sellers compete for consumers and government acts like a regulator. As a group, it might be optimal for the industry to make schools tough and maintain a high standard for the ability level of the diploma holders. However, it might be beneficial for each individual seller to make his school easier, save some costs, and attract more students. Pending on the level the employers can tell the difference between the qualities brought by diplomas offered by different schools, individual seller’s decision to make school easier would make other producers suffer (negative externality problem). And in the end, it might be the case that all schools choose to make school easier than optimal level (prisoner’s dilemma problem). Government as a regulator can try to mitigate the externality problem and the prisoner’s dilemma problem by imposing regulation on the operations of schools and the educational systems, such as introducing national-wide exams, standardizing curriculum, and subsiding/awarding well-behaved schools, and so on.

For us, this study is more than a game of applying the theory industrial organization to education systems. One the one hand, the conventional research tended to treat the education systems as a black box: money and time go in, knowledge and skills come out. For us, studying education process without considering people’s interests, their motivations, and their interactions is doomed to fail, since these are the most important elements in the studying process. We believe that treating the system as an industry where people behave strategically, could help us open the black box. On the other hand, this study should be able to contribute to the theory of the industrial
organization: studying the unique feature of the educational industry should bring many business strategies that cannot be found from other industries.

Finally, we would like to try and use our model to expand Eicher’s study on the relation between the informational problem in labor markets and issues related to international macroeconomics. Combining Eicher’s model with ours, we can explore the relationship between the performance of the sorting function of the educational system in an economy and the issues that studied in Eicher’s paper, such as economic growth, adoption of new technology, international trade, direct foreign investment, and so on. We believe this study would bring forth an interesting discussion about the relationship between education and economics growth. For example, combining the results of our study and Eicher’s, we would get some propositions like that lowering the admission rate of college would raise the average ability of college graduates, mitigate the informational problem in the modern sector, and make the economy more attractive to FDI and more able to adopt new technology.
Reference


Freeman, R.B. (1976), *The Overeducated American*, Cambridge (Mass.)


Roth, A. and Sotomayor, (1990), *Two sided matching, a Study in Game Theoretic Modeling and Analysis*. Cambridge.


Appendix

Appendix (A): the proof of the positively monotonic matching.

The following points are the results from the subsection 3.2.:

a. Employers and employees are paired in positively monotonic order. Namely, the employer with highest $\theta$ hire the employee with highest $s$, the employer with second highest $\theta$ hire the employee with second highest $s$, and so on.

b. Each employer with $\theta \geq 1$ hires a worker.

c. Employees are hired from the top of the ability distribution.

Proof:

a. Consider two employers, $\theta_H$ and $\theta_L$, and two employees, $s_H$ and $s_L$. We assume that $\theta_H > \theta_L$ and $s_H > s_L$. There are two possible ways to match $\theta$ and $s$: an “in-order” match $[(\theta_H, s_H); (\theta_L, s_L)]$ and an “out-of-order” match $[(\theta_H, s_L); (\theta_L, s_H)]$. Let the profits and wages earned in the in-order match be $\pi_H, W_H, \pi_L, W_L$; and those in the out-of-order match be $\pi'_H, W'_H, \pi'_L, W'_L$.

The following equations must all hold:

\[
\theta_H \cdot s_H = \pi_H + W_H, \\
\theta_L \cdot s_L = \pi_L + W_L, \\
\theta_H \cdot s_L = \pi'_H + W'_L, \\
\theta_L \cdot s_H = \pi'_L + W'_H.
\]

Because $\theta_H > \theta_L$ and $s_H > s_L$, it must be true that $\theta_H \cdot s_H + \theta_L \cdot s_L > \theta_H \cdot s_L + \theta_L \cdot s_H$.

\[
\rightarrow \pi_H + W_H + \pi_L + W_L > \pi'_H + W'_L + \pi'_L + W'_H.
\]

\[
\rightarrow \text{It must be true that either } \pi_H + W_H > \pi'_H + W'_H \text{ or } \pi_L + W_L > \pi'_L + W'_L \text{ (or both)}\]

\[
\rightarrow \text{Under the out-of-order match, there exists at least one pair of buyer and seller can do better by switching trading partners, and thus there is never be any out-of-order match in the equilibrium.}
\]

b. For any $\theta \geq 1$, it is true that $\theta \cdot s \geq s$. In other words, every employer with $\theta \geq 1$ is able to hire an employee, pay him a wage not less than his reservation wage, and make a non-negative profit. Given perfect information (and zero transaction costs), all these of mutual benefits would be realized.

c. Suppose there exist such two employees, $s_1$ and $s_2$. Assume that $s_1 > s_2$ and that $s_2$ is hired but not $s_1$. Then the worker with $s_1$ would always be willing and able to “bid” the job away from $s_2$ by offering a better deal to $s_2$’s employer. Since $s_2$ can at most create profit of $(\theta \cdot s_2 - s_2)$ to any employer, while $s_1$ can offer up to $(\theta \cdot s_1 - s_1)$. Obviously, $(\theta \cdot s_1 - s_1) > (\theta \cdot s_2 - s_2)$.

Q.E.D.

Appendix (B): the justification for Assumption 1 (A1).

The logic that the price-taker assumption is justified in a competitive market can be applied to validate one side of the assumption, but not the other side: it works well on the sellers’ (employees’) side, but not on the buyers’ (employers’) side.

First, consider the sellers’ side. There are many employees selling labor that looks indistinguishable to buyers. Thus no employee can ask for a higher wage and no employee would ask for a lower wage, as long as in equilibrium every worker who wants to sell labor at the wage level can sell his labor.

However, this logic cannot be applied as smoothly to the buyers’ side. Each employer has different capacity to make use of employee’s ability and thus might be willing to offer a wage either higher or lower than the equilibrium one to attract workers of different (expected) ability into their hiring pools. To defend the price-taker assumption on the employers’ side, first of all, we argue that it is impossible for a employer to offer a wage lower than the equilibrium level, since doing so will make it extremely hard for the employer to hire worker. The most difficult part of defending this assumption is to explain why employer
won’t raise wage to increase the expected ability when they prefer a higher wage-expected ability combination than the one brought by the equilibrium.

The explanation we offer is that there is a difference between offering a higher wage alone and offering a higher wage together with all other employers. If all employers offers a higher-than-the-equilibrium wage, $W^+ > W^E$, then the expected ability in each employer’s hiring pool would be $E(W^+)$. If an employer offers $W^+$ alone, due to a fixed cost of entering the labor market, then the expected ability level the employer gets would be much lower than $E(W^+)$. If there is only one employer offering $W^+$, since all employees look identical, the workers with reservation wage higher than $E(W)$ but lower than $W^+$ cannot be sure that they will get the high-pay job. If there is fixed cost of entering the labor market, then only those with relatively low reservation wage, namely, those with lower ability, would join the hiring pool. We assume the fixed cost is high enough, so the benefit of offering $W^+$ is low enough, and thus the employers would have no incentive to offer higher wages.

Appendix (C): the proof of Lemma 1

1. Consider the possibility of no employment. If no employment, then it must be that we cannot find a combination of $(W, E(W))$ that is profitable for even the employer with the highest $\theta$, $\bar{\theta}$, to hire workers.
   No employment $\Rightarrow$ No $(W, E(W))$ such that $\bar{\theta} \cdot E(W) - W \geq 0$
   However, if $W = \underline{s}$, then $E(W) = \underline{s}$ and it must also be true that $\bar{\theta} \cdot E(\underline{s}) - \underline{s} > 0$.
   (Since $\bar{\theta} > 1$ and $\frac{\underline{s}}{E(\underline{s})} = 1$)
   $\Rightarrow$ There at least exists one profitable employment: it is always profitable for $\bar{\theta}$ to hire $\underline{s}$ in our model.
   $\Rightarrow$ There must be some employments.

2. Consider the possibility of full employment. If full employment, it must be that the employer with the lowest $\theta$, $\underline{\theta}$, and the employee with the highest $s$, $\bar{s}$, are both in the market.
   $\Rightarrow$ If $\bar{s}$ is in the market, then all employees are in the market. The expected ability level in the market would be $E(\bar{s})$
   For $\bar{s}$ to be in the market, we need $W \geq \bar{s}$ . (*)
   For $\underline{\theta}$ to be in the market, we need $\underline{\theta} \cdot E(\bar{s}) - W \geq 0$, or $W \leq \underline{\theta} \cdot E(\bar{s})$ . (**)
   $\Rightarrow$ It is not possible to satisfy (*) and (**) at the same time.
   (Since $\underline{\theta} < 1$ and $\bar{s} > E(\underline{s})$ $\Rightarrow$ $\bar{s} > \underline{\theta} \cdot E(\underline{s})$)
   $\Rightarrow$ There must be some workers not employed. Q.E.D.

Appendix (D): the property of continuous

The existence and uniqueness of the solutions in the AS case and the ST case require some functions to be continuous. And this note is to shows that our assumptions in this paper ensure the continuous properties we need.

Two assumptions are made in the paper:
1. the probability density function of workers’ ability, $g(\cdot)$, and the probability density function of employers’ capacity, $f(\cdot)$, to be continuous (in the defined domains)
2. the educational cost function $C(\cdot)$ is continuous at $s$.

Before we demonstrate the continuity, we quote some basic rules about continuity from Simon and Blume (1994):
1. Let $\kappa$ and $\zeta$ be functions from $R^k$ to $R^n$. Suppose $\kappa$ and $\zeta$ are continuous at $x$. Then, $\kappa + \zeta$, $\kappa - \zeta$, and $\kappa \cdot \zeta$ are all continuous at $x$. (Theorem 13.4)

2. If $\zeta \neq 0$, then $\frac{\kappa}{\zeta}$ is also continuous at $x$.

3. Let $\kappa: R^k \rightarrow R^n$ be a continuous function at $x \in R^k$. Let $\zeta: R^m \rightarrow R^n$ be a continuous function at $\kappa(x) \in R^n$. Then, the composition $\kappa \circ \zeta: R^k \rightarrow R^n$ is continuous at $x$. (Theorem 13.7)

For the AS case, we need $\theta$ to be a continuous function of $s$ from the following two equations:

$$1 - F(\theta^{AS}) = G(s^{AS})$$

(1)

$$\theta^{AS} = \frac{s^{AS}}{E(s^{AS})}$$

(4)

To show $\theta$ is a continuous function at $s$ in (1), we can first rewrite the equation as $\theta^{AS} = F^{-1}(1 - G(s^{AS}))$. Since $g(\cdot)$ is continuous at $s$, $G(\cdot)$ and thus $(1 - G(\cdot))$ would be continuous at $s$. Also, $F(\cdot)$ is continuous and monotonically increasing, thus $F^{-1}(\cdot)$ is well defined and continuous. $\Rightarrow$ $\theta$ is a continuous function at $s$ in (1).

We then check if $\theta$ is a continuous function at $s$ in (4). First of all, $E(s) = \int xg(x)dx \cdot \left[G(s)\right]^{-1}$ is obviously continuous, since the integration part has to be continuous at $s$ and $G(s)$ is continuous at $s$ and doesn’t equal to zero (except when $s = \zeta$). And thus the right hand side of (4) is continuous at $s$ as long as $E(s) \neq 0$.

For the ST case, we need $\theta$ to be a continuous function of $s$ in the following four equations:

$$G(s^{ST}) = F(\theta^{ST}) - F(\theta^{AS})$$

(5)

$$G(s^{ST}) = F(\theta^{ST})$$

(6)

$$\theta^{AS} \cdot E_L = s^{AS}$$

(7)

$$\theta^{ST} \cdot (E_H - E_L) + s^{AS} = C(s^{ST}) + s^{ST}$$

(8)

One can show that $\theta$‘s in equations (6) and (7) are continuous at $s$ by the same method that shows continuity in (1) and (4).

For (5), we need $\theta^{AS}$ to be continuous at $s$. For a given $\theta^{ST}$, we can using similar method show that it is true. For (8), we can first rewrite it as $\theta^{ST} = \frac{[C(s^{ST}) + s^{ST} - s^{AS}]}{(E_H - E_L)}$. Every single term in the right hand side is continuous at $s$, and in all cases we study the denominator would not be zero. Applying the above mentioned rules, $\theta^{ST}$ is clearly continuous at $s$.

**Appendix (E):** the PST equilibrium

a). The model

Here we consider the sorting model in which the sorting mechanism doesn’t entirely reverse the relationship between ability and reservation wage. In other words, the effective opportunity cost curve has a hook like panel (B) and panel (C) in < Figure 3A >.
A hook won’t necessarily bring a different result from the model without it, and we would need to analyze the model in a different manner only when the hook lets some best workers quit the high market, like that case represented by panel (C).

Some best workers would choose not to join the high market (too cool for schools) when $W^H - C(s) < \bar{y}$, like the situation shown by <Figure 5A>.

When situation like this occurs, the high market would skim workers from the middle of the ability distribution, rather than attract workers from the top of the ability distribution. The employers’ problems, however, stay changed.

The equations that characterize the equilibrium will become

(A1) \[ \theta^{AS} \cdot E^L = W^L \]

(The marginal employer $\theta^{AS}$ is indifferent between hiring and not hiring workers. It also has to be true that $\theta \cdot E^L > W^L$ for all $\theta > \theta^{AS}$ and $\theta \cdot E^L < W^L$ for all $\theta < \theta^{AS}$.)

(A2) \[ \theta^{ST} \cdot E^L - W^L = \theta^{ST} \cdot E^H - W^H \]

(The marginal employer $\theta^{ST}$ is indifferent between hiring workers from the high market and hiring workers from the low market. It also has to be true that $\theta^{ST} \cdot E^L - W^L < \theta^{ST} \cdot E^H - W^H$ for all $\theta > \theta^{ST}$ and $\theta^{ST} \cdot E^L - W^L > \theta^{ST} \cdot E^H - W^H$ for all $\theta < \theta^{ST}$.)
(A3) \[ W^L = s^{45} \]
(The marginal worker \( s^{45} \) is indifferent between being hired by the low market and remaining self-employed. It also has to be true that \( W^L > s \) for all \( s < s^{45} \) and \( W^L < s \) for all \( s > s^{45} \).

(A4) \[ W^H = C(s^{ST}) + s^{ST} = C(s^{ST}) + \bar{s}^{ST} \]
(The marginal workers \( s^{ST} \) and \( \bar{s}^{ST} \) are both indifferent between being hired by the high market and remaining self-employed. We also need to have \( W^H - C(s) \geq s \) for all \( s^{ST} \leq s \leq \bar{s}^{ST} \) and \( W^H - C(s) < s \) for all \( s < s^{ST} \) and \( s > \bar{s}^{ST} \). Namely, we need to have \( |C'(s^{ST})| \geq 1 \) and \( |C'(\bar{s}^{ST})| < 1 \). The condition must be satisfied for a hook-equilibrium.)

(A5) \[ G(s^{45}) = F(\theta^{45}) - F(\theta^{ST}) \]
(The measure of labor supplied equals the measure of labor demanded in the low market.)

(A6) \[ G(s^{ST}) - G(s^{45}) = 1 - F(\theta^{ST}) \]
(The measure of labor supplied equals the measure of labor demanded in the high market.)

In this model, the \( E^L \) function remains the same and equals \( \int_{s^{45}}^{1} xg(x)dx \cdot [G(s^{45})]^{-1} \); but the \( E^H \) function becomes \( \int_{s^{ST}}^{\bar{s}^{ST}} xg(x)dx \cdot [G(s^{ST}) - G(s^{45})]^{-1} \).

By combining (A1) and (A3), we can get
\[(A7) \quad \theta^{45} \cdot E^L (= W^L) = s^{45} \]

By combining (A2) and (A4), we can get
\[(A8) \quad \theta^{ST} \cdot (E^H - E^L) + s^{45} (= W^H) = C(s^{ST}) + \bar{s}^{ST} \]
and \[(A8') \quad \theta^{ST} \cdot (E^H - E^L) + s^{45} (= W^H) = C(s^{ST}) + \bar{s}^{ST} \]

The equilibrium can be described by 5 equations: (A5), (A6), (A7), (A8), and (A8'). The characters of (A5) to (A7) are identical to those of (6) to (8), but (A8) and (A8') are quite different from (9). In the sorting equilibrium, \( \theta^{ST} \) and \( s^{ST} \) are negatively correlated (given \( \bar{s}^{ST} \) and \( s^{45} \)); an employer with greater \( \theta^{ST} \) can afford to offer higher wage and higher wage attracts more employees to high market, which means a smaller \( s^{ST} \). In the PST model, there are two sets of relationship between \( \theta^{ST} \) and \( s^{ST} \): \( \theta^{ST} \) and \( s^{ST} \) \((A8)) \) guides the lower bound of the K market; \( \theta^{ST} \) and \( \bar{s}^{ST} \) \((A8')) \) directs the upper bound. The curve represents the relationship between \( \theta^{ST} \) and \( s^{ST} \) still has a negative slope, but the slope for the one for \( \theta^{ST} \) and \( s^{ST} \) is not determined.

To see it mathematically, we total differentiate (A8) and (A8'),
\[ (E^H - E^L) \cdot d\theta^{ST} = \left[ C'(s^{ST}) + 1 - \theta^{ST} E^H \right] \cdot ds^{ST} \]
and get\[ \frac{d\theta^{ST}}{ds^{ST}} = \frac{\left[ C'(s^{ST}) + 1 - \theta^{ST} E^H \right]}{E^H - E^L} \]

For \( s^{ST} \), the slope is negative since we still have \( |C'(s^{ST})| > 1 \). However, for \( \bar{s}^{ST} \), since \( |C'(\bar{s}^{ST})| < 1 \), we cannot decide the sign of the slope.

In the ST model, \( \bar{s} \) is the most preferred worker to hire in the \( H \) market. Every employer would be happy to hire him, and employers with higher technology level could afford to include workers with less ability into the hiring pool, namely, to extend the \( H \) market further down. And the equilibrium is reached when the marginal worker \( (s^{ST}) \) happens to be the last one on the hiring pool of the marginal employer \( (\theta^{ST}) \). In this
case, $\tau$ is no longer the favorite worker to be hired in the $H$ market. Someone in the middle of the ability distribution would be the most preferred (the one that all employers want (afford) to include in their hiring pool), and employers with higher technology level could afford to include more workers from both sides of the most preferred into the hiring pool. Namely, the hiring pool gets bigger as employers’ capacity gets higher, and the pool is enlarged both further up and further down the ability distribution from a certain level. The equilibrium would now be reached when the marginal worker in the upper bound ($s^{ST}$) happens to be the highest one on the hiring pool of the best employer ($\theta$) and the marginal worker in the upper bound ($s^{ST}$) happens to be the lowest one on the hiring pool of the marginal employer ($\theta^{ST}$).

b). The implications of the PST model

Regarding the implications of the models, the main difference is that the externalities caused by individuals’ decisions to join the $H$ market are exclusively negative in the sorting equilibrium but is not so in the PST model. In the PST model, the $H$ market expands toward both directions. Those join from the bottom, like in the sorting equilibrium, would lower the average productivity of the $H$ market (negative externality). However, those join for the top would actually pull up the average ability of workers in the $H$ market (positive externality). We can not rule out the possibility that, in certain equilibrium, this extra positive externality outweighs the existing negative ones.

This difference doesn’t change the welfare analysis much. Nevertheless, the PST model has a quite different implication regarding over-education: the PST model does not necessarily generate an over-education result. Because of the positive externality and the possibility, even though could be a very small possibility, that it dominates the negative externality, a PST model can generate an equilibrium with under-education.

The results of the comparative static are also very similar. For the impacts of the changes in parameters on the sizes of the labor markets, we get the same results:
\[ \frac{\partial \Psi^H}{\partial \alpha} > 0, \frac{\partial \Psi^L}{\partial \alpha} < 0, \frac{\partial \Psi^{SE}}{\partial \alpha} < 0 \] , and \[ \frac{\partial \Psi^H}{\partial \phi} < 0, \frac{\partial \Psi^L}{\partial \phi} > 0, \frac{\partial \Psi^{SE}}{\partial \phi} > 0 \]

And, for the equilibrium wages, we get the same results:

\[ \frac{\partial W^H}{\partial \alpha} > 0 \text{ and } \frac{\partial W^L}{\partial \alpha} < 0, \frac{\partial W^{SE}}{\partial \alpha} < 0 \] \text{ and } \frac{\partial W^H}{\partial \phi} > 0 \text{ and } \frac{\partial W^L}{\partial \phi} > 0 .

The intuitive explanations for these results for the sorting equilibrium can also be validly applied to this model.

The most different, and also the most interesting, result brought by the PST model is the implication for the self-employed sector. The PST model predicts that there would be some workers whose self-employed income is too high for them to be employed. In our setup, they would still contribute more to the society if they are hired by any employer with \( \theta > 1 \). The reason that it is not beneficial for them to be hired is because that by being employed their productivity, and thus their wage, would be pooled with other low ability workers.

**Appendix (F): workers’ welfare change**

Consider < Figure 5B >, which is a duplicate of < Figure 5 > with the possible allocation of workers in the AS equilibrium added in the bottom.

Panel (A) of < Figure 5B > illustrates the case that the AS market and the \( H \) market are relatively small. In this case, all workers who are hired in the \( H \) market in the ST equilibrium are self-employed in the AS equilibrium, and we have some workers be self-employed in both equilibria. Panel (B) of < Figure 5B > illustrates the case that the AS market and the \( H \) market are relatively large and thus they overlap. In this case, some workers shift directly from the AS market to the \( H \) market.

The importance of making this distinction is can be seen in analyzing the welfare change. The thick/grey line in both panels shows workers’ income in the AS equilibrium, and the thick/black line in both panel shows workers’ income in the ST equilibrium. In the case of panel (A), the sorting mechanism makes all those work in the \( H \) market better off and make all those works in the AS market worse off. In the case of
panel (B), the result from panel (A) holds except for those “overlapping” workers: The high-ability part of the overlapping workers benefits from the mechanism, while the low-ability part of the overlapping workers suffers from the mechanism.

Appendix (G):

Steps to show that there are more employers hiring workers in the ST equilibrium than in the AS equilibrium:

1. For any given market situation (wage/ability combination), high capacity employers can make more profit than those with lower capacity.
2. The hiring employers are those from the top of the capacity distribution.
3. To see which equilibrium has more employers, we simply have to compare the capacity level of the marginal employer, the zero-profit employer, of the uneducated labor markets in the two equilibria.
4. The $L$ market in the ST equilibrium is almost identical to the labor market in the AS equilibrium, except that the $L$ market “loses” some high-capacity employers to the $H$ market.
5. Thus, we can simply use < Figure 2A > to illustrate it.
6. The left hand side of < Figure 2A > is simply a duplicate of < Figure 2 >. The right hand side of the figure shows the operation of the $L$ market in the ST equilibrium.
7. In the $L$ market, with top employers taken away, the market size will be smaller than the AS market. In both market, workers are hired from the bottom of the ability distribution, and thus a smaller market means the marginal worker has a lower level of ability.
8. According to our assumption, the wage/ability ratio gets higher as the market expands. With both marginal employers earning zero profit, the one in the smaller market has a smaller capacity.
9. This shows that the ST equilibrium uses more employers.

< Figure 2A >

Appendix (H): the employers’ welfare change

This part considers how employers’ market-participations and their profits change from the AS equilibrium to the ST equilibrium. All of the panels in < Figure 4A > are duplicates of < Figure 4 >. The two lines in < Figure 4 > show different levels of profit that employers’ can earn from hiring workers in either market. Here, we simply add an additional line, the thick/grey line, representing their profit from the AS market to check their welfare change.
The relationship among wages and ability levels in different market that we are sure is as the follows.

1. $W^L < W^{AS} < W^H$,
2. $E^L < E^{AS} < E^H$,
3. $W^L / E^L < W^H / E^H$, and
4. $W^L / E^L < W^{AS} / E^{AS}$

Thus, the only rules that guide us draw the grey line is to start at a wage level between $W^L$ and $W^H$, draw a line with a slope between $E^L$ and $E^H$, and cross the horizontal line on the right hand side of $W^L / E^L$. We thus have three possibilities: in panel (A) every employers hiring in the ST equilibrium are better off, in panel (B) some employers (the high-capacity ones in the $H$ market and the low-capacity ones in the $L$ market) are better off and others (those around the marginal employer) worse off, and in panel (C) every employer in the $H$ market is worse off.

The situation in panel (C) is an interesting point and it brings some important results later. Thus we would like to consider the condition for this to happen in more detail. The condition is $\bar{\sigma} \cdot E^H - W^H < \bar{\sigma} \cdot E^{AS} - W^{AS}$. The situation is more likely to happen when

1. workers in the top of the ability distribution are very distinguishable from the workers with medium ability
2. the best employers’ capacity is very high
3. the cost of education for the marginal worker in the $L$ market is low

**Appendix (I):** the welfare change of the economy as a whole

To have a bigger net benefit in aggregate

We can use the following diagram to compare the gross benefit.
The benefit is big when:
1. \( H \) market is bigger than AS market
2. \( E^H - E^{AS} \) is big
3. \( E^H - E^L \) is small (condition 2 is stronger) (also implies small gap between \( E^{AS} \) and \( E^L \))
4. The capacity of new employers is big
5. \( L \) market is big (small gap between old and new employers)

In terms of the environmental factors, the conditions for the benefit to be large are:
1. the capacity of employers drops slowly when \( \theta \) decreases from the highest level
2. the cost of education increases slowly when \( s \) decreases from the highest level
3. the ability level decreases slowly when \( s \) decreases from the highest level
4. the capacity of the new employers are high
5. the difference in ability between the workers at the upper margin of the AS market and the L market is large

**Appendix (J): Over-education or Under-education?**

The net output for the economy in the GST equilibrium is
\[
\Omega = \int_{\theta^{ST}}^{\theta^{AS}} E^L yf(y) dy + \int_{s^{ST}}^{s^{AS}} xg(x) dx + \int_{\theta^{ST}}^{\theta^{AS}} E^H yf(y) dy - \int_{s^{ST}}^{s^{AS}} C(x)g(x) dx.
\]

The optimal size of the \( H \) market is the one that maximizes the net output of the economy. Namely, we should have \( \frac{d\Omega}{d\theta^{ST}} = 0 \). To see how the size of the \( H \) market in the GST equilibrium is related to this optimal level, we measure the value of \( \frac{d\Omega}{d\theta^{ST}} \) around the equilibrium point. If \( \frac{d\Omega}{d\theta^{ST}} < 0 \) around the equilibrium, then the equilibrium \( H \) market is smaller than the optimal. And if \( \frac{d\Omega}{d\theta^{ST}} > 0 \) around the equilibrium, then the equilibrium \( H \) market is greater than the optimal.

In the ST equilibrium, the value of \( \theta^{ST} \) is related to all other critical values. If \( \theta^{ST} \) changes, first of all, \( s^{ST} \) would also have to change in order to clear the \( H \) market. Also, a change in \( \theta^{ST} \) means there are some employers moving into or out of the \( L \) market, and thus \( \theta^{AS} \) and \( s^{AS} \) would also be changed to maintain
the balance in the L market. To proceed, we write the other three critical values as a function of \( \theta^{ST} \), differentiate \( \Omega \) with respect to \( \theta^{ST} \), and then bring in the equilibrium conditions to evaluate the first order condition around the equilibrium.

We differentiate \( \Omega = \int_{\theta^{ST}} E^L \theta^{ST} dy + \int_{\theta^{ST}} xg(x)dx + \int_{\theta^{ST}} E^H \theta^{ST} dy - \int_{\theta^{ST}} C(x)g(x)dx \) with respect to \( \theta^{ST} \), and get

\[
\left[ \frac{d\Omega}{d\theta^{ST}} \right] = E^L \theta^{ST} f(\theta^{ST}) - E^L \theta^{ST} f(\theta^{ST}) \left( \frac{\partial \theta^{ST}}{\partial \theta^{ST}} \right)
+ \int_{\theta^{ST}} E^L \left( \frac{\partial \theta^{ST}}{\partial \theta^{ST}} \right) yf(y)dy
+ s^{ST} g(s^{ST}) \left( \frac{\partial \theta^{ST}}{\partial \theta^{ST}} \right) - s^{ST} g(s^{ST}) \left( \frac{\partial \theta^{ST}}{\partial \theta^{ST}} \right)
- E^H (\theta) s^{ST} g(s^{ST}) + \int_{\theta^{ST}} E^H \left( \frac{\partial \theta^{ST}}{\partial \theta^{ST}} \right) yf(y)dy + C(s^{ST}) g(s^{ST}) \left( \frac{\partial \theta^{ST}}{\partial \theta^{ST}} \right)

To have both markets cleared, both (6) and (7) must hold all the time.

We totally differentiate (6) \( F(\theta^{ST}) - F(\theta^{ST}) = G(s^{ST}) \), and get

\[
(6') \quad f(\theta^{ST}) \cdot d\theta^{ST} - f(\theta^{ST}) \cdot d\theta^{ST} = g(s^{ST}) \cdot ds^{ST}
\]

We then totally differentiate (7) \( F(\theta^{ST}) = G(s^{ST}) \), and get

\[
(7') \quad f(\theta^{ST}) \cdot d\theta^{ST} = g(s^{ST}) \cdot ds^{ST}, \quad f(\theta^{ST}) = g(s^{ST}) \cdot ds^{ST} / d\theta^{ST}
\]

(6') and (7') mean that the change in quantity demanded in both markets must equal the change in quantity supplied. We can combine (6”) and (7”)

\[
(6”) \quad f(\theta^{ST}) \cdot d\theta^{ST} = g(s^{ST}) \cdot ds^{ST}
\]

\[
(7”) \quad f(\theta^{ST}) = g(s^{ST}) \cdot ds^{ST} / d\theta^{ST}
\]

\[
(6””) \quad f(\theta^{ST}) = g(s^{ST}) \cdot ds^{ST} / d\theta^{ST} = f(\theta^{ST}) \cdot d\theta^{ST} + g(s^{ST}) \cdot ds^{ST}
\]

Or \( f(\theta^{ST}) = g(s^{ST}) \cdot \left( \frac{ds^{ST}}{d\theta^{ST}} \right) = f(\theta^{ST}) \cdot \left( \frac{d\theta^{ST}}{d\theta^{ST}} \right) + g(s^{ST}) \cdot \left( \frac{ds^{ST}}{d\theta^{ST}} \right) \)

(6””) has an intuitive explanation: the change in quantity of labor supplied in high market must equal the summation of the change in quantity of labor supplied in low market and the change in quantity of labor remains in self-employed.

For the L market to be in an equilibrium, (8) \( \theta^{AS} \cdot E^L = s^{AS} \) must also hold.

With (6””) and (8), we can rewrite \( \left[ \frac{d\Omega}{d\theta^{ST}} \right] \) as
\[
\int_0^\theta E^* (\psi_{\theta^*}) f(y) dy + \int_0^\pi E^H (\psi_{\theta^H}) f(y) dy + E^* f(\theta^*) - E^H f(\theta^H) \left( \frac{d\theta^H}{d\theta^*} \right) + s^{ST} g(s^{ST}) \left( \frac{ds^{ST}}{d\theta^*} \right) - s^{AS} g(s^{AS}) \left( \frac{ds^{AS}}{d\theta^*} \right) = \int_0^\theta E^* (\psi_{\theta^*}) f(y) dy + \int_0^\pi E^H (\psi_{\theta^H}) f(y) dy + E^* f(\theta^*) - E^H f(\theta^H) \left( \frac{d\theta^H}{d\theta^*} \right) + s^{ST} g(s^{ST}) \left( \frac{ds^{ST}}{d\theta^*} \right) + C(s^{ST}) g(s^{ST}) \left( \frac{ds^{ST}}{d\theta^*} \right)
\]

The first two terms are positive and can be thought as externalities from shrinking the \( H \) market. The last term is the “private” benefit and cost for those people involved. More precisely, consider the term within the bracket of the last term, \( E^* f(\theta^*) \left( \frac{d\theta^H}{d\theta^*} \right) \). It is easier to interpret the term by rewriting it as \( E^* f(\theta^*) \left( \frac{d\theta^H}{d\theta^*} \right) = \left[ E^* f(\theta^*) - E^H f(\theta^H) \right] \left( \frac{d\theta^H}{d\theta^*} \right) \). This is actually the condition that ensures a sorting equilibrium: the marginal employer \( (\theta^*) \) earns the same profit from both markets, and the marginal employee of the \( L \) market and the marginal employee in the \( H \) market are both indifferent between being employed and being self-employed. Therefore, the last term must be zero in the \( ST \) equilibrium. Since the first two terms (the externalities) are positive, the first order condition is positive around the \( ST \) equilibrium. Namely, For a GST equilibrium we always have \( d\Omega / d\theta^* > 0 \) around the equilibrium.

**Appendix (K):** the comparative statics

We rewrite (8) and (9) as (8’) and (9’), and then put it together with the other three equations that guide the sorting equilibrium.

(6) \( F(\theta^*) - F(\theta^H) = G(s^{AS}) \)

(7) \( F(\theta^H) = G(s^{ST}) \)

(8’) \( \alpha \cdot \theta^{AS} \cdot L = s^{AS} \)

(9’) \( \alpha \cdot \theta^{ST} \cdot (L - \theta^{ST}) + s^{AS} = C(s^{ST}, \phi) + s^{ST} \)

a). Comparative statics for the cutting values
\[
\begin{bmatrix}
    f(\theta^{st}) & -f(\theta^{as}) & 0 & -g(s^{as}) \\
    f(\theta^{as}) & 0 & -g(s^{st}) & 0 \\
    0 & \alpha \cdot E^s & 0 & \alpha \cdot \theta^{as} \cdot E^s -1 \\
    \alpha \cdot (E^s - E^l) & 0 & \alpha \cdot \theta^{as} \cdot E^s - C_i(s^{st}) - 1 & 1 - \alpha \cdot \theta^{st} \cdot E^s \\
\end{bmatrix}
\begin{bmatrix}
    d\theta^{st} \\
    d\theta^{as} \\
    d\theta^{s} \\
    d\phi \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
    0 & -f(\theta^{as}) & 0 & -g(s^{as}) \\
    0 & 0 & -g(s^{st}) & 0 \\
    -\theta^{as} \cdot E^s & \alpha \cdot E^s & 0 & \alpha \cdot \theta^{as} \cdot E^s -1 \\
    -\theta^{st} \cdot (E^s - E^l) & 0 & \alpha \cdot \theta^{st} \cdot E^s - C_i(s^{st}) - 1 & 1 - \alpha \cdot \theta^{st} \cdot E^s \\
\end{bmatrix}
\begin{bmatrix}
    d\theta^{st} \\
    d\theta^{as} \\
    d\theta^{s} \\
    d\phi \\
\end{bmatrix}
\]

The signs of the denominator matrix can be shown as

\[
\begin{bmatrix}
    (+) & (-) & 0 & (-) \\
    (+) & 0 & (-) & 0 \\
    0 & (+) & 0 & (-) \\
    (+) & 0 & (+) & (?) \\
\end{bmatrix}
\]

Given the assumption that the undetermined term, \((1 - \alpha \cdot \theta^{st} \cdot E^s)\), is not significantly negative, the sign of the denominator can decided:

\[
\begin{align*}
    &\det
    \begin{bmatrix}
        f(\theta^{st}) & -f(\theta^{as}) & 0 & -g(s^{as}) \\
        f(\theta^{as}) & 0 & -g(s^{st}) & 0 \\
        0 & \alpha \cdot E^s & 0 & \alpha \cdot \theta^{as} \cdot E^s -1 \\
        \alpha \cdot (E^s - E^l) & 0 & \alpha \cdot \theta^{st} \cdot E^s - C_i(s^{st}) - 1 & 1 - \alpha \cdot \theta^{st} \cdot E^s \\
    \end{bmatrix} \\
    &= f(\theta^{as}) \cdot \det
    \begin{bmatrix}
        f(\theta^{st}) & -g(s^{st}) & 0 \\
        f(\theta^{as}) & 0 & -g(s^{as}) \\
        0 & \alpha \cdot E^s & \alpha \cdot \theta^{as} \cdot E^s -1 \\
        \alpha \cdot (E^s - E^l) & \alpha \cdot \theta^{st} \cdot E^s - C_i(s^{st}) - 1 & 1 - \alpha \cdot \theta^{st} \cdot E^s \\
    \end{bmatrix} \\
    &\quad -\alpha \cdot E^s \cdot \det
    \begin{bmatrix}
        f(\theta^{st}) & 0 & -g(s^{as}) \\
        f(\theta^{as}) & -g(s^{st}) & 0 \\
        0 & \alpha \cdot E^s & \alpha \cdot \theta^{as} \cdot E^s -1 \\
        \alpha \cdot (E^s - E^l) & \alpha \cdot \theta^{st} \cdot E^s - C_i(s^{st}) - 1 & 1 - \alpha \cdot \theta^{st} \cdot E^s \\
    \end{bmatrix} \\
    &= f(\theta^{as}) \cdot \left[ -g(s^{st}) \cdot (\alpha \cdot \theta^{as} \cdot E^s -1) - \alpha \cdot (E^s - E^l) \right] - f(\theta^{st}) \cdot (\alpha \cdot \theta^{as} \cdot E^s -1) \cdot (\alpha \cdot \theta^{st} \cdot E^s - C_i(s^{st}) - 1) \\
    &\quad -\alpha \cdot E^s \cdot \left[ f(\theta^{as}) \cdot -g(s^{st}) - f(\theta^{st}) \cdot (\alpha \cdot \theta^{as} \cdot E^s -1) \cdot -g(s^{as}) \cdot \alpha \cdot (E^s - E^l) \right] \\
    &> 0
\end{align*}
\]

We keep the assumption that \((1 - \alpha \cdot \theta^{st} \cdot E^s)\) is not significantly negatively throughout this comparative statics.
The numerator:

\[
\begin{vmatrix}
0 & -f(\theta^{ST}) & 0 & -g(s^{ST}) \\
0 & 0 & -g(s^{ST}) & 0 \\
-\theta^{AS} \cdot E^L & \alpha \cdot E^L & 0 & \alpha \cdot \theta^{AS} \cdot E^L - 1 \\
-\theta^{ST} \cdot (E^{H} - E^L) & 0 & \alpha \cdot \theta^{ST} \cdot E^{H} - C_i(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E^L
\end{vmatrix}
\]

\[= f(\theta^{ST}) \cdot \det
\begin{vmatrix}
0 & -g(s^{ST}) & 0 \\
0 & -\theta^{AS} \cdot E^L & 0 & \alpha \cdot \theta^{AS} \cdot E^L - 1 \\
0 & -\theta^{ST} \cdot (E^{H} - E^L) & \alpha \cdot \theta^{ST} \cdot E^{H} - C_i(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E^L \\
0 & -g(s^{ST}) & 0 & \alpha \cdot \theta^{AS} \cdot E^L - 1
\end{vmatrix}
\]

\[-\alpha \cdot E^L \cdot \det
\begin{vmatrix}
0 & 0 & -g(s^{ST}) \\
0 & 0 & -g(s^{ST}) \\
-\theta^{ST} \cdot (E^{H} - E^L) & \alpha \cdot \theta^{ST} \cdot E^{H} - C_i(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E^L \\
0 & -g(s^{ST}) & 0 & \alpha \cdot \theta^{AS} \cdot E^L - 1
\end{vmatrix}
\]

\[= f(\theta^{ST}) \cdot \left\{ g(s^{ST}) \cdot (\alpha \cdot \theta^{AS} \cdot E^L - 1) \cdot [\theta^{ST} \cdot (E^{H} - E^L)] - g(s^{ST}) \cdot \theta^{AS} \cdot E^L \cdot [1 - \alpha \cdot \theta^{ST} \cdot E^L] \right\}
\]

\[= \alpha \cdot E^L \cdot \left\{ [g(s^{ST}) \cdot g(s^{ST})] \cdot \theta^{ST} \cdot (E^{H} - E^L) \right\}
\]

\[< 0
\]

\[\Rightarrow \frac{d\theta^{ST}}{d\alpha} < 0.
\]

\[
\begin{vmatrix}
f(\theta^{ST}) & 0 & 0 & -g(s^{ST}) \\
f(\theta^{ST}) & 0 & -g(s^{ST}) & 0 \\
\alpha \cdot (E^{H} - E^L) & -\theta^{ST} \cdot (E^{H} - E^L) & \alpha \cdot \theta^{ST} \cdot E^{H} - C_i(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E^L \\
\end{vmatrix}
\]

\[
\begin{vmatrix}
f(\theta^{ST}) & -f(\theta^{ST}) & 0 & -g(s^{ST}) \\
f(\theta^{ST}) & 0 & -g(s^{ST}) & 0 \\
0 & \alpha \cdot E^L & 0 & \alpha \cdot \theta^{AS} \cdot E^L - 1 \\
\alpha \cdot (E^{H} - E^L) & -\theta^{ST} \cdot (E^{H} - E^L) & \alpha \cdot \theta^{ST} \cdot E^{H} - C_i(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E^L
\end{vmatrix}
\]

\[= g(s^{ST}) \cdot \det
\begin{vmatrix}
f(\theta^{ST}) & 0 & -g(s^{ST}) \\
f(\theta^{ST}) & 0 & 0 \\
0 & -\theta^{AS} \cdot E^L & \alpha \cdot \theta^{AS} \cdot E^L - 1 \\
\alpha \cdot (E^{H} - E^L) & -\theta^{ST} \cdot (E^{H} - E^L) & \alpha \cdot \theta^{ST} \cdot E^{H} - C_i(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E^L
\end{vmatrix}
\]

\[= [\alpha \cdot \theta^{ST} \cdot E^{H} - C_i(s^{ST}) - 1] \cdot \det
\begin{vmatrix}
f(\theta^{ST}) & 0 & -g(s^{ST}) \\
f(\theta^{ST}) & 0 & 0 \\
0 & -\theta^{AS} \cdot E^L & \alpha \cdot \theta^{AS} \cdot E^L - 1
\end{vmatrix}
\]

\[= g(\theta^{ST}) \cdot \left\{ (\theta^{ST}) \cdot (\theta^{AS} \cdot E^L) \cdot [\theta^{ST} \cdot (E^{H} - E^L)] + f(\theta^{ST}) \cdot [\alpha \cdot \theta^{AS} \cdot E^L - 1] \cdot [\theta^{ST} \cdot (E^{H} - E^L)]
\]

\[= [\alpha \cdot \theta^{ST} \cdot E^{H} - C_i(s^{ST}) - 1] \cdot [g(s^{ST}) \cdot f(\theta^{ST}) \cdot \theta^{AS} \cdot E^L]
\]

\[< 0
\]
\[
\frac{d\theta^{AS}}{d\alpha} < 0.
\]

\[
d_{s^{ST}} = \frac{d}{d\alpha} \begin{bmatrix}
    f(\theta^{ST}) & -f(\theta^{AS}) & 0 & -g(s^{AS}) \\
    f(\theta^{ST}) & 0 & 0 & 0 \\
    0 & \alpha \cdot E^L & -\theta^{AS} \cdot E^L & \alpha \cdot \theta^{AS} \cdot E^L -1 \\
    \alpha \cdot (E^H - E^L) & 0 & -\theta^{ST} \cdot (E^H - E^L) & 1 - \alpha \cdot \theta^{ST} \cdot E^L
  \end{bmatrix}
\]

The numerator:

\[
= \det \begin{bmatrix}
    f(\theta^{ST}) & -f(\theta^{AS}) & 0 & -g(s^{AS}) \\
    f(\theta^{ST}) & 0 & 0 & 0 \\
    0 & \alpha \cdot E^L & -\theta^{AS} \cdot E^L & \alpha \cdot \theta^{AS} \cdot E^L -1 \\
    \alpha \cdot (E^H - E^L) & 0 & -\theta^{ST} \cdot (E^H - E^L) & 1 - \alpha \cdot \theta^{ST} \cdot E^L
  \end{bmatrix}
\]

\[
= f(\theta^{AS}) \cdot \det \begin{bmatrix}
    f(\theta^{ST}) & -f(\theta^{AS}) & 0 & -g(s^{AS}) \\
    0 & -\theta^{AS} \cdot E^L & \alpha \cdot \theta^{AS} \cdot E^L -1 \\
    \alpha \cdot (E^H - E^L) & -\theta^{ST} \cdot (E^H - E^L) & 1 - \alpha \cdot \theta^{ST} \cdot E^L
  \end{bmatrix}
\]

\[
= f(\theta^{AS}) \cdot \{ f(\theta^{ST}) \cdot \{ -\theta^{AS} \cdot E^L \cdot [1 - \alpha \cdot \theta^{ST} \cdot E^L] \} + f(\theta^{ST}) \cdot [\alpha \cdot \theta^{AS} \cdot E^L -1] \cdot [\theta^{ST} \cdot (E^H - E^L)] \}
\]

\[
< 0
\]

\[
\frac{d\theta^{AS}}{d\alpha} < 0
\]

\[
d_{s^{AS}} = \frac{d}{d\alpha} \begin{bmatrix}
    f(\theta^{ST}) & -f(\theta^{AS}) & 0 & 0 \\
    f(\theta^{ST}) & 0 & -g(s^{ST}) & 0 \\
    0 & \alpha \cdot E^L & 0 & -\theta^{AS} \cdot E^L \\
    \alpha \cdot (E^H - E^L) & 0 & \alpha \cdot \theta^{ST} \cdot E^L - C_s(s^{ST}) -1 & -\theta^{ST} \cdot (E^H - E^L)
  \end{bmatrix}
\]

The numerator:

\[
= \det \begin{bmatrix}
    f(\theta^{ST}) & -f(\theta^{AS}) & 0 & 0 \\
    f(\theta^{ST}) & 0 & -g(s^{ST}) & 0 \\
    0 & \alpha \cdot E^L & 0 & -\theta^{AS} \cdot E^L \\
    \alpha \cdot (E^H - E^L) & 0 & \alpha \cdot \theta^{ST} \cdot E^L - C_s(s^{ST}) -1 & -\theta^{ST} \cdot (E^H - E^L)
  \end{bmatrix}
\]
\[
\begin{align*}
&= f(\theta^S) \cdot \det \begin{bmatrix}
 f(\theta^S) & -g(s^{ST}) & 0 \\
 0 & 0 & -\theta^{4S} \cdot E^L \\
 \alpha \cdot (E^H - E^L) & \alpha \cdot \theta^{ST} \cdot E_i^H - C_i(s^{ST}) - 1 & -\theta^{ST} \cdot (E^H - E^L)
\end{bmatrix} \\
&- \alpha \cdot E^L \cdot \det \begin{bmatrix}
 f(\theta^S) & 0 & 0 \\
 f(\theta^S) & -g(s^{ST}) & 0 \\
 \alpha \cdot (E^H - E^L) & \alpha \cdot \theta^{ST} \cdot E_i^H - C_i(s^{ST}) - 1 & -\theta^{ST} \cdot (E^H - E^L)
\end{bmatrix} \\
&= f(\theta^S) \cdot \left\{ g(\theta^S) \cdot \alpha \cdot (E^H - E^L) \right\} + f(\theta^S) \cdot (\theta^{4S} \cdot E_i^L \cdot \left[ \alpha \cdot \theta^{ST} \cdot E_i^H - C_i(s^{ST}) - 1 \right]) \\
&\quad - \alpha \cdot E^L \cdot \left[ f(\theta^S) \cdot \alpha \cdot \theta^{ST} \cdot (E^H - E^L) \right]
\end{align*}
\]

\( \forall \theta^S \cdot [f(\theta^S) \cdot \alpha \cdot \theta^{ST} \cdot (E^H - E^L)] \leq 0 \)

\( \Rightarrow \quad \frac{ds^{4S}}{d\alpha} \geq 0 \).

\[
\begin{align*}
\frac{d\theta^S}{d\phi} &= \frac{\det \begin{bmatrix}
 0 & -f(\theta^{4S}) & 0 & -g(s^{4S}) \\
 0 & 0 & -g(s^{ST}) & 0 \\
 0 & \alpha \cdot E^L & 0 & \alpha \cdot \theta^{4S} \cdot E_i^L - 1 \\
 C_i(s^{ST}) & 0 & \alpha \cdot \theta^{ST} \cdot E_i^H - C_i(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E_i^L
\end{bmatrix}}{
\det \begin{bmatrix}
 f(\theta^{4S}) & -f(\theta^{4S}) & 0 & -g(s^{4S}) \\
 f(\theta^{4S}) & 0 & -g(s^{ST}) & 0 \\
 0 & \alpha \cdot E^L & 0 & \alpha \cdot \theta^{4S} \cdot E_i^L - 1 \\
 \alpha \cdot (E^H - E^L) & 0 & \alpha \cdot \theta^{ST} \cdot E_i^H - C_i(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E_i^L
\end{bmatrix}}
\end{align*}
\]

The numerator:

\[
\begin{align*}
&= \det \begin{bmatrix}
 0 & -f(\theta^{4S}) & 0 & -g(s^{4S}) \\
 0 & 0 & -g(s^{ST}) & 0 \\
 0 & \alpha \cdot E^L & 0 & \alpha \cdot \theta^{4S} \cdot E_i^L - 1 \\
 C_i(s^{ST}) & 0 & \alpha \cdot \theta^{ST} \cdot E_i^H - C_i(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E_i^L
\end{bmatrix} \\
&= -C_i(s^{ST}) \cdot \det \begin{bmatrix}
 -f(\theta^{4S}) & 0 & -g(s^{4S}) \\
 0 & -g(s^{ST}) & 0 \\
 \alpha \cdot E^L & 0 & \alpha \cdot \theta^{4S} \cdot E_i^L - 1
\end{bmatrix} \\
&= -C_i(s^{ST}) \cdot \left\{ f(\theta^{4S}) \cdot g(s^{ST}) \cdot [\alpha \cdot \theta^{4S} \cdot E_i^L - 1] - g(s^{4S}) \cdot g(s^{ST}) \cdot \alpha \cdot E^L \right\} \\
&> 0
\end{align*}
\]

\( \Rightarrow \quad \frac{d\theta^S}{d\phi} > 0 \)
\[
\frac{d\theta^A}{d\phi} = \frac{\det \begin{bmatrix}
  f(\theta^T) & 0 & 0 & -g(s^{45}) \\
  f(\theta^T) & 0 & -g(s^T) & 0 \\
  0 & 0 & 0 & \alpha \cdot \theta^{45} \cdot E^L_i - 1 \\
\end{bmatrix}}{\det \begin{bmatrix}
  f(\theta^T) & -f(\theta^{45}) & 0 & -g(s^{45}) \\
  f(\theta^T) & 0 & -g(s^T) & 0 \\
  0 & \alpha \cdot E^L_i & 0 & \alpha \cdot \theta^{45} \cdot E^L_i - 1 \\
\end{bmatrix}}.
\]

The numerator:

\[
= \det \begin{bmatrix}
  f(\theta^T) & 0 & 0 & -g(s^{45}) \\
  f(\theta^T) & 0 & -g(s^T) & 0 \\
  0 & 0 & 0 & \alpha \cdot \theta^{45} \cdot E^L_i - 1 \\
\end{bmatrix}
\]

The numerator:

\[
= C_{\phi}(s^{45}) \cdot \det \begin{bmatrix}
  f(\theta^T) & -f(\theta^{45}) & 0 & -g(s^{45}) \\
  f(\theta^T) & 0 & -g(s^T) & 0 \\
  0 & \alpha \cdot E^L_i & 0 & \alpha \cdot \theta^{45} \cdot E^L_i - 1 \\
\end{bmatrix}
\]

\[
= C_{\phi}(s^{45}) \cdot \{f(\theta^{45}) \cdot [-g(s^T)] \cdot [\alpha \cdot \theta^{45} \cdot E^L_i - 1]\}
\]

\[
> 0
\]

\[
\Rightarrow \frac{d\theta^A}{d\phi} > 0
\]

\[
\frac{ds^T}{d\phi} = \frac{\det \begin{bmatrix}
  f(\theta^T) & -f(\theta^{45}) & 0 & -g(s^{45}) \\
  f(\theta^T) & 0 & -g(s^T) & 0 \\
  0 & \alpha \cdot E^L_i & 0 & \alpha \cdot \theta^{45} \cdot E^L_i - 1 \\
\end{bmatrix}}{\det \begin{bmatrix}
  f(\theta^T) & -f(\theta^{45}) & 0 & -g(s^{45}) \\
  f(\theta^T) & 0 & -g(s^T) & 0 \\
  0 & \alpha \cdot E^L_i & 0 & \alpha \cdot \theta^{45} \cdot E^L_i - 1 \\
\end{bmatrix}}
\]

The numerator:

\[
= \det \begin{bmatrix}
  f(\theta^T) & -f(\theta^{45}) & 0 & -g(s^{45}) \\
  f(\theta^T) & 0 & -g(s^T) & 0 \\
  0 & \alpha \cdot E^L_i & 0 & \alpha \cdot \theta^{45} \cdot E^L_i - 1 \\
\end{bmatrix}
\]

\[
= -C_{\phi}(s^{45}) \cdot \det \begin{bmatrix}
  f(\theta^T) & -f(\theta^{45}) & -g(s^{45}) \\
  f(\theta^T) & 0 & 0 \\
  0 & \alpha \cdot E^L_i & \alpha \cdot \theta^{45} \cdot E^L_i - 1 \\
\end{bmatrix}
\]

\[
= -C_{\phi}(s^{45}) \cdot \{-g(s^{45}) \cdot f(\theta^T) \cdot \alpha \cdot E^L_i + f(\theta^{45}) \cdot f(\theta^T) \cdot [\alpha \cdot \theta^{45} \cdot E^L_i - 1]\}
\]

\[
> 0
\]
\[ \frac{ds^{ST}}{d\phi} > 0 \]

\[
\begin{bmatrix}
f(\theta^{ST}) & -f(\theta^{AS}) & 0 & 0 \\
f(\theta^{ST}) & 0 & -g(s^{ST}) & 0 \\
0 & \alpha \cdot E^L & 0 & 0 \\
\alpha \cdot (E^{ST} - E^L) & 0 & \alpha \cdot \theta^{ST} \cdot E^{ST} - C_s(s^{ST}) - 1 & C_s(s^{ST})
\end{bmatrix}
\]

\[
\frac{ds^{AS}}{d\phi} = \frac{\det}{\det}
\begin{bmatrix}
f(\theta^{ST}) & -f(\theta^{AS}) & 0 & -g(s^{AS}) \\
f(\theta^{ST}) & 0 & -g(s^{ST}) & 0 \\
0 & \alpha \cdot E^L & 0 & 0 \\
\alpha \cdot (E^{ST} - E^L) & 0 & \alpha \cdot \theta^{ST} \cdot E^{ST} - C_s(s^{ST}) - 1 & 1 - \alpha \cdot \theta^{ST} \cdot E^L
\end{bmatrix}
\]

The numerator:

\[ = \det \begin{bmatrix} f(\theta^{ST}) & -f(\theta^{AS}) & 0 & 0 \\ f(\theta^{ST}) & 0 & -g(s^{ST}) & 0 \\ 0 & \alpha \cdot E^L & 0 & 0 \\ \alpha \cdot (E^{ST} - E^L) & 0 & \alpha \cdot \theta^{ST} \cdot E^{ST} - C_s(s^{ST}) - 1 & C_s(s^{ST}) \end{bmatrix} \]

\[ = C_s(s^{ST}) \cdot \det \begin{bmatrix} f(\theta^{ST}) & -f(\theta^{AS}) & 0 \\ f(\theta^{ST}) & 0 & -g(s^{AS}) \\ 0 & \alpha \cdot E^L & 0 \end{bmatrix} \]

\[ = C_s(s^{ST}) \cdot [f(\theta^{ST}) \cdot g(s^{AS}) - \alpha \cdot E^L] \]

\[ > 0 \]

\[ \Rightarrow \frac{ds^{AS}}{d\phi} > 0 \]

b). Comparative statics for market characters

The size of the H market depends on the value of \( \theta^{ST} \). Since \( \frac{d\theta^{ST}}{d\alpha} < 0 \), an increase in \( \alpha \) will enlarge the H market (\( \frac{\partial y^H}{\partial \alpha} > 0 \)). The size of the L market can be determined from the value of \( s^{AS} \). Since \( \frac{ds^{AS}}{d\alpha} > 0 \), an increase in \( \alpha \) can either expand or shrink the L market (\( \frac{\partial y^L}{\partial \alpha} < 0 \)). Even if \( \frac{\partial y^L}{\partial \alpha} < 0 \), the contraction is not going to outweigh the expansion of the H market. Thus, we expect to see a smaller self-employed sector when \( \alpha \) increases (\( \frac{\partial y_{SE}}{\partial \alpha} < 0 \)).

Similarly, the size of the H market depends on the value of \( \theta^{ST} \). Since \( \frac{d\theta^{ST}}{d\phi} < 0 \), an increase in \( \phi \) will shrink the H market (\( \frac{\partial y^H}{\partial \phi} < 0 \)). The size of the L market can be determined from the value of \( s^{AS} \). Since \( \frac{ds^{AS}}{d\phi} > 0 \), an increase in \( \phi \) will expand the L market (\( \frac{\partial y^L}{\partial \phi} > 0 \)). Again, the expansion of the L market is
smaller the contraction of the H market. Thus, we expect to see a larger self-employed sector when $\phi$ increases ($\frac{\partial W^{SE}}{\partial \phi} > 0$).

The wage level of the $H$ market is the effective reservation wage of the marginal worker. Since $\frac{ds^{ST}}{d\alpha} < 0$, an increase in $\alpha$ will lower the value of $s^{ST}$ and thus raise the effective reservation wage for the marginal worker ($\frac{\partial W^H}{\partial \alpha} > 0$). The wage in the $L$ market similarly depends on the reservation wage of the marginal worker. Since $\frac{ds^{AS}}{d\alpha} \geq 0$, we cannot determine whether wage in the $L$ market will go up or down ($\frac{\partial W^L}{\partial \alpha} \leq 0$).

The wage level of the $H$ market is the effective reservation wage of the marginal worker. Even though $\frac{ds^{ST}}{d\phi} > 0$, we cannot determine whether wage in the $H$ market will go up or down because an increase in $\phi$ itself will increase cost of education ($\frac{\partial W^H}{\partial \phi} \geq 0$). However, an increase in $\phi$ will raise the value of $s^{AS}$ and thus raise the wage in the $L$ market ($\frac{\partial W^L}{\partial \phi} > 0$).