Becker Meets Ricardo: Multisector Matching with Communication and Cognitive Skills*

Robert J. McCann, Xianwen Shi, Aloysius Siow, Ronald Wolthoff†

January 9, 2015

Abstract

This paper presents a tractable framework for studying frictionless matching in education and labor markets when individuals have heterogeneous communication and cognitive skills. In the model, there are gains to specialization and team production, but specialization requires communication and coordination between team members. Individuals accumulate cognitive skills in schools when young. As adults, they decide whether to work as a manager or a worker in a firm or become a teacher in a school. Individuals with more communication skills will become either managers or teachers and earn higher wages. Each manager manages several workers and each teacher teaches several students, with their span of control being determined by their communication skill. These individuals also invest discretely more in education than marginally different individuals who become workers. Equilibrium is equivalent to the solution of an utilitarian social planner solving a linear programming problem.

1 Introduction

A large empirical literature in psychology and a smaller one in economics document the quantitative importance of non-cognitive skills in affecting individual behavior and outcomes.[1] For example, Heckman et al. (2006) show that non-cognitive skills strongly influence schooling decisions, occupational choice, and wages. Kuhn and Weinberger (2005) identify one specific non-cognitive skill that captures the ability to manage other people. They call it "leadership skill." They show that "after controlling for cognitive skills, men who occupied leadership roles in high school earn more as adults" and "high school leaders are more likely to occupy managerial occupations as adults, and leadership skills command a higher wage premium within managerial occupations than elsewhere." In addition, leadership skill plays an important role in individuals' schooling choices as well as in schools' admission decisions. Kuhn and Weinberger (2005) find that individuals with high school leadership experience are 24% more likely to have attained a college degree and 38% more likely to have completed a graduate degree than other high-school graduates with the same cognitive skill. At the same time, selective universities often screen students based on their educational and extra-curricular achievements in addition to other criteria (Duffy and Goldberg, 1998, Klitgaard, 1985).

Despite this empirical evidence, Lazear (2012) concludes from his reading of the literature that there are few, if any, analytic and testable models of leadership.[2] Instead, the theoretical literature has primarily used one-factor, cognitive-skill, models of employee heterogeneity to study the organization of the labor market (e.g., Lucas, 1978; Rosen, 1982; Acemoglu and Autor, 2011; Eeckhout and Kircher, 2012).

This paper therefore develops a framework that captures the above empirical patterns in the labor and education markets by modeling leadership skill as communication skill. Our model builds on three classic ideas in team production: specialization, task assignment, and matching. Adam Smith argued that when workers specialize in different tasks, they can produce more output per worker than if each worker does every task. However, as various authors have pointed out, individuals who specialize and engage in team production typically have to coordinate in order to realize the potential gains to specialization. Such coordination is often costly, either because the pre-agreed action plan may lack the flexibility to tailor production

---

[1]Borghans et al. (2008) and Almlund et al. (2011) review this literature for economists.

[2]He comments on page 92: "The approaches described above cover virtually every aspect of leadership and are rich in description and breadth. Their shortfall, to the extent that there is one, is that the literature does not lend itself well to the type of scientific analysis and proof that could add additional insight into our understanding of the area."

to local information as in Dessein and Santos (2006), or because individuals have to incur explicit communication or coordination costs as in Becker and Murphy (1992), Garicano (2000), and Garicano and Rossi-Hansberg (2004, 2006), among others.

We follow the latter approach and assume that individuals form teams and choose tasks, trading off the gains from specialization against communication costs. Most of the literature assumes that communication cost is a time cost and is the same across individuals. However, as the empirical literature indicates, individuals in general differ in their ability to lead a team, in addition to their cognitive skill. In order to understand how individuals' cognitive skill and communication skill interact and determine outcomes in both labor and education markets, we build a model in which individuals are heterogeneous in both dimensions.

Specifically, we consider a production process in which team production is more effective than production alone but requires individuals to coordinate their actions. These communication activities ("task C") take time away from production ("task P"). Following Garicano (2000), we assume that communication cost is one-sided and is only borne by the team member in task C. Individuals with higher communication skill use less time to communicate with their teammates and therefore have a comparative advantage in task C. The amount of output produced depends on the time spent on task P and the cognitive abilities of the team members, which we assume to be complements in production, as in Becker (1973, 1974).

When young, individuals can augment their cognitive skills in the education market. For analytic convenience, communication skill is assumed to be constant throughout life. Each school in the education market consists of one adult teacher (who does task C) and several students (who do task P). Similar to the labor market, we assume that the teacher's cognitive skill is complementary to his students' initial cognitive skills in education output. Different schools with different teachers may charge different tuition fees. Students choose schools to maximize their future labor market earnings minus tuition cost.

In the labor market equilibrium, we obtain full specialization in tasks, that is, an individual with high communication skill specializes in task C while an individual with low communication skill specializes in task P. Within a team, only one individual will do task C and all the remaining teammates do task P. We refer to the former individual as the manager, and refer to the latter ones as workers. Therefore, our model generates many-to-one teams, a matching pattern commonly observed in the real world. Managers who have more communication skills and thus lower communication cost will manage larger teams (a larger span of control), while the communication skills of workers are irrelevant. Further, the labor market equilibrium exhibits positive assortative matching (PAM) between managers and workers by cognitive skills, which is line with recent evidence by Lazear et al. (2012).

In the education market equilibrium, students with higher initial cognitive skill will choose better teachers (or more education), because of complementarity between students' initial cognitive skill and their teachers' cognitive skill. More interestingly, students with higher communication skill will also demand better teachers, even though students' communication skill does not enter the educational production function. The reason is that these students will become managers or teachers later in life. While the wage of a future worker is independent of his communication skill, a future manager/teacher with more communication skill will have a larger span of control, which increases the return to investment in cognitive skill.[3] As a result, there is a discontinuity in the demand for education. That is, students who expect to become managers or teachers in the future will invest discretely more than marginally different students who expect to become workers in the future. The model also generates an endogenous positive correlation between communication and cognitive skills of adults. Note that in our model, teachers and managers require a similar set of skills. This is in line with the observation in Lazear et al. (2012) that the primary means by which managers matter is through teaching.

Compared to one-factor model, our two-factor model can potentially better fit empirical data in both labor market and education market. For example, in the one-factor model such as Garicano and Rossi-Hansberg (2004, 2006), the labor earnings of employees are fully explained by their cognitive skills. In contrast, our two-factor model predicts that the earnings may depend on both the adult cognitive and communication skills, which is consistent with the finding in Kuhn and Weinberger (2005) who find that "most of the wage effects of leadership skill operate within, not between, very detailed occupational groups." Similarly, in the education market, our two-factor structure allows the possibility that an individual with low initial cognitive ability may have higher aspiration for education attainment than an individual with higher initial cognitive ability, if the former individual has higher leadership skill. Since many selective universities take extra-curriculum activities into account in making admission decisions, our two-factor structure may explain the educational attainment data better.

A two-factor model is also natural for our focus on specialization and endogenous occupation choice. Different from most of existing literature on matching, we assume that the two sides of the market are not exogenous, rather they come from the same pool of individuals. Individuals choose occupations according to their comparative advantage determined by their skills. There are two driving forces for the equilibrium matching pattern and occupation choice: complementarity and span of control. To see this, consider an individual with high

---

[3]Smeets and Warzynski (2008) finds a positive relationship between managers' wages and the number of individuals they supervise.

cognitive ability but moderate communication skill. On one hand, complementarity in cognitive skills implies that it may be beneficial to assign this individual as a worker and match him with a manager with high cognitive skill. On the other hand, it may also be beneficial to make this individual a manager himself. He can then manage multiple workers who all produce more output because of his high cognitive ability. The equilibrium strikes a balance between these two forces, exhibiting dispersion in both team size and productivity.

To illustrate the equilibrium properties, we simulate the equilibrium of the model with a bivariate uniform distribution of initial cognitive abilities and communication skills. The simulation shows that there are two distinct groups of teachers: one group teaches future workers while the other teaches future teachers and managers. The equilibrium generates an earnings distribution which is qualitatively consistent with a single peaked right skewed distribution.

Finally, the paper also makes a methodological contribution. We show that the equilibrium of our multi-factor multi-market matching model with endogenous occupational choice is equivalent to the solution of a utilitarian social planner solving a linear programming problem, substantially extending the existing results. It also facilitates the proof for the existence of equilibrium and makes it easy to numerically compute the equilibrium.

A caveat is in order. In this paper, the role for communication skill in team production is narrow and we identify it as the ability to lead. In reality, leadership skill is broader and it can include many other non-cognitive skills, such as persistence, reliability, self-discipline, and inter-personal skills. Therefore, our model should be viewed as a first-pass theory of frictionless assignment with cognitive and non-cognitive skills. Due to the lack of a reliable measure of non-cognitive skills, empirical studies on how they affect market outcomes are rather limited, but as we briefly outlined above, their findings are broadly consistent with our equilibrium characterization.

This paper proceeds as follows. Section 2 introduces the model and discusses the micro-foundation of the production technology. In Section 3, we characterize the equilibrium in each market and formulate the planner's problem as a linear programming problem. In Section 4, we illustrate the properties of the equilibrium by simulation. Section 5 discusses related literature. Several extensions of the baseline model are briefly discussed in Section 6. Section 7 concludes.

2 Model

2.1 Setting

We consider the steady-state in an economy with discrete time periods and infinite horizon. In each period, a unit measure of risk-neutral individuals is born and lives for two periods. At birth, each individual draws a two-dimensional skill vector (n, a) from a common non-degenerate distribution with support on $[\underline{n}, \overline{n}] \times [\underline{a}, \overline{a}]$, where $0 < \underline{n} < \overline{n}$ and $0 < \underline{a} < \overline{a}$. We call n an individual's "communication skill" and a his "initial cognitive ability."

In the first period of their lives, individuals enter the education market as students in order to augment their cognitive ability. Education takes place in schools, each consisting of one teacher and a certain number of students. A school charges each of its students a tuition fee τ which it uses to hire the teacher. Schools can enter the education market freely, so they make zero profit in equilibrium, i.e., the tuition fees exactly cover the teachers' wage ω. The initial cognitive ability a of a student and his school choice determine the final cognitive skill of this student upon graduation, which we denote by k, with $k \in [\underline{k}, \overline{k}]$, $0 < \underline{k} < \overline{k}$. The communication skill n is fixed for the entire life of the individual and cannot be changed.[4]

After graduating from school, students become adults in the second period and enter the labor market. Each adult can become either a teacher in one of the schools or an employee in a firm. Each firm employs a certain number of individuals to produce output which is sold at a normalized price of 1. Firms can freely enter the labor market. Therefore, each firm makes zero profit in equilibrium, and its output exactly covers the sum of the wages ω of its employees.

We assume that all markets in the economy are perfectly competitive. Furthermore, we assume that individuals can freely borrow from a perfect capital market and there is no discounting. Therefore, individuals' lifetime payoffs are equal to the sum of their labor market earnings minus tuition costs, $\omega - \tau$. Individuals choose who to match with in each of the two markets to maximize these payoffs.

2.2 Labor Market

In the labor market, output is produced by the completion of productive tasks. Individuals can produce either alone (i.e., a firm of size 1) or in a team (firm size larger than 1), and each

---

[4]Empirical evidence shows that communication skill and other non-cognitive skills can be acquired when young (e.g., Heckman et al. 2006, 2012). Our framework can accommodate this acquisition but we ignore it here to keep the model simple.

individual is endowed with one unit of time. Team production is potentially advantageous because team members can help and direct each other, but such collaboration is costly. In order to introduce our production technology, we discuss the output produced first by an individual, then by a team of two individuals, and lastly by a team of three or more individuals.

Consider first an individual i of type (n_i, k_i) who works alone. The output that he produces is assumed to be proportional to his cognitive skill k_i:

$$Y(\theta_i; n_i, k_i) = \beta k_i \theta_i, \tag{1}$$

where $\theta_i \in [0, 1]$ is the amount of time that he spends on production and $\beta \in (0, 1)$ is a penalty for solitary production, capturing the idea that producing alone may be less effective. Notice that individual i's communication skill n_i does not enter the production function, because by working alone, there is no need for communication.

Communication becomes relevant when individuals produce in teams. We model the role of communication in team production as follows. If a type-(n_i, k_i) team member i encounters a problem, he can turn to one of his teammates (say, member j) for help. However, offering help is time-consuming, because it takes away the amount of time that member j can allocate to production. A helper with higher communication skill is assumed to be more effective in helping: the amount of time taken away from production is smaller. Formally, if member j with communication skill n_j spends $\psi_j \in [0, 1]$ units of time on helping, his effective helping time is $n_j \psi_j$. To capture the idea that the probability for a member to encounter problems in team production is proportional to his time on production, we assume that, to ensure smooth production, on average the "effective" helping time incurred by the helper must be equal to the production time of the assisted individual. Specifically, if member i spends θ_i units of time on production, member j has to spend θ_i/n_j units of time on helping member i.[5] We assume, as in Garicano (2000), that the time cost of communication is only incurred by the helper (i.e., member j).

In a two-member team consisting of individual i of type (n_i, k_i) and individual j of type (n_j, k_j), the two individuals can only help each other. Hence, if i and j spend θ_i and θ_j units of time on production, respectively, then i has to spend $\psi_i = \theta_j/n_i$ units of time in helping j, and j has to spend $\psi_j = \theta_i/n_j$ units of time in helping i. Total team output is given by

$$Y(\theta_i, \theta_j; n_i, k_i, n_j, k_j) = \sqrt{k_i k_j}\theta_i + \sqrt{k_i k_j}\theta_j \tag{2}$$

---

[5] In other words, we assume that time spend on production and time spend on helping are perfect complements in production. As we discuss in section 6.1, this assumption is mostly for analytical convenience; it can be relaxed without qualitative implications.

subject to $\theta_i + \psi_i \leq 1$ and $\theta_j + \psi_j \leq 1$.[6]

Note that the two-member team production function (2) differs from the working alone production function (1) in two ways. First, we omit the parameter β to reflect the benefits of team production with communication. Second, the output produced now depends on the cognitive skill of both team members. In particular, we assume the cognitive skills of the two team members are complementary in producing output, which is captured by the term $\sqrt{k_i k_j}$.[7] As is well known from Becker, all other things equal, this complementarity will induce positive assortative matching (PAM) in two-member teams by cognitive skills in that sector. We employ the complementarity specification here for that reason.

Note that in the two-member team, both individuals need to spend some time on communication, that is, there is no full specialization in tasks. It may therefore be advantageous to extend a two-member team to a three-member team with a third member (say member m) who specializes in communication. The output of a three-member team is given by

$$\sqrt{k_i k_m}\theta_i + \sqrt{k_j k_m}\theta_j \tag{3}$$

subject to $\theta_i + \theta_j = n_m \psi_m$ with $\theta_i, \theta_j, \psi_m \in [0, 1]$. Note that if team member m has a sufficiently large n_m, she may have extra time left which she can use to help more workers, increasing the team size even further. In particular, if n_m is an integer,[8] then individual m can help n_m workers who each spend one unit of time on production. In this case, the output is given by

$$\sum_{i=1}^{n_m} \sqrt{k_i k_m}.$$

Given these production functions, firms need to decide which employees they wish to hire and how to allocate them to the two different tasks. To keep the exposition as simple as possible, we will often refer to the productive task as "task P" and to the communicative task ("helping") as "task C." We assume throughout that

$$\underline{n} \geq 2 \text{ and } \beta \leq 2/3 \tag{4}$$

so that team production is always superior to working alone.[9]

---

[6]If $\bar{k} \leq 1$, one can, as in Kremer (1993), interpret $\sqrt{k_i}$ as the probability for individual i to succeed in performing task P while $\sqrt{k_j}$ the probability for individual j to succeed in performing task C.

[7]This specific functional form is again not essential. We only require constant return to scale and supermodularity in k_i and k_j, as we discuss in more detail in section 6.1.

[8]Given constant return to scale, we will ignore the integer problem throughout the paper to ease exposition.

[9]To see why condition (4) is sufficient, note that a type-(n, k) individual produces output βk by working

8

2.3 Education Market

In the education market, task assignment is exogenous: students learn and teachers instruct. Let individual i be a student and individual j be a teacher. The inputs in the production function are the initial cognitive skill a_i of student i and the adult cognitive skill k_j of the teacher j, while output is the adult cognitive skill k_i of individual i. If a type-(n_i, a_i) student spends θ_i units of time on learning, then his teacher has to spend $\psi_j = \theta_i / n_j$ units of time on teaching. Hence, the production technology can be adapted to the education market as follows

$$k_i = \sqrt{a_i k_j} \theta_i \tag{5}$$

subject to $\theta_i = n_j \psi_j$ with $\theta_i, \psi_j \in [0, 1]$.

We conclude the model description with a brief discussion of the restrictions and implications of several simplifying assumptions we have imposed on the production technology for tractability. First, we assume that effective helping is linear in both communication skill and the amount of helping time. This implies that the production has constant return to scale in helping time, and hence we get many-to-one matching in teams. Without this linearity assumption, the analysis is no longer tractable. Second, we assume that communication cost is one-sided: only the helper has to incur cost to help while the helped incur no cost. As a result, the communication skill of the workers does not matter for production, and thus does not factor into their wages. This assumption is standard in the literature (e.g., Garicano 2000), and can be partially relaxed at a cost of increased analytical complexity. Third, we assume that the production technology in the education market has a similar structure as in the labor market. In particular, students' communication skill matters for choosing schools but does not matter for improving their cognitive skills conditional on their teacher choice. In addition, teachers' communication skill determines the number of students they each can mentor but it does not matter for individual students in accumulating their cognitive skills. In other words, we assume that, from a student's perspective, a teacher with poor communication skill can do as well as one with better communication skill by spending more time with the student. Again this assumption is restrictive, but it substantially improves tractability and offers strong predictions in terms of schooling patterns and occupation choice.

---

alone. Now suppose he forms a three-person team with two other type-(n, k) individuals: two of them perform task P, and the third does task C. If $n \geq 2$, this team produces at least $2k$ according to our team production function (3). Therefore, team production is better than working alone as long as (4) holds.

3 Equilibrium

3.1 Labor Market Equilibrium

It is convenient for us to first analyze the labor market. An adult in the labor market is characterized by his communication and cognitive skills, (n, k). Here we take the distribution $\alpha(n, k)$ of adult skills as given, but we will endogenize it when we study the education market.

Let $\omega(n, k)$ denote the equilibrium wage for a type-(n, k) adult. Each adult can either work for a firm as an employee or work for a school as a teacher. If a firm hires the adult for $\theta \in [0, 1]$ units of time, then the firm will pay the adult $\theta\omega(n, k)$. The adult will supply $(1 - \theta)$ units of time to other firms and earn $(1 - \theta)\omega(n, k)$ from them. If a school hires a type (n, k) adult as a teacher, it also has to pay $\omega(n, k)$ per unit of time.

There is a perfectly elastic supply of firms and schools in the labor market. So in equilibrium, all firms and schools make zero profit, and the wage function $\omega(n, k)$ must satisfy the demand of firms and schools for adults. For now, we focus only on the problem that firms solve.

A firm is a collection of teams. Given our production function, there is no interaction across teams, so we can study the problem of one team in a firm. A team is a collection of employees chosen by the firm who assigns tasks to them and pays them market wages. We first establish that our production function implies full specialization in task assignment in the labor market.

Proposition 1. *Each team's profit is maximized by allocating every employee to a specific task, either P or C, for the entire length of the production process.*

Proof. Consider a team which employs a type-(n, k) adult for a short time interval Δ. The firm can allocate the employee to either task P or task C. If employee (n, k) is allocated to task P during the time interval Δ, then the firm has to hire another adult (n', k') from the labor market to perform task C for $\Delta(n')^{-1}$ units of time in order to produce output $\sqrt{kk'}\Delta$. Choosing (n', k') optimally, the firm's profits of having (n, k) in task P for Δ time interval is given by

$$\pi^P(n, k, \Delta) = \max_{(n', k')} \sqrt{kk'}\Delta - \omega(n', k')(n')^{-1}\Delta - \omega(n, k)\Delta. \tag{6}$$

If employee (n, k) is instead allocated to task C in period Δ, then the firm needs to hire another adult (n'', k'') to do task P for $n\Delta$ units of time to produce output $\sqrt{kk''}n\Delta$. The

10

associated profits $\pi^C(n, k, \Delta)$ are

$$\pi^C(n, k, \Delta) = \max_{(k'', n'')} \sqrt{kk''} n\Delta - \omega(n'', k'') n\Delta - \omega(n, k)\Delta. \tag{7}$$

Therefore, the firm would assign employee (n, k) to task P if and only if $\pi^P(n, k, \Delta) - \pi^C(n, k, \Delta) \geq 0$. The sign of $\pi^P(n, k) - \pi^C(n, k)$ is independent of Δ, the length of time that is available for production. Consequently, the firm's profits are maximized by allocating employee (n, k) to either P or C for the entire duration of the production process. $\qquad \square$

Proposition 1 indicates that within a firm each employee will specialize in performing either task P or task C. Moreover, since all firms have access to the same production technology, Proposition 1 also implies that task assignments must be the same across firms: if one firm strongly (weakly) prefers to assign an employee of type (n, k) to task P (C), all other firms will do the same. As a result, it is without loss of generality to assume that each team has only one team member performing task C.[10] This proposition also implies that task assignment within a team is determined by comparative advantage.

Corollary 1. *Consider any two members with types (n_i, k_i) and (n_j, k_j) in a team. Member i will be assigned to task P if*

$$\omega(n_i, k_i)(1 - n_i^{-1}) < \omega(n_j, k_j)(1 - n_j^{-1})$$

The above corollary is obtained by comparing the profits from assigning member i to task P and j to task C to produce a fixed amount of output, versus the reverse assignment. The insight of the above corollary is known since Ricardo.

Finally, for fixed cognitive skill k, task assignment is sorted according to individuals' communication skills.

Proposition 2. *For each cognitive skill level k, there exists a cutoff value $\widehat{n}(k) \in [\underline{n}, \overline{n}]$ such that individuals with communication skill $n < \widehat{n}(k)$ perform task P, and individuals with communication skill $n \geq \widehat{n}(k)$ perform task C.*

Proof. Applying the envelope theorem to (6) and (7) yields

$$\frac{d}{dn}\left(\pi^P(n, k, \Delta) - \pi^C(n, k, \Delta)\right) = -\left[\sqrt{kk''} + \omega(n'', k'')\right]\Delta < 0.$$

---

[10]If there are $q > 1$ team members performing task C in a particular team, the firm hiring these team members can split and re-organize the team such that each (new) team has only one member performing task C, without lowering profits.

Hence, the value of $\pi^P(n, k, \Delta) - \pi^C(n, k, \Delta)$ crosses zero only once and from above. $\square$

We will call the employees who are optimally assigned to task P, i.e. $\pi^P(n, k, \Delta) \geq \pi^C(n, k, \Delta)$, 'workers' and denote their occupation by w. Note that the amount of team output produced in Δ time interval is $\sqrt{kk'}\Delta$ which is independent of n. In other words, the firm does not value a worker's communication skill and thus will not be willing to pay for it. Instead, it only pays attention to the worker's cognitive skill k. Therefore, the equilibrium wage of workers of skill (n, k), which we denote by $\omega_w(n, k)$, is independent of n. To simplify notation, we will write $\omega(k) \equiv \omega_w(n, k)$.

On the other hand, if the employee (n, k) is assigned to task C, the profit $\pi^C(n, k, \Delta)$ from hiring (n, k) depends on n. We call these employees 'managers' and denote their occupation by m. Their wages will depend on both n and k and are denoted by $\omega_m(n, k)$.

Consider a team with a manager of type (n, k) where $n \geq 2$. According to Proposition 1, this manager is only matched with other employees who perform task P, i.e., workers. Hence, all teams in the labor market consist of many-to-one matchings. The number of workers that a manager supervises, which can be interpreted as the span of control or the capacity of the manager, is exactly equal to the manager's communication skill.

Let the team with a type-(n, k) manager choose n workers with respective types $(k_1, ..., k_n)$ in order to maximizes its profits. The team solves the following maximization problem:[11]

$$\max_{(k_1,...,k_n)} \sum_{i=1}^{n} \left[\sqrt{kk_i} - \omega(k_i) \right] - \omega_m(n, k).$$

Given the additive separability of the total output, the optimal choice of workers satisfies $k_1^* = ... = k_n^* = \mu(k)$ with

$$\mu(k) \in \arg\max_{k'} \sqrt{kk'} - \omega(k'). \tag{8}$$

Therefore, we have proved the following result.[12]

Lemma 1. *In equilibrium, it is optimal for a team to hire workers with the same cognitive skill.*

---

[11]For expositional purposes, we treat n as an integer here. More generally, we can write the maximization problem as follows: $\max_{\{k_i\}_{i \in [0,n]}} \int_0^n \left[\sqrt{k \cdot k_i} - \omega(k_i) \right] di - \omega_m(n, k)$.

[12]This result indicates that all team members, other than the manager, are homogenous in their cognitive skill. This is consistent with the findings of Lopes de Melo (2013), who documents a substantial degree of clustering and segregation of coworkers in terms of their (cognitive) skills.

The function $\mu(k)$ determines the worker type matched to a type-(n, k) manager. It depends on the manager's cognitive skill k, but not on his or her communication skill n, and fully captures the sorting between workers and managers in the labor market. Hence, we call $\mu(k)$ the equilibrium matching function in the labor market. Given Lemma 1, we can rewrite the profits of the team with manager (n, k) as follows:

$$n\left[\sqrt{k\mu(k)} - \omega(\mu(k))\right] - \omega_m(n, k).$$

The free-entry condition for firms implies that the above expression must be zero. Therefore, the manager's wage is given by $\omega_m(n, k) = n\phi(k)$, where

$$\phi(k) \equiv \sqrt{k\mu(k)} - \omega(\mu(k)) = \max_{k'} \sqrt{kk'} - \omega(k') \tag{9}$$

denote the the profits per worker generated by the manager.

Next, we address the issue of sorting in labor market. Both workers and managers are heterogeneous in their cognitive skill, so an important question is which worker types work for which manager. Applying the envelope theorem to equation (8), we obtain:

Lemma 2. *The equilibrium matching function $\mu(k)$ is strictly increasing.*

Given that $\mu(k)$ is weakly increasing, we can define the generalized inverse function $\mu^{-1}(\cdot)$ of $\mu(\cdot)$ as

$$\mu^{-1}(k) = \min\{k' : \mu(k') = k\}.$$

That is, $\mu^{-1}(k)$ is the lowest cognitive skill among managers hiring type-k workers. Now we can link the equilibrium wage $\omega(k)$ and $\phi(k)$ with the equilibrium matching function $\mu(k)$.

Lemma 3. *Given an equilibrium matching function $\mu(k)$, wages $\phi(k)$ and $\omega(k)$ must satisfy the following differentiation equations:*

$$\frac{d\phi(k)}{dk} = \frac{1}{2}\sqrt{\frac{\mu(k)}{k}} \text{ and } \frac{d\omega(k)}{dk} = \frac{1}{2}\sqrt{\frac{\mu^{-1}(k)}{k}}.$$

Proof. We can apply the envelope theorem to (9) and obtain that

$$\frac{d\phi(k)}{dk} = \frac{1}{2}\sqrt{\frac{\mu(k)}{k}}. \tag{10}$$

13

Furthermore, the necessary first-order condition of the maximization problem (9) is

$$\frac{d\omega\left(k'\right)}{dk'}|_{k'=\mu(k)} = \frac{1}{2}\sqrt{\frac{k}{k'}},$$

which can be rewritten as

$$\frac{d\omega\left(k\right)}{dk} = \frac{1}{2}\sqrt{\frac{\mu^{-1}\left(k\right)}{k}}. \tag{11}$$

$\square$

To conclude the analysis of labor market, we should also characterize the equilibrium occupation choice. However, since schools also compete in the labor market for teachers, we will carry out the analysis of occupation choice when we investigate the education market.

3.2 Education Market

We now turn to the education market. Given our production function, let a type-(n_t, k_t) teacher manage up to n_t students. Since the adult cognitive skill that a student accumulates depends only on the student's initial cognitive ability a_s and the cognitive skill of her teacher k_t, the tuition charged by a school with teacher (n_t, k_t) does not depend on the communication skill of the teacher, n_t, and we can write tuition as $\tau\left(k_t\right)$. The profit for a school with a type-(n_t, k_t) teacher and n_t students is $n_t\tau\left(k_t\right) - \omega_t\left(n_t, k_t\right)$, so the free entry condition for schools implies that the teacher's wage must equal

$$\omega_t\left(n_t, k_t\right) = n_t\tau\left(k_t\right). \tag{12}$$

A simple arbitrage argument then shows that $\tau\left(k_t\right)$ must be increasing in k_t.

In order to study the education choice of students, we need to compute the return to schooling which depends on labor market earnings. Since equilibrium wages in the labor market vary across occupations, it is useful to first discuss the equilibrium occupation choice in the labor market. Note that a type-(n, k) adult can choose to become a worker, a manager, or a teacher, so his wage must be

$$\omega\left(n, k\right) = \max\left\{\omega_w\left(n, k\right), \omega_m\left(n, k\right), \omega_t\left(n, k\right)\right\} = \max\left\{\omega\left(k\right), n\phi\left(k\right), n\tau\left(k\right)\right\}.$$

For a given cognitive skill level k, there exists a cutoff $\widehat{n}\left(k\right) \in [\underline{n}, \overline{n}]$ such that adults with communication skill $n < \widehat{n}\left(k\right)$ become workers, and individuals with communication skill $n \geq \widehat{n}\left(k\right)$ become either managers or teachers. In particular, if $\widehat{n}\left(k\right)$ is interior, then a

14

type-$(\widehat{n}(k), k)$ adult will be either indifferent between becoming a worker and becoming a manager, or indifferent between becoming a worker and becoming a teacher. For the former case, we must have

$$\widehat{n}(k) = \omega(k)/\phi(k), \tag{13}$$

while for the latter, we must have

$$\widehat{n}(k) = \omega(k)/\tau(k). \tag{14}$$

Note that if, for a particular value of k, there are both managers and teachers, then we must have $\phi(k) = \tau(k)$. In this case, we cannot separate managers from teachers in terms of communication skill n, because a manager with communication skill $(n_1 + n_2)$ can switch his position with two teachers with respective communication skill n_1 and n_2, and vice versa. In this case, the equilibrium masses of managers and teachers are indeterminate.[13]

Now consider the schooling decision of a type-(n_s, a_s) student. He chooses the type of school k_t that maximizes his lifetime income. Going to a school with a better teacher k_t results in a higher value of cognitive skill and thus a higher payoff in the labor market, but comes at a higher cost of tuition. Let $\rho(n_s, a_s)$ denote the equilibrium school choice of a type-(n_s, a_s) student. Formally, $\rho(n_s, a_s)$ is defined as

$$\rho(n_s, a_s) \in \arg\max_{k_t} \left[\max\left\{ \omega\left(\sqrt{a_s k_t}\right), n_s\phi\left(\sqrt{a_s k_t}\right), n_s\tau\left(\sqrt{a_s k_t}\right) \right\} - \tau(k_t) \right]$$

The next lemma addresses sorting in the education market, that is, how the equilibrium school choice $\rho(n_s, a_s)$ varies with respect to a student's characteristics.

Lemma 4. *Given n_s and conditional on becoming a worker or a manager, a student with higher initial cognitive skill a_s will choose a teacher with higher k_t, that is, $\frac{\partial \rho(n_s, a_s)}{\partial a_s} > 0$. Given a_s and conditional on becoming a worker, $\rho(n_s, a_s)$ is independent of n_s. However, given a_s and conditional on becoming a manager, a student with higher communication skill n_s will choose a teacher with higher k_t, that is, $\frac{\partial \rho(n_s, a_s)}{\partial n_s} > 0$.*

Proof. See Appendix. □

Therefore, the education choice of future managers is increasing in their initial cognitive ability and communication skill. The latter effect may seem surprising, because communication skill n_s does not enter the education production function. Future managers with higher

---

[13]In the simulation that we describe in section 4, however, the equilibrium masses of managers and teachers are fully determined.

n_s, however, demand more education (or better teachers), because the gains from education are proportional to their span of control: by increasing their education choice today, they increase their adult cognitive ability, which will in turn increase the productivity of every worker in their future team.

In equilibrium, students with different cognitive abilities may choose the same school because of compensating differences in their communication skill. Given a school with teachers of cognitive skill k_t, the students' cognitive achievements, k_s, are weakly decreasing in n_s and increasing in a_s. That is, within a school, cognitive skill achievements by students are inversely correlated with their communication skills. Lazear (2012) found that "there is a large and negative association between being a leader, as measured by 'CLEVEL' (CEO, COO, CFO...), and grade point average" among Stanford MBA students. Given how selective the Stanford MBA program is, it is surprising that our prediction holds for such a selected group of students.[14]

Finally, we want to highlight one interesting feature of the equilibrium education choice in our model. Specifically, we show that on the margin students who become managers or teachers invest discretely more in schooling than students who become workers. Intuitively, with endogenous occupation choice, the return to schooling has a kink for the marginal student type who is indifferent between two occupations, creating a wedge between the optimal education choice for students with communication skill just below and above this marginal type.

Proposition 3. *Let $k_t = \rho\left(n_s, a_s\right)$ denote the equilibrium education choice of a type-(n_s, a_s) student. If $\widehat{n}'(k) \neq 0$ at $k = k_t$, then $\rho\left(n_s, a_s\right)$ is discontinuous at $n_s = \widehat{n}(\sqrt{a_s k_t})$.*

Proof. See Appendix. ∎

The polarization of education choice is illustrated in Figure 1. The x-axis represents the students' eventual cognitive ability $k_s = \sqrt{a_s k_t}$, and $c\left(k_s\right)$ represents the education cost for a student type $\left(a_s, n_s\right)$. Both $c\left(k_s\right)$ and $\omega\left(k_s\right)$ do not depend the students' communication skills n_s, while the wage for managers increases in n_s. Fix the student's ability a_s. Then the two curves $c\left(k_s\right)$ and $\omega\left(k_s\right)$ are fixed, and thus the optimal education choice k_s is determined by maximizing the distance $\omega\left(k_s\right) - c\left(k_s\right)$. Let k_s^* denote the optimal eventual education choice for student a_s if he aims to be a worker. By varying n_s, there exist one n_s^* such that the wage curve for manager type (n_s^*, k_s^*) passes through the point $(k_s^*, \omega\left(k_s^*\right))$. If in equilibrium

---

[14]Hence, our model suggests that George Bush Jr., unlike popular belief, went to Yale to augment his cognitive skill rather than to build his social network.

there exist adults of type (n_s^*, k_s^*), then k_s^* must also be optimal choice for students of type (a_s, n_s^*) who want to be a manager. This is generically impossible, as illustrated in the figure: the distance $n\phi(k_s) - c(k_s)$ will not be maximized at k_s^*.
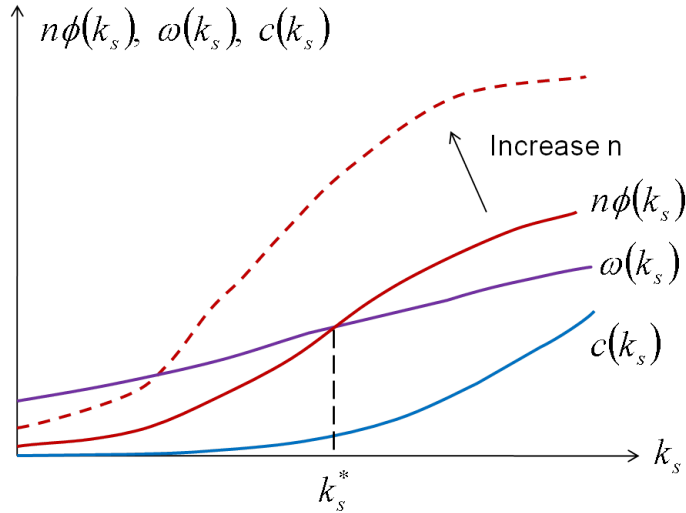

Figure 1: Polarization in the education choice

In the empirical literature, both the benefit and cost functions to schooling are typically assumed to be smooth functions of the amount of schooling. Thus individuals who differ slightly in cognitive abilities will obtain similar levels of schooling. Bunching in schooling, where the marginal student who is indifferent between stopping with high school graduation or going to junior college for two years rather than an additional year is usually attributed to asymmetric information as in the Spence signalling model. This model generates bunching in schooling without any informational problems or non-competitive behavior.

3.3   Equilibrium and Linear Programming Formulation

We can now formally define the equilibrium in our model. Recall that, although our model has an overlapping-generation structure, we focus on the steady state equilibrium by treating it as a two-period model.

Definition 1. A (steady state) equilibrium consists of wages $\omega(n, k)$, tuition $\tau(k)$, matching functions $\mu(k)$ and $\rho(n, k)$ for the labor market and education market, respectively, and a distribution $\alpha(n, k)$ of adult types, such that

1. Profit maximization: firms and schools choose the number and types of individuals to form teams to maximize their profits, given wages and tuition.

17

2. Free entry: the number of firms and schools is such that each firm and school earns zero profits.

3. Utility maximization: individuals choose who to match with in each sector and how to divide tasks to maximize their lifetime payoff, given wages and tuition.

4. Market clearing: wages and tuition are such that demand equals supply for each type of adult and/or student in each of the two sectors.

5. Consistency: the distribution of adult types is consistent with educational choices and the distribution of student types.

Because of the consistency requirement, solving the equilibrium may seem a daunting task. Fortunately, one can formulate the planner's optimization problem as a linear programming problem and the decentralized equilibrium as its dual. Since markets are competitive and there are no externalities, the planner's solution can be implemented in decentralized markets. The equilibrium wages and utilities can be described as Lagrange multipliers corresponding to the constraints in the planner's problem, and they must solve the dual to the planner's problem. We briefly describe below how one can formulate the planner's problem as a linear programming problem, as this approach greatly facilitates our numerical simulation in the next section. We refer interested readers to Erlinger et al. (2014) for a detailed analysis of the linear programming approach.

Let $A \equiv [\underline{n}, \overline{n}] \times [\underline{k}, \overline{k}]$ denote the type space for adults. Let $S \equiv [\underline{n}, \overline{n}] \times [\underline{a}, \overline{a}]$ denote the type space for students. Let $\sigma(n_s, a_s) \geq 0$ denote the exogenously given, absolutely continuous probability measure of student types with $\sigma(S) = 1$. Let $\varepsilon \geq 0$ denote the joint measure on $S \times A$ of many-to-one student-teacher pairings in the education market, and let $\lambda \geq 0$ denote the joint measure on $A \times A$ of many-to-one pairings of workers to managers in the labor market. The supply and demand constraint in the education market requires that the total number of type-(n_s, a_s) students in all schools cannot exceed the total supply of type-(n_s, a_s) students. Since a type-(n_t, k_t) teacher can mentor n_t students, we must have, for all (n_s, a_s),

$$\int_{(n_t, k_t) \in A} n_t \varepsilon(n_s, a_s; dn_t, dk_t) \leq \sigma(n_s, a_s). \tag{15}$$

Similarly, in the labor market, the total demand of type-(n, k) workers, type-(n, k) managers, and type-(n, k) teachers must not exceed the total supply of type-(n, k) adults. Since a manager of type-(n_m, k_m) has the capacity to supervise up to n_m workers, we must have, for all

(n, k),

$$\int_{(n_m, k_m) \in A} n_m \lambda(n, k; dn_m, dk_m) + \int_{(n_w, k_w) \in A} \lambda(dn_w, dk_w; n, k)$$
$$+ \int_{(a_s, n_s) \in S} \varepsilon(dn_s, da_s; n, k) \leq \int_{(n_t, k_t) \in A} \frac{2k}{k_t} n_t \varepsilon(n, k^2/k_t; dn_t, dk_t). \qquad (16)$$

where the term $2k/k_t$ on the right hand side is due to a change of variable $a_s = k^2/k_t$.

Given our production technology, the revenue for a team consisting of a type-(n_m, k_m) manager and n_m type-(n_w, k_w) workers is $n_m \sqrt{k_w k_m}$, independent of the workers' communication skill n_w. Thus the planner's primal linear program is given by

$$\sup_{\varepsilon, \lambda} \int_{A \times A} n_m \sqrt{k_w k_m} \lambda(dn_w, dk_w; dn_m, dk_m) \qquad (17)$$
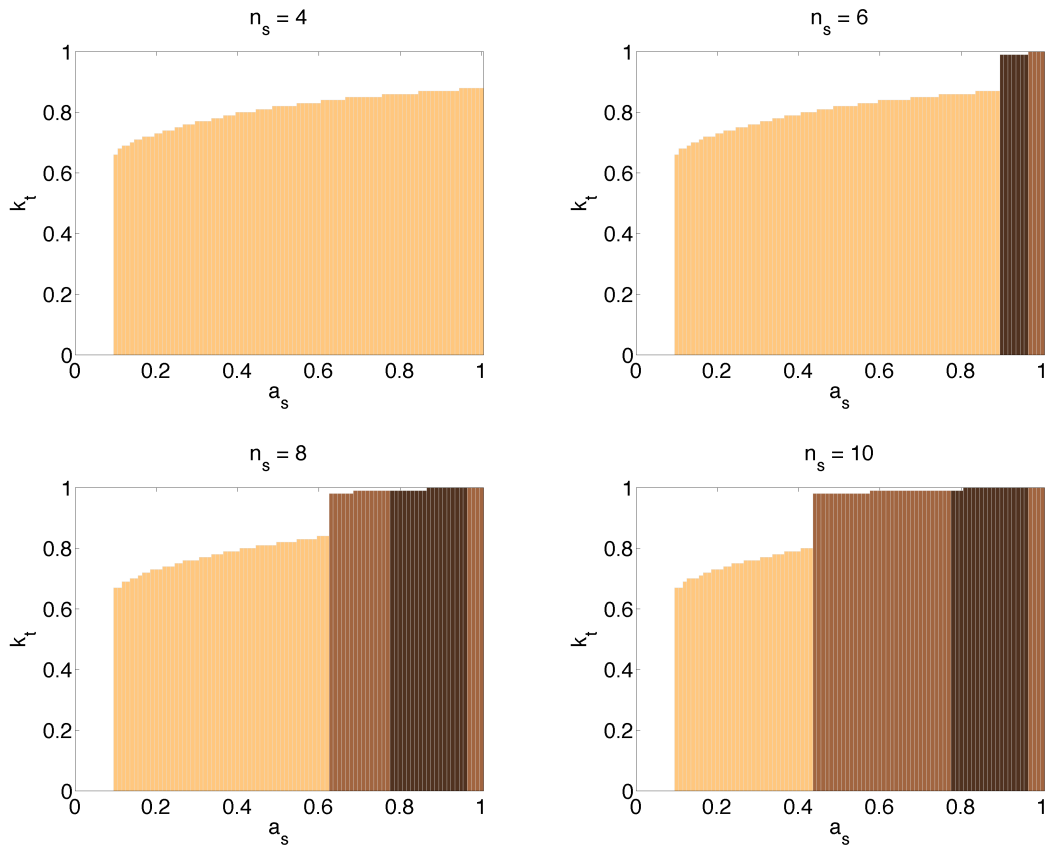
given the constraints $\varepsilon \geq 0$, $\lambda \geq 0$, (15), and (16). The Lagrange multipliers attached to (16) will be wages $w(n, k)$, and the multipliers attached to (15) will be student indirect utilities $u(n, a)$, equal to the difference between their future labor wages and their tuition costs.

4   Simulation

We illustrate the properties of the equilibrium by simulating the model. The simulation is computationally straightforward because of the equivalence between the market equilibrium and the solution of the planner's problem, which is a linear program as shown in the previous section. We solve the primal problem (17), which gives us the equilibrium utilities as the multipliers.

For the simulation, we consider a mass 1 of students with uniformly distributed skills. A bivariate uniform distribution is not necessarily a realistic assumption. However, it helps to highlight that 1) no right tail in the skill distribution is required to obtain a right tail in the earnings distribution, and 2) no correlation in initial abilities is required to observe a positive correlation in the adult population. The support of the distribution of communication skill n and initial cognitive skill a is assumed to be equal to $[\underline{n}, \overline{n}] \times [\underline{a}, \overline{a}] = [2, 10] \times [0.1, 1]$. The distribution of adult cognitive skill k and its support $[\underline{k}, \overline{k}]$ are endogenously determined through the students' education choices. We discretize the support of both student and adult skill for computational reasons, by using 17 grid points for n and 91 grid points for a and k.

The results of the simulations are presented in figure 2-5. Figure 2 displays the education

The cognitive skill $k_t = \rho(n_s, a_s)$ of the teacher that is chosen by a student with communication skill n_s and initial cognitive skill a_s. Yellow students become workers later in life, light-brown students become teachers, and dark-brown students become managers.

Figure 2: Education choice as function of initial cognitive skill

choice $\rho\left(n_s, a_s\right)$. It shows the cognitive skill k_t of the teacher chosen by students with initial cognitive skill a_s for four different values of the communication skill, $n_s \in \{4, 6, 8, 10\}$. Students who become workers later in life are displayed in yellow, while students who will be teachers or managers are shown in light-brown and dark-brown respectively. The figure confirms several of the equilibrium properties. For example, students with higher values of a match with teachers with higher values of k, as we described in Lemma 4. Further, the figure shows the educational gap between workers and managers/teachers that we described in Proposition 3.

Figure 2 also provides insight in the occupational choices of individuals. Consider for example the individuals with communication skill $n_s = 10$. Although these individuals are the most able communicators in the economy, they will become workers later in life if their initial cognitive ability is low, as being a manager or a teacher requires sufficient skill along both dimensions. The set of future managers divides the set of future teachers into two.[15] Individuals with relatively low initial cognitive skill become teachers who will educate future workers, while individuals with relatively high initial cognitive skill will become teachers teaching future managers and future teachers. The gap again reflects the discontinuity in the demand for teachers with different cognitive skills: students who are future managers and teachers invest discretely more in education than students who are future workers.
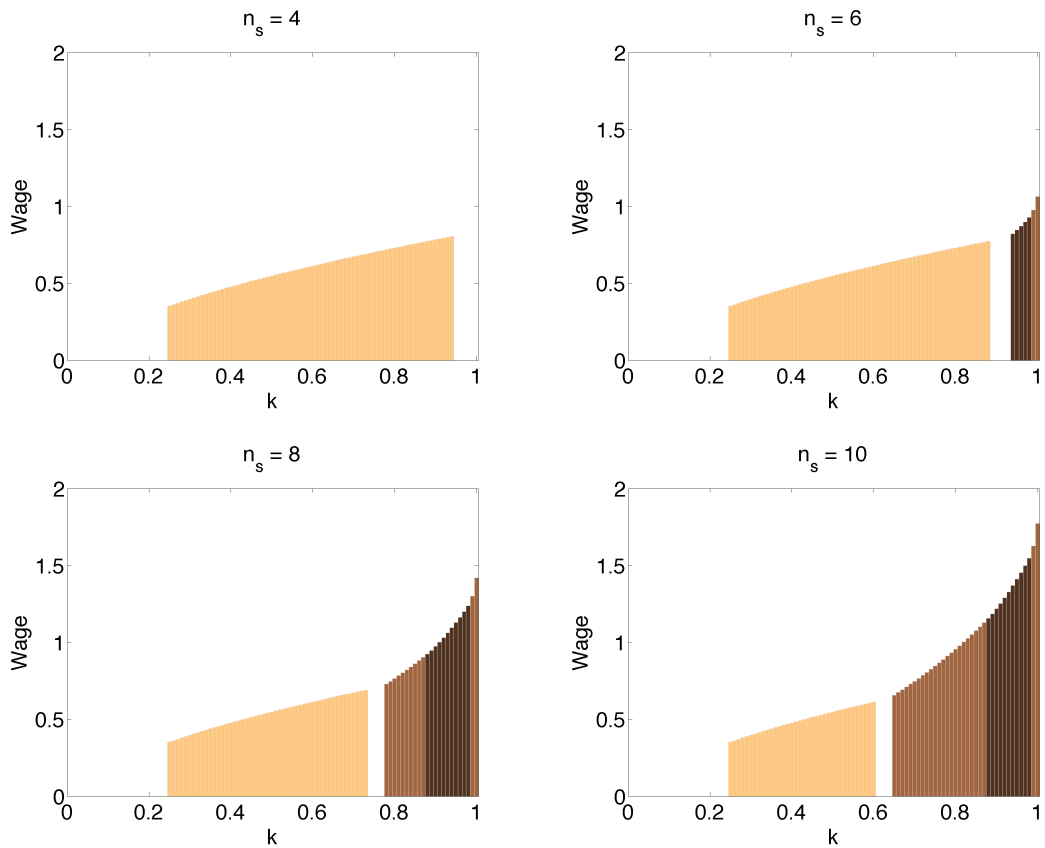
Figure 3 displays the wages earned in the labor market. More precisely, it shows the wage $\omega\left(n, k\right)$ for an adult of type-(n, k), again using yellow to indicate workers and brown to indicate managers / teachers. As discussed in Section 3, wages are increasing in cognitive skill k. In particular, they are concavely increasing for workers and convexly increasing for managers and teachers.[16] In particular, close comparison of the four panels confirms that a worker's wage does not vary with n, while the wages of managers and teachers are linearly increasing in n.[17] Note further that the worker with the highest wages may earn more than the least-paid managers and teachers.

Figure 3 further shows that certain combinations of n and k are absent in the labor market. For example, no adult has a very low value of k since everyone invests at least a certain

---

[15]When using a discrete grid, indifference between two occupations may arise for certain values of a_s in order to satisfy the aggregate demand/supply conditions. The measure of individuals that experiences indifference tends to go to zero when the number of grid points increases.
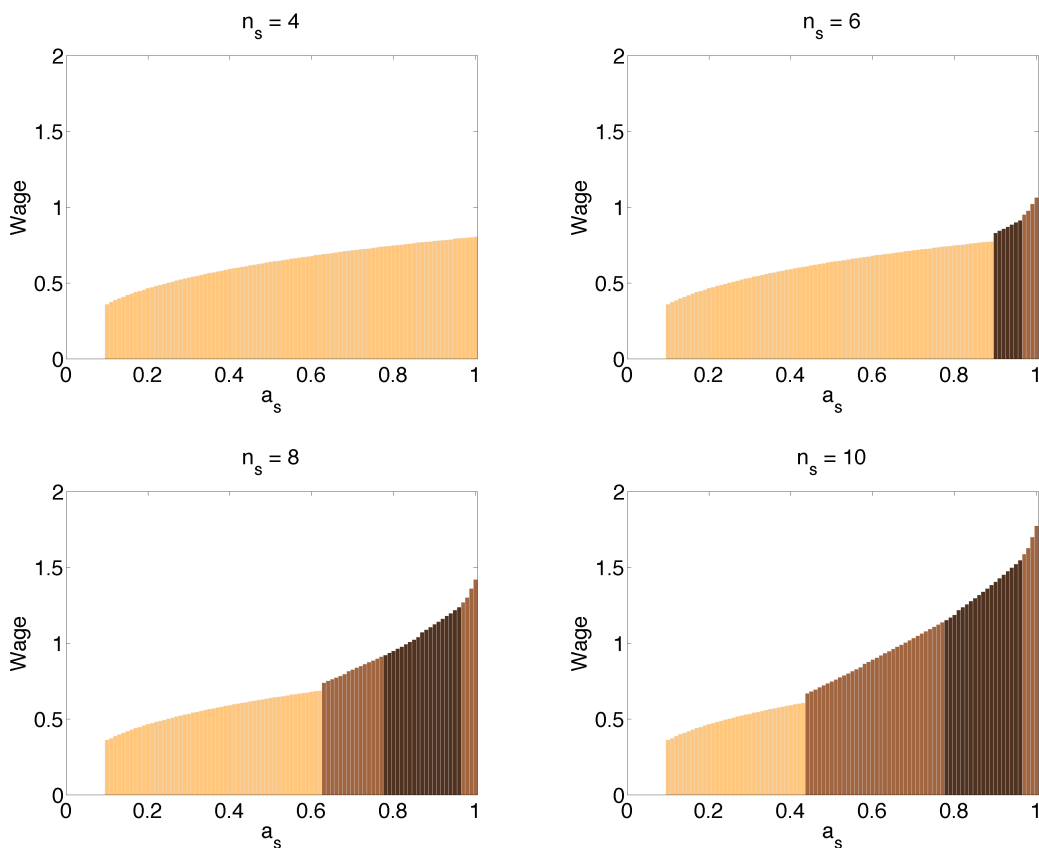
[16]However, the relationship between wages and $\log k$ is convex for all occupations, as predicted by Erlinger et al. (2014). Figure 3 also seems to confirm their prediction that, for our parameter values, the partial derivative of the wage with respect to cognitive skill tends to infinity for $k \to 1$.

[17]In this sense, our model is consistent with work by Cattan (2014), who argues that allowing for a differential impact of (both cognitive and non-cognitive) skills across occupations is important for understanding individuals' wages.

The wage $\omega\,(n,k)$ of an adult with communication skill n and cognitive skill k. Yellow adults are workers, light-brown adults are teachers, and dark-brown adults are managers.

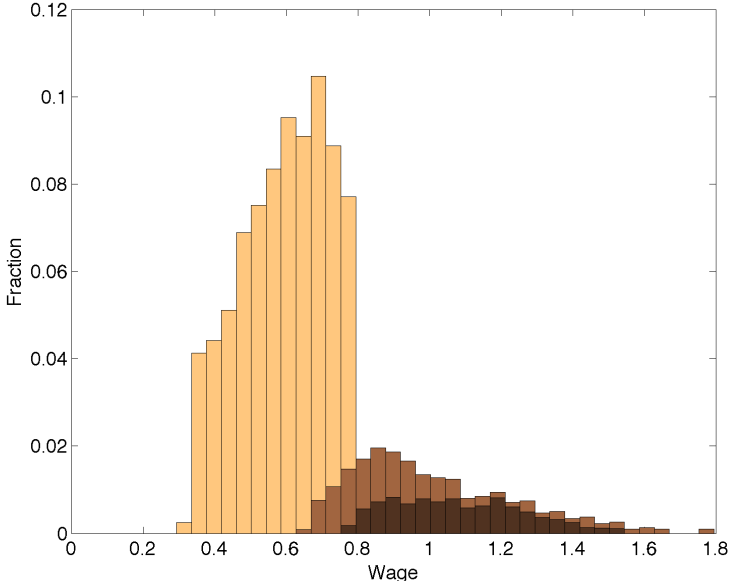Figure 3: Wages as function of adult cognitive skill

The future wage $\omega\left(n_s, \sqrt{a_s \rho\left(n_s, a_s\right)}\right)$ of a student with communication skill n_s and initial cognitive skill a_s. Yellow students become workers later in life, light-brown students become teachers, and dark-brown students become managers.

Figure 4: Wages as function of initial cognitive skill

amount in education. Similarly, low values of n combined with very high values of k do not arise, since students with low values of n realize that they will never become managers or teachers in the second period. Therefore, they are generally not willing to attain very high levels of education. Overall, education choices cause the correlation between cognitive skill and communication skill to be positive in the adult population, even though they were uncorrelated at birth. Finally, the figure clearly shows the educational gap between workers and managers/teachers as an empty band between the two corresponding sets of adult types.

Figure 4 combines figure 2 and 3 to show an individual's wage $\omega\left(n_s, \sqrt{a_s \rho\left(n_s, a_s\right)}\right)$ as a function of communication skill n_s and initial cognitive ability a_s. For each n, a discontinuity can be observed between the wage of the future worker with the highest a_s and the wage of

the future teacher / manager with the lowest a_s. This discontinuity once again reflects the gap in the education choice: future managers and teachers must be compensated for the fact that they choose a discretely better teacher than future workers and therefore pay a discretely higher tuition.[18]

Cross-sectional wage density. Yellow adults are workers, light-brown adults are teachers, and dark-brown adults are managers.

Figure 5: Labor market wage density

Figure 5 shows the cross-sectional wage density. One feature stands out: in line with empirical evidence from real labor markets, the simulated wage density is asymmetric with a short left tail and a long right tail. It is important to stress that this occurs even though we assumed a bivariate uniform distribution for a and n. Introducing any asymmetry in this distribution would only further strengthen this result. Note that workers, even with very high cognitive skills, do not receive the highest wages. The highest wages go to managers or teachers with strong communication and cognitive skills. On average, managers earn the most of all three occupations.[19] It turns out that heterogeneity in communication skills is important for gener-

---

[18]Of course, the magnitude of the discontinuity is relatively small compared to the earnings differences among teachers, as the cost of a better teacher is shared by n_t students.

[19]Nevertheless, the highest wage earner is a teacher, which may seem counterfactual. Clearly, our model is very stylized and does not capture many important features of real-life education and labor markets. For

ating the right skewness in the wage distribution.[20]

Finally, we simulated one comparative static with the model. We extend the support of communication skill by increasing $\overline{n}$. As a result of this change, the equilibrium number of managers and teachers falls and the skewness of log wages increases. This result is reminiscent of Rosen (1982) as well as Garicano and Rossi-Hansberg's (2006) concern about advances in communication technology.

5 Related literature

In addition to what we have discussed in the introduction, our model is related to several strands of literature. First, we contribute to the recent economics literature on cognitive and non-cognitive skills. See Borghans et al. (2008) and Almlund et al. (2011) for recent surveys. While our model is stylized in the sense that we only consider one non-cognitive skill, it provides the first attempt to explicitly model how non-cognitive factors can affect matching and occupation choice in the labor and education markets.

Second, our idea that task assignment should be based on comparative advantage dates back to Ricardo. Roy (1951) was the first to apply this concept to occupation choice based on occupation-specific skills. These ideas have been formalized and extended by various authors, with Sattinger (1975) being an early example. In Sattinger's model, however, the two sides of the market are exogenously determined. In that sense, our model is closer to contributions by Lucas (1978), Rosen (1978, 1982), Garicano (2000), and Garicano and Rossi-Hansberg (2006), in which agents endogenously choose their roles within the firm.

Third, as discussed in the introduction, there is a large organizational behavior literature on leadership in labor markets. Lazear (2012) surveys this literature, discusses economic models of leadership and presents one of his own. However, these models do not study task assignments, multifactor and multisector matching in a unified framework as we do here.

Fourth, our model predictions are in line with the empirical literature on the span of

---

example, firms often have more than two levels of hierarchy, creating scope for the manager at the top to indirectly manage much more than n workers. In the education market, we do not consider the non-pecuniary considerations are important for both teachers and students, and schools are often not operated on a for-profit basis, while in our model all participants maximize their expected monetary payoff. In addition, education markets are heavily regulated (tuition, for example), but we abstract from this. The goal of this paper, therefore, is not to provide a realistic account of how the compensation of the most able individual is determined in the market, but rather to provide a tractable framework to study the interaction among multidimensional matching, occupation choice and educational investment.

[20]For example, we simulated a model with a fixed $n = 5$ for all individuals and did not obtain a long right tail in the wage distribution.

control and the wage distribution. For example, Smeets and Warzynski (2008) show on the basis of a survey data that wages and bonuses of managers are increasing in the number of workers they supervise (i.e., span of control). They also show that individuals with higher communication skills are more likely to be managers. Furthermore, using survey data from a panel of more than 300 large U.S. companies over the period 1986-1999, Rajan and Wulf (2006) find a simultaneous increase in the span of control of CEOs and wage inequality.

As shown in the simulation, our model can generate a non-monotonic wage distribution that is skewed to the right, which is consistent with empirical regularities summarized in Neal and Rosen (2000). Sattinger (1975), Rosen (1978, 1982) and Waldman (1984) have used one-factor models to show that task assignment can generate right tail skewness in the earnings distribution.[21] These papers, however, do not explore whether the predicted earnings distributions are qualitatively consistent with the entire earnings distribution (roughly lognormal with a fat right tail).[22] Acemoglu and Autor (2011) argues that task assignments and skill biased technical change are needed to explain the recent evolution of the US labor market.

In addition, students sort in both dimensions of skills, so students with different combination of communication skill and initial cognitive skill are enrolled in the same school. Thus, our model provides an explanation to the puzzle why high and low cognitive ability students are in the same school.[23]

Fifth, multidimensional skills has been considered in a few other papers. In Papageorgiou (2013), each worker has two-dimensional skills which correspond to their productivity in one of two occupations and can be correlated. There are both information frictions and search frictions: workers have to learn their skills over time through employment, and workers have to search for employers, one occupation a time. While employed, workers observe their outputs, update their beliefs about their skills, and decide whether to switch occupations. He finds that his model predictions are consistent with the empirical patterns of occupational mobility documented in the literature (e.g., Kambourov and Manovskii, 2008). His analysis builds on the earlier work by Moscarini (2005) which in turn draws insight from the seminal work by Jovanovic (1979). Eeckhout and Weng (2010) consider a similar setup but abstract from search frictions and assume that the rate of learning could differ across occupations.

---

[21]See Gibbons and Waldman (1999) for a survey, and Gabaix and Landier (2008) for a more recent contribution.

[22]Li (2012) shows how a model of task assignment can approximate the fat right tail and the evolution of empirical wage distribution by introducing Pareto learning.

[23]A common answer in the literature is peer effects. In their surveys on peer effects among college students, however, Epple and Romano (2011) and Sacerdote (2011) do not find quantitatively large academic student peer effects for students with low cognitive skills.

They prove that supermodularity in production is necessary and sufficient for the equilibrium to exhibit positive assortative matching. Our model differs from aforementioned models in several aspects. First, in our model both sides of the match (managers and workers) are drawn from the same pool of the agents, while in their models one side of the match (firms or occupations) is exogenously given. Second, agents in different tasks/occupations in our model have to form teams in order to produce output, and thus our model features the commonly seen one-to-many matching. In contrast, agents in different occupations in these papers do not interact in production. Third, our model does not allow learning but we have an education market to allow individuals to accumulate cognitive skills. This allows us to form predictions regarding the education choice of agents with different combinations of cognitive and communication skills.

Finally, there are several papers which show the equivalence between a social planner's linear programming problem and frictionless matching. For example, Chiappori, McCann, Nesheim (2010) show that the frictionless multifactor marriage matching model is in this class.[24] Prescott and Townsend (2006) show that a one-sector frictionless matching labor model with one-factor matching between workers and firms, organizational design, occupational choice and moral hazard within firms is also in the same class. We extend these equivalences to a frictionless multisector multifactor many-to-one matching framework with occupational choice.

6 Extensions

In this section, we will briefly discuss two variations of the baseline model. First, we will describe a more general team production function and argue that all our quantitative results remain valid. Second, we will consider an alternative education process which does not involve teachers, as common in the literature.

6.1 General Team Production Function

It follows from (2) and our earlier analysis that if an individual i of type (n_i, k_i) spends θ_i units of time on production and another individual j of type (n_j, k_j) spends ψ_j units of time

---

[24]The use of linear programming to analyze matching problems goes back to Shapley and Shubik (1972) and Gretsky, Ostroy and Zame (1992). See also Li and Suen (2001), and McCann and Trokhimtchouk (2010) for a single sector, unidimensional model with endogenous occupation choice.

on helping, then they jointly produce

$$\widetilde{Y}\left(\theta_i, \psi_j; n_i, k_i, n_j, k_j\right) = \sqrt{k_i k_j} \min\left\{\theta_i, n_j \psi_j\right\}.$$

While this specific functional form improves analytical tractability, it is not crucial for our results. For example, we can assume

$$\widetilde{Y}\left(\theta_i, \psi_j; n_i, k_i, n_j, k_j\right) = R_1\left(k_i, k_j\right) \cdot R_2\left(\theta_i, \psi_j, n_j\right)$$

where both R_1 and R_1 have constant elasticity of substitution (CES). That is,

$$R_1\left(k_i, k_j\right) = \left[\delta_1 k_i^{1/s} + (1 - \delta_1) k_j^{1/s}\right]^s,$$

and

$$R_2\left(\theta_i, \psi_j, n_j\right) = \left[\delta_2 \theta_i^{\frac{\xi-1}{\xi}} + (1 - \delta_2)\left(n_j \psi_j\right)^{\frac{\xi-1}{\xi}}\right]^{\frac{\xi}{\xi-1}},$$

where $s \geq 1$ is a measure of supermodularity between k_i and k_j, while $\xi \geq 0$ is the elasticity of substitution between time allocated to task P and time allocated to task C. As usual, the weights δ_1 and δ_2 are between 0 and 1. Note that our baseline specification corresponds to the special case with $s = \infty$, $\xi = 0$ and $\delta_1 = \delta_2 = 1/2$.

In this case, individuals will continue to specialize and choose their occupation according to the cutoff rule. Moreover, it follows from the analysis of Eeckhout and Kircher (2012) that the labor market equilibrium will exhibit positive assortative matching if and only if

$$s \geq \frac{1}{1 - \xi}.$$

Under this more general specification, however, a type-(n, k) manager will not necessarily hire n workers, in contrast to our baseline model. The difference arises because the new production function introduces a meaningful tradeoff between the intensive margin and the extensive margin in the time allocation of the manager.

6.2   Education Market

In our baseline model, we assume that human capital k is accumulated according to a team production function similar to the one in the labor market. Alternatively, as common in the literature, we can assume that individuals accumulate human capital by themselves, without

(explicitly) involving teachers. For example,

$$k_i = E\left(a_i, \tau\right)$$

where τ is the educational investment, and the function E is assumed to be increasing in both arguments. If E is sufficiently concave in τ, then there exists a unique optimal education level for each type of individual. With this alternative specification, students who will become managers as adults will invest discretely more in education than students who will become workers in the future, as in our baseline model.

7   Conclusion

This paper presents a tractable multi-sector matching model of cognitive and communication skills. We analyze the matching patterns in the education and labor markets, and derive a large number of empirical predictions. In order to keep the analysis as transparent and tractable as possible, we make several simplifying assumptions. Our model is therefore best viewed as an initial exploration of how cognitive and communication skills may interact in the various environments.

The model can be extended by relaxing some of our assumptions. For example, additional non-cognitive factors could be considered. The introduction of two-sided communication costs or interaction among workers may help generate richer predictions regarding matching in the labor market. In addition, one could allow individuals to accumulate not only cognitive skills but also communication skills. It would also be interesting to consider more heterogeneity on the firm side, e.g. with respect to firm resources or price of output. Such heterogeneity may interact with the communication skill of a manager, generating different spans of control for managers with the same n. Finally, more elaborate task structures could be considered, e.g. with a larger number of tasks, with an endogenous division of tasks to workers, or with a multi-level hierarchy of tasks. The last extension may yield multi-level hierarchies among a firm's employees as well.

Appendix: Omitted Proofs

Proof of Lemma 4. Note that we can write

$$\phi(k) = \max_{k'} \sqrt{kk'} - \omega(k').$$

The necessary first-order and second-order conditions are

$$\frac{1}{2}\sqrt{\frac{k}{k'}} - \omega'(k') = 0 \text{ and } -\frac{1}{4k'}\sqrt{\frac{k}{k'}} - \omega''(k') \leq 0,$$

which imply that

$$\omega''(k') \geq -\frac{1}{2k'}\omega'(k').$$

Now first suppose student (n_s, a_s) will become a worker. Define

$$\Pi(n_s, a_s; n_t, k_t) \equiv \omega\left(\sqrt{a_s k_t}\right) - \tau(k_t).$$

Then we have

$$
\begin{aligned}
\frac{\partial^2 \Pi(n_s, a_s; n_t, k_t)}{\partial a_s \partial k_t} &= \frac{1}{4}\omega''\left(\sqrt{a_s k_t}\right) + \frac{1}{4\sqrt{a_s k_t}}\omega'\left(\sqrt{a_s k_t}\right) \\
&\geq -\frac{1}{4}\frac{1}{2\sqrt{a_s k_t}}\omega'\left(\sqrt{a_s k_t}\right) + \frac{1}{4\sqrt{a_s k_t}}\omega'\left(\sqrt{a_s k_t}\right) \\
&= \frac{1}{8\sqrt{a_s k_t}}\omega'\left(\sqrt{a_s k_t}\right) \\
&> 0
\end{aligned}
$$

Therefore, $\Pi(n_s, a_s; n_t, k_t)$ is supermodular in k_t and a_s, which implies that the equilibrium matching $k_t = \rho(a_s, n_s)$ is increasing in a_s. On the other hand,

$$\frac{\partial^2 \Pi(n_s, a_s; n_t, k_t)}{\partial n_s \partial k_t} = 0$$

so the equilibrium matching $k_t = \rho(a_s, n_s)$ does not depend on n_s.

Next consider the case where student (n_s, a_s) chooses to become a manager. Note that we can write

$$\omega(k) = \max_{k'} \sqrt{kk'} - \phi(k').$$

The necessary first-order and second-order conditions imply that

$$\phi''(k') \geq -\frac{1}{2k'}\phi'(k')$$

Let's define

$$\widehat{\Pi}(n_s, a_s; n_t, k_t) \equiv n_s \phi\left(\sqrt{a_s k_t}\right) - \tau(k_t)$$

Then we have

$$
\begin{aligned}
\frac{\partial^2 \widehat{\Pi}(n_s, a_s; n_t, k_t)}{\partial a_s \partial k_t} &= n_s \frac{1}{4}\phi''\left(\sqrt{a_s k_t}\right) + n_s \frac{1}{4\sqrt{a_s k_t}}\phi'\left(\sqrt{a_s k_t}\right) \\
&\geq -\frac{1}{4}\frac{1}{2\sqrt{a_s k_t}}\phi'\left(\sqrt{a_s k_t}\right) + \frac{1}{4\sqrt{a_s k_t}}\phi'\left(\sqrt{a_s k_t}\right) \\
&= \frac{1}{8\sqrt{a_s k_t}}\phi'\left(\sqrt{a_s k_t}\right) \\
&> 0
\end{aligned}
$$

and

$$\frac{\partial^2 \widehat{\Pi}(n_s, a_s; n_t, k_t)}{\partial n_s \partial k_t} = \frac{1}{2}\sqrt{\frac{a_s}{k_t}}\phi'\left(\sqrt{a_s k_t}\right) > 0$$

Therefore, $\widehat{\Pi}(n_s, a_s; n_t, k_t)$ is supermodular in k_t and a_s, and supermodular in k_t and n_s. Thus, the equilibrium matching $k_t = \rho(a_s, n_s)$ is increasing in both a_s and n_s. ∎

Proof of Proposition 3. Consider a type-(a_s, n_s) student whose equilibrium school choice is $k_t = \rho(n_s, a_s)$. If this student becomes a manager eventually, the optimal school choice k_t must satisfy

$$\frac{1}{2}n_s \phi'\left(\sqrt{a_s k_t}\right)\sqrt{\frac{a_s}{k_t}} + \frac{n_s}{n_s + 1}\frac{1}{2}\sqrt{\frac{a_s}{k_t}} - \tau'(k_t) = 0. \tag{18}$$

In contrast, if this student eventually becomes a worker, k_t must satisfy

$$\frac{1}{2}\omega'\left(\sqrt{a_s k_t}\right)\sqrt{\frac{a_s}{k_t}} + \frac{n_s}{n_s + 1}\frac{1}{2}\sqrt{\frac{a_s}{k_t}} - \tau'(k_t) = 0. \tag{19}$$

Now suppose $n_s = \widehat{n}(\sqrt{a_s k_t})$. If this student indeed chooses education level k_t, then k_t must solve both (18) and (19), which implies that

$$\widehat{n}(\sqrt{a_s k_t})\phi'\left(\sqrt{a_s k_t}\right) = \omega'\left(\sqrt{a_s k_t}\right). \tag{20}$$

31

Note that, by definition of $\widehat{n}(k)$, we have

$$\widehat{n}'(k) = \frac{\omega'(k)\,\phi(k) - \omega(k)\,\phi'(k)}{[\phi(k)]^2} = \frac{\omega'(k) - \widehat{n}(k)\,\phi'(k)}{\phi(k)}$$

So we have

$$\frac{\omega'(k)}{\phi'(k)} = \widehat{n}(k) + \frac{\phi(k)\,\widehat{n}'(k)}{\phi'(k)}$$

Therefore, condition (20) reduces to $\widehat{n}'\left(\sqrt{a_s k_t}\right) = 0$, a contradiction. ∎

References

[1] Acemoglu, D. and D. Autor (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics*, vol. 4B, Amsterdam, Elsevier, 1043-1172.

[2] Almlund, M., A. Duckworth, J. Heckman, and T. Kautz (2011): "Personality Psychology and Economics," *Handbook of the Economics of Education*, vol. 4, chapter 1, New York, Elsevier.

[3] Becker, G. S. and K. M. Murphy (1992). "The Division of Labor, Coordination Costs, and Knowledge," *The Quarterly Journal of Economics* 107 (4): 1137-1160.

[4] Borghans, L., A. L. Duckworth, J. J. Heckman, and B.T. Weel (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources* 43(4): 972-1059.

[5] Cattan, S. (2014): "Psychological Traits and the Gender Wage Gap," Working Paper, Institute for Fiscal Studies.

[6] Chiappori, P. A., R. J. McCann, and L. Nesheim (2010): "Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology, and Uniqueness," *Economic Theory*, 42(2), 317-354.

[7] Dessein, W. and T. Santos (2006): "Adaptive Organizations," *Journal of Political Economy* 114 (5): 956-995.

[8] Duffy, E. A. and I. Goldberg (1998): *Crafting a Class: College Admissions and Financial Aid, 1955-1994.* Princeton University Press.

[9] Eeckhout, J., and P. Kircher (2012): "Assortative Matching with Large Firms: Span of Control over More versus Better Workers," LSE working paper.

[10] Epple, D., and R.E. Romano (2011): "Peer Effects in Education: A Survey of the Theory and Evidence," In *Handbook of Social Economics* (11), Elsevier B.V., 1053-1163.

[11] Erlinger, A., R.J. McCann, X. Shi, A. Siow, and R. Wolthoff (2014): "Academic Wages and Pyramid Schemes: A Mathematical Model," Working paper, University of Toronto.

[12] Gabaix, X., and A. Landier (2008): "Why Has CEO Pay Increased So Much?"*Quarterly Journal of Economics* 123(1): 49-100.

[13] Garicano, Luis (2000): "Hierarchies and the Organization of Knowledge in Production,"*Journal of Political Economy* 108 (5): 874-904.

[14] Garicano, L., and E. Rossi-Hansberg (2004): "Inequality and the Organization of Knowledge." *American Economic Review* 94(2): 197-202.

[15] Garicano, L., and E. Rossi-Hansberg (2006): "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics* 121(6): 1383-1435.

[16] Gibbons, R., and M. Waldman (1999): "Careers in Organizations: Theory and Evidence," *Handbook of Labor Economics* 3: 2373–2437.

[17] Gretsky, N.E., J.M. Ostroy and W.R. Zame (1992): "The Nonatomic Assignment Model," *Economic Theory* 2: 103–127.

[18] Heckman, J. J., J. Stixrud and S. Urzua (2006): "The Effects Of Cognitive and Noncognitive Abilities On Labor Market Outcomes and Social Behavior," *Journal of Labor Economics* 24(3), 411-482.

[19] Heckman, J. J., R Pinto and P. Savelyev (2012): "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," NBER working paper no. 18581.

[20] Jovanovic, B. (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87(5), 972-990.

[21] Kambourov, G. and I. Manovskii (2008): "Rising Occupational and Industry Mobility in the United States: 1968-1997," *International Economic Review* 49(1), 41-79.

[22] Klitgaard, R. (1985): *Choosing Elites*. Basic Books, New York, NY

[23] Kremer, M. (1993): "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, 551-575.

[24] Kuhn, P, and C. Weinberger (2005): "Leadership Skills and Wages," *Journal of Labor Economics* 23(3): 395-436.

[25] Lazear, Edward P. (2012): "Leadership: A Personnel Economics Approach."*Labour Economics* 19, Issue 1, 92–101.

34

[26] Lazear, E. P., K. L. Shaw, and C. T. Stanton (2012): "The Value of Bosses."NBER Working Paper # 18317.

[27] Li H. and W. Suen (2001): "Technological Change and Labor Market Segregation," Working Paper, University of Toronto.

[28] Li, J. (2012): "A Theory of Wage Distribution and Wage Dynamics with Assignment and Pareto Learning," Working Paper, Northwestern University.

[29] Lopes de Melo, R. (2013): "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence," Working Paper, University of Chicago.

[30] Lucas, R. E. (1978): "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9(2): 508-523.

[31] McCann R.J., and M. Trokhimtchouk (2010): "Optimal Partition of A Large Labor Force into Working Pairs," *Economic Theory* 42, 375-395.

[32] Mortensen, D. (2003): *Wage Dispersion: Why Similar Workers Paid Differently*, The MIT Press.

[33] Moscarini, G. (2005), "Job Matching and the Wage Distribution," *Econometrica* 73(2), 481-516.

[34] Neal, D., and S. Rosen (2000): "Theories of the Distribution of Earnings," *Handbook of Income Distribution* 1: 379-427.

[35] Papageorgiou, T. (2013): "Learning Your Comparative Advantages," *Review of Economic Studies*, forthcoming.

[36] Prescott, E. S. and R. M. Townsend (2006): "Firms as Clubs in Walrasian Markets with Private Information."*Journal of Political Economy* 114 (4): 644-671.

[37] Rajan, R.G., and J. Wulf. (2006): "The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies." *Review of Economics and Statistics* 88(4): 759-773.

[38] Rosen, S. (1978): "Substitution and Division of Labour," *Economica* 45 (179): 235-250.

[39] Rosen, S. (1982): "Authority, Control, and the Distribution of Earnings." *Bell Journal of Economics* 13(2): 311-323.

[40] Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxord Economic Papers* (New Series), 3, 135-146.

[41] Sacerdote, B. (2011): "Peer effects in education: How might they work, how big are they and how much do we know thus far?"*Handbook of the Economics of Education*, 3, 249-277.

[42] Sattinger, M. (1975): "Comparative Advantage and the Distributions of Earnings and Abilities," *Econometrica*: 455-468.

[43] Shapley, L.S. and M. Shubik (1972): "The Assignment Game I: The Core," *International Journal of Game Theory* 1: 111–130.

[44] Smeets, V., and F. Warzynski (2008): "Too Many Theories, Too Few Facts? What the Data Tell Us about the Link between Span of Control, Compensation and Career Dynamics," *Labour Economics* 15(4): 687-703.

[45] Waldman, M. (1984): "Worker Allocation, Hierarchies and the Wage Distribution," *Review of Economic Studies* 51(1): 95-109.