Academic wages, singularities, phase transitions and pyramid schemes

Robert J. McCann*

Abstract. In this lecture we introduce a mathematical model which couples the education and labor markets, in which steady-steady competitive equilibria turn out to be characterized as the solutions to an infinite-dimensional linear program and its dual. In joint work with Erlinger, Shi, Siow and Wolthoff, we use ideas from optimal transport to analyze this program, and discover the formation of a pyramid-like structure with the potential to produce a phase transition separating singular from non-singular wage gradients.

Wages are determined by supply and demand. In a steady-state economy, individuals will choose a profession, such as worker, manager, or teacher, depending on their skills and market conditions. But these skills are determined in part by the education market. Some individuals participate in the education market twice, eventually marketing as teachers the skills they acquired as students. When the heterogeneity amongst student skills is large, so that it can be modeled as a continuum, this feedback mechanism has the potential to produce larger and larger wages for the few most highly skilled individuals at the top of the market. We analyze this phenomena using the aforementioned model. We show that a competitive equilibrium exists, and it displays a phase transition from bounded to unbounded wage gradients, depending on whether or not the impact of each teacher increases or decreases as we pass through successive generations of their students. We specify criteria under which this equilibrium will be unique, and under which the educational matching will be positive assortative. The latter turns out to depend on convexity of the equilibrium wages as a function of ability, suitably parameterized.

Mathematics Subject Classification (2010). Primary 91B66; Secondary 49N15, 35Q91, 91B68, 91B69

Keywords. steady state, competitive equilibrium, multisector matching, heterogeneous types, education, labor, convex wages, singularity, phase transition, pyramid scheme, optimal transport, infinite-dimensional linear program

---

*The author thanks Gary Becker and Yann Brenier for key references, plus Rosemonde Lareau-Dussault and his collaborators Alice Erlinger, Xianwen Shi, Aloysius Siow and Ronald Wolthoff for many stimulating interactions. He acknowledges partial support of his research by Natural Sciences and Engineering Research Council of Canada Grant 217006-08.

1. Introduction

The last half century has seen much fruitful interaction between economics and mathematics. Still, the relationship between these two subjects is far less developed than the long standing affair between mathematics and physics. This is good news for mathematicians, in the sense that much work remains to be done: economics provides a ready source of interesting mathematical problems, so far only modestly tapped, as well as an area of application where mathematical developments again have a chance to prove transformative, as they did during the development of the theories of gravitation and quantum mechanics in the last century, and more recently in areas like statistical physics and string theory.

For mathematicians working in optimal transportation, it has been a source of considerable satisfaction to discover that their work has found diverse applications in economics; see [6] [7] [9] and [3] for examples. My work with Figalli and Kim [11] in particular suggests that curvature may have a heretofore unanticipated role to play in economic theory. It is also the case that economists have exploited ideas from optimal transportation in surprising ways [14] [25], and in some cases have been ahead of mathematicians in anticipating significant developments, such as the gradient flow framework discussed in Sonnenschein's myopic price dynamics [28], some twenty-five years before it was linked to the heat equation by Jordan, Kinderlehrer and Otto [19]. (The triangle with physics is complete, in the sense that Cullen and Purser were using Hamiltonian dynamics in the same energy landscape as Sonnenschein to study atmospheric and oceanic flows [8].) All this is perhaps less surprising given the fact that one of the cornerstones of optimal transport — the Kantorovich-Koopmans duality theorem — lead its mathematician and physicist-trained inventors to share the 1975 Nobel Memorial Prize in Economics. A few years before that, future Nobel Laureate Lloyd Shapley had worked out the fundamental results concerning stable matchings in the non-transferable and transferable utility settings with Gale [13] and with Shubik [27], respectively. As Shapley and Shubik discovered, the transferable utility version of this problem can also be reformulated as a discrete optimal transport problem. This discovery was generalized and extended to a continuum setting by Gretzky, Ostroy and Zame [18].

In the present synopsis I will sketch the results of a joint work with three economist colleagues at the University of Toronto, and one former PhD student, which draws inspiration from the foregoing. It concerns a matching model coupling the education and labor sectors. For precise statements and proofs of all the ideas discussed here, see our joint paper [10].

As is often the case in economic models, the problem is to understand how supply equilibrates with demand to determine prices — or in this case wages – in each of these markets. What is supplied and demanded in these markets are skills; we focus primarily on cognitive skills for simplicity. The phenomenon of interest to us is whether, in the limit of a large population displaying a bounded range of skills, competition may lead to wages which display singularities: for example, does the the ratio of the highest to the average salary tend to a finite or an infinite number, as the size of the population tends to infinity? We investigate this question in a competitive equilibrium model, which means individuals choose those teachers, occupations and partners who reward them most generously, under the assumption that such decisions are made at the level of individuals (or groups) too small to affect market conditions such as the wage profile $v(k)$ as a function of skill level $k \in [0, \bar{a}[$. Skills can be continuously distributed in this interval, as a reflection of the large size of the population.

The education market plays a special role in our study. The output produced by this market is the enhancement of salable skills, taking initial student and teacher skills as its input. However, some participants match twice in the education market, first as students, and later as teachers. This creates a feedback mechanism which — depending the effectiveness of the educational technology — has the potential to create a pyramid, in which there is greater and greater demand for the most skilled teachers, who occupy positions closer and closer to the apex. Even if we assume that wages outside the education market are bounded, in a steady-state model the wage which a highly skilled teacher can command depends on the potential future earnings of their students, some of whom will be teachers, who may have many generations of students, whose wages must be determined in part by the internal dynamics of the education market insofar as these students again will again have the choice to become teachers — and will do so provided they can find other students willing to pay them sufficient tuition. As adults on the labor market, they also have alternatives to teaching: namely they could engage in production by working or managing a team consisting of N' workers and one manager, whose output is a known function $N'b_L$ of the team members' skills.

Our model depends on $c \geq 0$ and positive parameters θ, θ', N, N' and $\bar{a}$ satisfying $\max\{\theta, \theta\} < 1 \leq N$. Here N represents the number of students each teacher can teach, and the extent θ to which she succeeds at transmitting her cognitive skills to her students. Similarly, N' represents the number of workers each manager can manage, and θ' represents the extent to which a manager's skills influences the quality of the work produced by those whom she manages. All skills lie in the interval $A = [0, \bar{a}[$ or its closure $\bar{A}$.

A worker of skill $a \in \bar{A}$ working with a manager of skill $k \in \bar{A}$ produces a continuously differentiable output $b_L((1-\theta')a + \theta' k)$ given by $b_L \in C^1(\bar{A})$. A student of skill a studying with a teacher of skill k becomes an adult of skill $z(a, k) = (1-\theta)a + \theta k$. The acquired skill z may have some external value $cb_E(z)$ to the student — in addition to the wage $v(z)$ it commands on the market. Both b_L and $b_E \in C^1(\bar{A})$ are assumed to be fixed uniformly convex increasing functions hereafter, satisfying

$$0 < \underline{b}_{E/L} = b_{E/L}(0), \tag{1}$$

$$0 < \underline{b}'_{E/L} = b'_{E/L}(0), \tag{2}$$

$$0 < \underline{b}''_{E/L} = \inf_{a \in A} b''_{E/L}(a), \tag{3}$$

where $\underline{b}''_{E/L}$ is defined as the largest constant for which $b_{E/L}(a) - \underline{b}''_{E/L}|a|^2/2$ is convex on $\bar{A}$. A typical example would be $b_E(a) = e^a = b_L(a)$ as in [22].

At each period of our model, students are born into the population with cognitive skills distributed randomly throughout the closure $\bar{A}$ of an interval $A = [0, \bar{a}[\subset$

R according to some Borel probability distribution $\alpha \geq 0$. They live only for two periods. During the first period, each student $a \in \bar{A}$ seeks to enhance his cognitive skill by studying with whichever teacher $k \in \bar{A}$ provides the best educational value to him. During the second period, the student becomes an adult armed with skill $z = z(a, k) = (1 - \theta)a + \theta k$, who then enters the workforce as a worker, manager or teacher earning a wage $v(z)$, whose steady-state value we seek to determine.

Although we cannot find $v(z)$ explicitly, we are able to characterize it as the solution of a variational problem: an infinite-dimensional linear program whose non-standard form complicates its analysis considerably. What has been achieved with Erlinger, Shi, Siow and Wolthoff [10] is an analysis of the existence, uniqueness, and characteristics of the solutions to this linear program under suitable technical hypotheses. These include requiring the initial distribution α of student skills to satisfy a doubling condition at the top skill type $\bar{a}$, meaning there exists $C < \infty$ such that

$$\int_{[\bar{a}-2\Delta a, \bar{a}]} \alpha(da) \leq C \int_{[\bar{a}-\Delta a, \bar{a}]} \alpha(da) \tag{4}$$

for all $\Delta a > 0$. Under suitable assumptions, we deduce the wage profile $v(k)$ is strictly convex and increasing, but displays a phase transition from having bounded to unbounded gradients as the product $N\theta$ increases through 1. More precisely, if $N\theta > 1$, so that the net impact of a teacher increases as one passes from each generation of their students to the next, then $v'(a) \sim |\bar{a} - a|^{-\frac{\log \theta}{\log N} - 1}$ as $a \to \bar{a}$ (unless an even stranger pathology occurs — see Theorem 6.1). Since this singularity is integrable, our analysis suggests $\lim_{a \to \bar{a}} v(a)$ is finite.

Along the way, we establish various conclusions about the behavior of the model in different parameter regimes, such as which ranges of skill types will be displayed by workers, managers and teachers, and who will match with whom in each sector (education and labor) of our market. Before describing our model and conclusions more precisely, let us mention some antecedents.

The role of teachers and the individual investment required to develop human capital has been examined in the context of a steady-growth model for specialization by Becker and Murphy, who also recognized the relevance of the long lineages of teachers which may form [2]. The economics of superstars had been analyzed before that by Rosen [26]; convexity of wages play a key role in his study, as they do in ours. The formation of finite-depth pyramids (or management layers) in the context of a labor market model has also been investigated by Garicano [16] and followed up with Rossi-Hansberg [17], though the absence of feedback makes their model quite different from ours. Another explanation for inflated levels of executive compensation has been proposed by Gabaix and Landier [12]. Finally, the possibility of allowing parameters such as θ, θ', N and N' to vary endogenously across the population to model heterogeneity of communication skills was a feature of our original four-author model [22], which we have chosen to suppress in the logarithmically reparameterized five-author sequel [10]. This suppression facilitates a more penetrating analysis of some phenomena of interest in their simplest form.

2. A competitive equilibrium model

Fix θ, θ', N, N' and $\bar{a}$ positive with $\max\{\theta, \theta'\} < 1 \leq N$ and $A = [0, \bar{[}$, a probability measure α on $\bar{A}$, and education and labor production functions $b_{E/L} \in C^1(\bar{A})$ satisfying (1)–(3). Set $K = [0, \bar{k}[= A$ and $c \geq 0$.

A competitive equilibrium requires the wage profile $v(k)$ as a function of skill level $k \in \bar{A}$ to be related in a certain way to other quantities which reflect the behavior of students and adults in our model. Educational decisions of such agents are captured by a probability measure $d\epsilon(a,k) \geq 0$ on $\bar{A}^2$, which represents the fraction of students with skill a who choose to study with a teacher of skill k, and vice versa. Labor market decisions are recorded by a measure $d\lambda(a,k) \geq 0$ on $\bar{A}^2$, representing the number of workers of skill a who choose to work with a manager of skill k, and vice versa. The net lifetime utility of a student of skill a will be denoted by $u(a)$.

The *support* of any (Borel) measure μ on $\mathbf{R}^m$ refers to the smallest closed set $\operatorname{Spt}\mu \subset \mathbf{R}^m$ carrying the full mass of μ. The *push-forward* of any measure μ on $\mathbf{R}^m$ through a Borel map $f : \mathbf{R}^m \longrightarrow \mathbf{R}^n$ refers to the measure $f_{\#}\mu$ assigning mass $\mu[f^{-1}(N)]$ to each set $N \subset \mathbf{R}^n$. Thus $\epsilon^1 = \pi^1_{\#}\epsilon$ and $\epsilon^2 = \pi^2_{\#}\epsilon$ denote the marginal projections of ϵ through the coordinate maps $\pi^i(x_1, x_2) = x_i$.

We say $\epsilon \geq 0$ and $\lambda \geq 0$ represent a *steady-state* for our model if

$$\epsilon^1 = \alpha \qquad \text{and} \tag{5}$$

$$\lambda^1 + \frac{1}{N'}\lambda^2 + \frac{1}{N}\epsilon^2 = z_{\#}\epsilon, \tag{6}$$

where $z(a,k) = (1-\theta)a + \theta k$ is the skill attained by a student a through studying with teacher k. Here the first identity requires the initial distribution of student skills to be given by α, while the second requires that the current distribution of (worker + manager + teacher) skills in the population will be reproduced at the next generation through education. This is the steady-state constraint. We denote the set of non-negative measures (ϵ, λ) satisfying (5)–(6) by $R(\alpha)$ — which of course depends also on N, N' and θ. Note that λ will not be a probability measure; rather its mass coincides with the fraction $(1 - \frac{1}{N})/(1 + \frac{1}{N'})$ of the adult population who choose to become workers.

We say a pair of payoffs $u, v : \bar{A} \longrightarrow]0, \infty]$ are *stable* if

$$u(a) + \frac{1}{N}v(k) \geq cb_E(z(a,k)) + v(z(a,k)) \qquad \text{and} \tag{7}$$

$$v(a) + \frac{1}{N'}v(k) \geq b_L((1-\theta')a + \theta' k) \qquad \text{on } \bar{A} \times \bar{K}, \text{ and} \tag{8}$$

$$\tfrac{N}{N-1}(u(k) - cb_E(k)) \geq v(k) \geq \frac{N'}{N'+1}b_L(k) \qquad > 0. \tag{9}$$

The wage constraint (8) reflects the stability of matchings in the labor sector. If the reverse inequality held, the output $N'b_L$ produced by N' adults of skill a and one of skill k would be sufficient to allow all $N'+1$ of them to improve their wages by abandoning their occupations to collaborate by forming N' new worker-manager pairs. Similarly, constraint (7) is a stability constraint on the education market, which ensures that no N students of ability a plus one teacher of ability k all have the incentive to abandon their institutions to form a school with each other. Together, these two constraints imply (9) at any point $a = k$ where v is finite; we have included it only to show the payoffs are positive and that v cannot diverge unless u does.

We must also specify in what class of functions the payoffs u, v must lie. Since we wish to allow for the possibility that the payoffs $u, v : A \longrightarrow [0, \infty[$ become unbounded at the upper endpoint $\bar{a}$ of the half-open interval A, we define the feasible set F_0 to consist of pairs $(u, v) = (u_0 + u_1, v_0 + v_1)$ satisfying (7)–(9) which differ from bounded continuous functions $u_0, v_0 \in C(\bar{A})$ by non-decreasing lower semicontinuous functions $u_1, v_1 : \bar{A} \longrightarrow [0, \infty]$ which are real-valued on A.

Together, a pair of stable payoffs $(u, v) \in F_0$ and steady-state matchings $(\epsilon, \lambda) \in R(\alpha)$ form a *competitive equilibrium* if they satisfy the budget constraint

$$\text{equality holds } \epsilon\text{-a.e. in (7), and } \lambda\text{-a.e. in (8).} \tag{10}$$

In other words, the productivity of λ-a.e. worker-manager team must be sufficient to the pay the worker's wage plus a fraction $\frac{1}{N'}$ of the manager's; similarly ϵ-a.e. student-teacher pair must generate future earnings v for the student, which together with any non-labor compensation cb_E for skills acquired through education, must be sufficient to leave utility $u(a)$ for the student after a fraction $\frac{1}{N}$ of his teacher's salary has been paid.

3. A variational approach

Since it is not obvious whether such equilibria exist or how to find them, we begin by recharacterizing them variationally. Consider the problem of minimizing the expected net utility $\alpha(u)$ over the population α of students:

$$LP_* := \inf_{(u,v) \in F_0} \int_{[0,\bar{a}]} u(a) d\alpha(a). \tag{11}$$

This is a linear minimization over the convex set of stable payoffs $(u, v) \in F_0$. As an infinite-dimensional linear program whose domain includes pairs of continuous functions on $\bar{A}$ satisfying two stability contraints, this problem has a linear programming dual, which turns out to be a maximization involving pairs of measures on $\bar{A}^2$:

$$LP^* := \max_{(\epsilon,\lambda) \in R(\alpha)} \int_{\bar{A} \times \bar{K}} [cb_\theta(a, k) d\epsilon(a, k) + \tilde{b}_{\theta'}(a, k) d\lambda(a, k)], \tag{12}$$

where $b_\theta(a, k) = b_E(z(a, k))$ and $\tilde{b}_{\theta'}(a, k) = b_L((1-\theta')a + \theta' k)$. It can be interpreted as a social planners problem, which is to maximize the production $c\epsilon(b_\theta) + \lambda(\tilde{b}_{\theta'})$

of the two sectors in question (education and labor) over steady-state measures $(\epsilon, \lambda) \in R(\alpha)$.

If F_0 consisted solely of continuous bounded functions $u, v \in C(\bar{A})$, it would be easy to see $LP_* \geq LP^*$ via the argument of the proposition below; equality would then follow from a standard application of the Fenchel-Rockafellar duality theorem. The fact that F_0 includes unbounded functions makes the 'obvious' inequality $LP_* \geq LP^*$ much more subtle to prove. It is for this purpose that we exploit the doubling condition (4) on α to establish $LP_* = LP^*$ [10].

We also assert that Shapley and Shubik's insight [27] extends from single-sector, single-stage matching problems to the current multisectorial steady-state setting:

Proposition 3.1 (Optima v. equilibria). *The pair* $(u, v) \in F_0$ *and* $(\epsilon, \lambda) \in R(\alpha)$ *constitutes a competive equilibrium* (5)–(10) *if and only if* (u, v) *minimizes the primal problem* (11) *and* (ϵ, λ) *maximizes its dual problem* (12).

Idea of proof. We sketch a proof here, side-stepping the subtlety mentioned above, by assuming boundedness of u and v to ensure that all integrals in question converge. Integrating the stability constraint (7) for the education market against ϵ yields

$$\alpha(u) - c\epsilon(b_\theta) \geq (z_{\#}\epsilon)(v) - \frac{1}{N'}\epsilon^2(v) \tag{13}$$

$$\geq \lambda^1(v) + \frac{1}{N'}\lambda^2(v) \tag{14}$$

$$\geq \lambda(\tilde{b}_{\theta'}), \tag{15}$$

where $(\epsilon, \lambda) \in R(\alpha)$ has been used to obtain (13)–(14), and the stability constraint for the labor market (8) has been used in (15). This shows $LP_* \geq LP^*$. Moreover, the conditions for equality in (13) and (15) coincide precisely with the budget constraints (10). Thus any competitive equilibrium forces (u, v) to minimize the primal linear program, and (ϵ, λ) to maximize its dual. Conversely, since we have independently deduced $LP^* = LP_*$ using the Fenchel-Rockafellar duality theorem, any bounded pair of optimizers $(u, v) \in F_0$ and $(\epsilon, \lambda) \in R(\alpha)$ must saturate the chain of inequalities above, hence satisfy the budget constaint and form a competitive equilibrium. □

Having established the equivalence between equilibrium and optimality, it is natural to want to establish the existence of minimizers for the primal problem and maximizers for the dual. As is typically the case in Fenchel-Rockafellar duality, existence of optimizers for the dual problem comes for free: it set in the Banach space dual to $(C(\bar{A}^2), \|\cdot\|_\infty)$, which is a space of measures whose unit ball is well-known to be weak-$*$ compact. Since ϵ and λ both belong to this unit ball, it is easy to extract a subsequential limit from a maximizing sequence, and this limit is the maximizer. To show the primal infimum is attained is much more subtle, since the only obvious bound on u (and hence v) is in $L^1(\bar{A}, \alpha)$. To address it, we shall need to learn more about what to expect in terms of the structure of any optimal (u, v).

4. Existence and structure of optimal wages

Given stable $(u, v) \in F_0$, the convex functions $b_E(z)$, $z(a,k) = (1-\theta)a + \theta k$ and $\tilde{b}_{\theta'}(k', k) := b_L((1-\theta')k' + \theta' k)$ can be used to define the wages implicitly available to an individual of cognitive skill k employed as a worker, manager, or teacher, respectively:

$$v_w(k) := \max_{k' \in \bar{A}} \tilde{b}_{\theta'}(k, k') - \tfrac{1}{N'} v(k'), \tag{16}$$

$$v_m(k) := N' \max_{k' \in \bar{A}} \tilde{b}_{\theta'}(k', k) - v(k'), \qquad \text{and} \tag{17}$$

$$v_t(k) := N \max_{a \in \bar{A}} c b_E(z(a,k)) + v(z(a,k)) - u(a) \qquad \text{where} \tag{18}$$

$$\infty - \infty := \infty. \tag{19}$$

Notice that v_m and v_w are suprema of convex functions of k; hence inherit uniform convexity directly from (3). It is not obvious whether or not v_t is convex — unless v is convex, in which case v_t is convex and inherits uniform convexity from b_E when $c > 0$. Similarly, convexity of

$$\bar{u}(a) := \max_{k \in \bar{A}} c b_E(z(a,k)) + v(z(a,k)) - \tfrac{1}{N} v(k) \tag{20}$$

is not obvious, unless v is convex, in which case $\bar{u}$ is a convex function which inherits uniform convexity when $c > 0$. These observations play a crucial role in our proof that $\alpha(u)$ attains its minimum on F_0.

Our strategy is the following: first we minimize (11) on the smaller set $F_0 \cap C_0$ consisting of pairs of *convex* non-decreasing functions $(u, v) \in F_0$. For $c > 0$, we then hope to show the minimizer over this restricted set is actually uniformly convex and increasing, in the sense that its first two derivatives are bounded away from zero. In this case the convexity and monotonicity constraints do not bind, so the minimum over the smaller set $F_0 \cap C_0$ also minimizes $\alpha(u)$ over the larger set F_0. (The existence of minimizers in case $c = 0$ can then be handled by taking a limit $c \to 0^+$ and relying on the compactness properties of the set of convex functions. The question of whether or not *uniform* convexity of u and v remains true in this limit requires a more subtle analysis in [10]; its conclusion is appended to Theorem 4.1 below.)

Stability (7) implies that the students' net lifetime utility satisfies $u \geq \bar{u}$, which corresponds to the fact that, in a competitive equilibrium, every student chooses to study with the teacher who represents the best educational investment for him. On the other hand, since we seek to minimize the expectation $\alpha(u)$, it costs no generality to assume this bound is saturated, meaning $u = \bar{u}$. Stability (7)–(9) also implies $v \geq \bar{v} := \max\{v_w, v_m, v_t\}$, which corresponds to the fact that, in a competitive equilibrium, each adult chooses the most financially rewarding occupation and professional partners for him or herself. Under the plausible hypothesis $v = \bar{v}$, our existence argument would be complete (at least in case $c > 0$). Unfortunately, we can only really expect $v = \bar{v}$ on the set of skills represented in the adult population, which might form a complicated subset of A and vary considerably along a minimizing sequence in F_0. Our strategy for circumventing this difficulty is to perturb both the primal and dual problems artificially, to ensure that adult skills populate the entire range A at some minimal level $\delta > 0$, solve the perturbed problems, and then take a limit $\delta \to 0$. In this way, we arrive at:

Theorem 4.1 (Existence of minimizing wages). *Fix $c \geq 0$ and positive θ, θ', N, N' and $\bar a = \bar k$ with $\max\{\theta, \theta'\} < 1 \leq N$ and $A = [0, \bar a] = K$. Let α be a Borel probability measure on $\bar A$ satisfying the doubling condition (4) at $\bar a$, and define $z(a,k) = (1-\theta)a + \theta k$, $b_\theta = b_E \circ z$ and $\tilde b_{\theta'}(a,k) = b_L((1-\theta')a + \theta' k)$, where $b_{E/L} \in C^1(\bar A)$ satisfy (1)–(3). Then infimum (11) is attained by functions $(u,v) \in F_0$ satisfying $v = \max\{v_w, v_m, v_t\}$ and $u = \bar u$ on $\bar A = [0, \bar a]$, where the $v_{w/m/t}$ and $\bar u$ are defined by (16)–(20); here $u, v : \bar A \longrightarrow]0, \infty]$ are continuous, convex, non-decreasing, and — except perhaps at $\bar a$ — real-valued. For $j \in \{1,2\}$, if $N\theta^j \geq 1$ then $d^j v/dk^j \geq \underline{b}_L^{(j)} \min\{(1-\theta')^j, (\theta')^j N'\}$.*

5. Who matches with whom?

We next try to understand which adults will choose to become workers, managers, or teachers, and with whom they will collaborate. The convex wages we have just shown to exist provide a key tool in this endeavor.

On the one hand, the slopes of v_w and v_m are inherited from $(1-\theta')b_L$ and $\theta' N' b_L$ according to (16)–(18), so if $(1-\theta')$ is very different from $\theta' N'$, we will have each worker being more skilled than each manager, or vice versa. How different these parameters must be depends on the range of slopes possessed by $b_L \in C^1(\bar A)$, as reflected in the ratio $\bar b'_L / \underline{b}'_L$. Here

$$\bar b'_{E/L} = b'_{E/L}(\bar a) = \sup_{a \in \bar A} b'_{E/L}(a)$$

and $\underline{b}'_L$ is from (2). Similarly, v_t inherits its slopes from $(cb_E + v)N\theta$, so taking $N\theta c \underline{b}'_E$ large enough relative to the parameters mentioned above ensures that the cognitive skills of each teacher will exceed those of all managers and workers. However if c is small or vanishes, meaning education has little or no value outside the labor market, things become more subtle, as in our Proposition 5.2 below. See also the numerical simulations of [22].

The other major tool that we have at our disposal is the knowledge that the functions

$$f(a,k) := u(a) + \frac{1}{N} v(k) - cb_E(z(a,k)) - v(z(a,k))$$
$$g(a,k) := v(a) + \frac{1}{N'} v(k) - b_L((1-\theta')a + \theta' k)$$

are non-negative throughout $\bar A^2$ by the stability of $(u,v) \in F_0$, yet f vanishes ϵ a.e. and g vanishes λ-a.e. by the budget constraint (10). In other words, ϵ is supported on the set where f is minimized, and λ on the set where g is minimized. Thus we expect the derivatives of f and g to vanish ϵ-a.e. and λ-a.e. respectively, provided these derivatives exist; that is, we expect

$$\frac{u'(a)}{1-\theta} = [cb'_E + v']_{(1-\theta)a+\theta k} = \frac{v'(k)}{N\theta} \qquad \text{to hold } \epsilon\text{-}a.e. \text{ and} \tag{21}$$

$$\frac{v'(a)}{1-\theta'} = b'_L((1-\theta')a+\theta' k) = \frac{v'(k)}{N'\theta'} \qquad \text{to hold } \lambda\text{-}a.e. \tag{22}$$

Similarly, we expect the Hessians of f and g to be non-negative definite ϵ-a.e. and λ-a.e. respectively, provided these derivatives exist:

$$\frac{u''(a)}{(1-\theta)^2} \geq [cb''_E + v'']_{(1-\theta)a+\theta k} \leq \frac{v''(k)}{N\theta^2} \qquad \text{should hold } \epsilon\text{-}a.e. \text{ and} \tag{23}$$

$$\frac{v''(a)}{(1-\theta')^2} \geq b''_L((1-\theta')a+\theta' k) \leq \frac{v''(k)}{N'(\theta')^2} \qquad \text{should hold } \lambda\text{-}a.e., \tag{24}$$

plus $\det D^2 f \geq 0$ and $\det D^2 g \geq 0$ should hold ϵ-a.e. and λ-a.e. respectively. In fact, for each $(a,k) \in \operatorname{Spt}\epsilon$ we can show the first equality in (21) holds provided $a \in \operatorname{Dom} Du$, while the second equality holds provided $k \in \operatorname{Dom} Dv$; here $\operatorname{Dom} Du$ denotes the subset of $]0,\bar a[$ where u is differentiable, and $\operatorname{Dom} D^2 u$ denotes the further subset where the non-decreasing function $u'(a)$ has a derivative in the sense of Lebesgue. (It is straightforward to see for each $(a,k) \in \operatorname{Spt}\lambda$ that the first equality in (22) holds provide $a \in \operatorname{Dom} Du$, and the second equality holds provided $k \in \operatorname{Dom} Dv$.)

Assuming α has no atoms, convexity of u ensures $\operatorname{Dom} Du$ constitutes a set of full measure; likewise $\operatorname{Dom} D^2 u$ is a set of full measure when α is absolutely continuous with respect to Lebesgue. If $c > 0$ or convexity of v is strict, (21) can be solved to identify the skill

$$k = k_t(a) = \frac{1}{\theta}(cb'_E + v')^{-1}\left(\frac{u'(a)}{1-\theta}\right) - \frac{1-\theta}{\theta} a \tag{25}$$

of each teacher who teaches students of skill $a \in \operatorname{Dom} Du$. It is less transparent to see that the convex function v is differentiable at almost every adult skill level, since the distribution of adult skills $z_{\#}\alpha$ is not prescribed, but rather determined by the model. It is therefore useful to know whether or not $z_{\#}\alpha$ can have atoms, assuming α does not. The following lemma rules out atoms in $z_{\#}\alpha$ provided ϵ is positive assortative; it also shows $z_{\#}\alpha$ inherits absolute continuity with respect to Lebesgue from α in that case. *Positive assortativity* simply means $\operatorname{Spt}\epsilon$ is a non-decreasing subset of $\mathbf{R}^2$, so that the skill of each teacher cannot decrease as a function of the skill of the students they teach.

Lemma 5.1 (Endogenous distribution of adult skills). *Fix $\theta \in]0,1[$ and a Borel probability measure $\alpha \geq 0$ on $\bar A$ with $A = [0,\bar a[$. Set $z(a,k) = (1-\theta)a + \theta k$. If $\epsilon \geq 0$ on $\bar A^2$ has $\alpha = \epsilon^1$ as its left marginal, then for each $\bar a - \Delta a \in A$ the corresponding distribution $\kappa = z_{\#}\epsilon$ of adult skills satisfies*

$$\int_{[\bar a - \Delta a, \bar a]} d\kappa(a) \leq \int_{[\bar a - \frac{1}{1-\theta}\Delta a, \bar a]} d\alpha(a). \tag{26}$$

Thus κ has no atom at $\bar{a}$ unless α does.

In addition, if ϵ is positive assortative and α has no atoms, then κ has no atoms and $\epsilon = (id \times k_t)_{\#}\alpha$ for some non-decreasing map $k_t : \bar{A} \longrightarrow \bar{A}$. uniquely determined α-a.e. by κ. Moreover, if $d\alpha(a) = \alpha^{ac}(a)da$ is given by a density $\alpha^{ac} \in L^1(A)$, then $d\kappa(a) = \kappa^{ac}(a)da$ is given by a related density $\kappa^{ac} \in L^1(A)$ satisfying

$$\alpha^{ac}(a) = (1 + \theta(k_t'(a) - 1))\, \kappa^{ac}(z(a, k_t(a))) \tag{27}$$

for Lebesgue-a.e. $a \in A$. In this case $\|\kappa^{ac}\|_{L^\infty(A)} \leq \frac{1}{1-\theta}\|\alpha^{ac}\|_{L^\infty(A)}$.

Our next theorem shows that positive assortativity of ϵ holds as long the equilibrium payoffs $(u, v) \in F_0$ are *strictly* convex. It also explains when and in what sense equilibria will be unique. Before stating it we cite a proposition which details more elaborate consequences of the foregoing analysis concerning who will work, manage and teach. For the phase transition which we plan to describe, it is particularly relevant to have criteria such as (c) below, ensuring that even for $c = 0$, the most skilled individuals will be teachers. It is also essential for the theorem which follows, to know that the skill levels of the academic descendants of almost every given teacher are only finite in number. By contrast, there will certainly be teachers whose academic ancestors populate countably many skill types.

Note that in the following proposition, (c) and (d) together imply (e), meaning at least one of the two inequalities $N\theta \geq 1$ or $c \geq 0$ is strict. Also note $N'\theta' \geq \bar{b}_L'/\underline{b}_L'$ and $N\theta \geq \bar{b}_L'/\underline{b}_L'$ are sufficient for (b) and (c), respectively.

Proposition 5.2 (Specialization by type; the educational pyramid). *Fix $A = [0, \bar{a}[$ with $\bar{a} > 0$, and $c \geq 0$. Extend convex, nondecreasing $u, v : A \longrightarrow \mathbf{R}$ lower semicontinuously to $\bar{A}$ and suppose $v = \max\{v_w, v_m, v_t\}$, where $v_{w/m/t}$ are from* (16)–(19).

If (a) $N\theta c\underline{b}_E' \geq \bar{b}_L' \max\{N'\theta', 1 - \theta'\}$ then all teacher types lie weakly above all of the manager and worker types.

If (b) $N'\theta' > (1 - \theta') \sup_{a\in A} b_L'(1 - \theta')a + \theta'\bar{a})/b_L'(\theta' a)$ then all of the worker types lie weakly below all of the manager types.

If (c) $N\theta \geq \sup_{0\leq z\leq k} b_L'((1-\theta')z+\theta'\bar{a})/(b_L'(\theta' z)+\frac{c}{N'\theta'}b_E'(z))$ and (b) holds, and $f(a, k) := u(a) + \frac{1}{N}v(k) - cb_E(z(a, k)) - v(z(a, k))$ vanishes at some $(a, k) \in A \times A$ where $v(z(a, k)) = v_m(z(a, k))$, then $v > v_m$ on $]k, \bar{a}]$. In other words, no manager (or worker) can have a type higher than a teacher of managers.

If (d) $N\theta \geq 1$, then any student of type $a \in A$ will be weakly less skilled than his teacher, and strictly less skilled if (e) either $c > 0$ or $N\theta > 1$ in addition.

If (f) either $c > 0$ or $v'(0) > 0$, then (d)–(e) imply all academic descendants of a teacher with skill $k \in A$ will display one of at most finitely many $d = d(k)$ distinct skill types, unless differentiability of v fails at k. However, $d(k)$ may diverge as $k \to \bar{a}$, in which case $v'(k) \to +\infty$ at a rate we can estimate.

We are finally in a position to state our positive assortativity and uniqueness results.

Theorem 5.3 (Positive assortative and unique optimizers). *Adopting the hypotheses and notation of Theorem 4.1, if $(\epsilon, \lambda) \in R(\alpha)$ maximize the dual problem* (12), *then the labor matching λ is positive assortative. Moreover, there exist a pair of maximizers (ϵ, λ) for which the educational matching ϵ is also positive assortative.*

If there exist minimizing payoffs $(u, v) \in F_0$ for the primal problem (11) which are strictly convex and increasing, (as for example if either $c > 0$ or $N\theta^2 \geq 1$), then any maximizing ϵ and λ are positive assortative. If, in addition, α is free from atoms then the maximizing ϵ and λ are unique. If, in addition, hypotheses (d)-(f) from Proposition 5.2 hold, then u' and v' exist and are uniquely determined α-a.e. and $(z_{\#}\epsilon)$-a.e. respectively. If, in addition, α dominates some absolutely continuous measure whose support fills $\bar{A}$, and $(u_0, v_0) \in F_0$ is any other minimizer with $v_0 : A \longrightarrow \mathbf{R}$ locally Lipschitz then $u_0 = u$ holds α-a.e., meaning u_0 is unique.

Regarding the marginals of λ as fixed, positive assortativity follows from the fact that λ is chosen to maximize a surplus $\tilde{b}_{\theta'}(a, k) = b_L((1-\theta)a+\theta k)$ whose cross-partial derivatives are positive (owing to the uniform convexity of b_L). Such results have played a celebrated role in the economics literature since the work of Mirrlees on taxation [24], Becker on marriage [1], and Spence on educational signaling in the labor market [29]; in the mathematical literature they date to Lorentz' earlier work on rearrangement inequalities [20]. The positive assortativity of ϵ cannot be derived in quite the same way, since it is $z_{\#}\epsilon - \epsilon^2/N$ rather than ϵ^2 which is fixed in the maximization (12). However, it is strongly suggested by (23) and (25), and can be rigorously derived from cross-partial derivatives of the expressions appearing in the suprema (18)–(20), whose positivity relies on the uniform convexity of the endogenous wage profile $v(k)$ established in Theorem 4.1. In view of (23) and (24), one can also view this convexity as propagating from the wage of each adult to the wage of their teacher; it takes finitely many steps to reach any teacher's skill level in the educational pyramid, by Proposition 5.2(f).

Uniqueness of ϵ and λ follow from positive assortativity once their marginal distributions are known. My favorite proof of this fact appears in [21]. The distribution of student skills $\epsilon^1 = \alpha$ is specified a priori, and the distribution of teacher skills ϵ^2 can be worked out from the equilibrium payoffs (u, v) using the student-teacher skill correspondence $k = k_t(a)$ given by (25). This follows a strategy which has become standard in optimal transportation since the work of Brenier [4], Caffarelli [5], Gangbo and myself [15]. To specify the marginals of λ uniquely requires sorting out who will be a worker and who will be a manager, allowing for the possibility that their skill distributions λ^1 and λ^2 may overlap. A precedent for deriving uniqueness in such settings appears in work with Trokhimtchouk [23].

6. Transition to unbounded wage gradients

Finally, we are in a position to address our motivating question, which is the possibility of singularities in the wage profile at the apex $\bar{a}$ of the skills pyramid $A = [0, \bar{a}[$. By analyzing the recursion (21) relating the wage of each teacher to the future earnings of their students, we are able to prove the following singularity alternative:

Theorem 6.1 (Wage behavior and density of top-skilled adults). *Adopting the hypotheses and notation of Theorem 4.1, let α be given by a Borel probability density $\alpha^{ac} \in L^\infty(A)$ which is continuous and positive at the upper endpoint of $A = [0, \bar{a}[$. Suppose $(\epsilon, \lambda) \in R(\alpha)$ and convex $(u, v) \in F_0$ optimize the primal and dual problems (11)–(12), and (i) $\bar{a} \in (\mathrm{Spt}\,\epsilon^2) \setminus \mathrm{Spt}(\lambda^1 + \lambda^2)$, meaning all adults with sufficiently high skills become teachers; (ii) the educational matching ϵ is positive assortative, meaning a non-decreasing correspondence $k = k_t(a)$ relates the ability of α-a.e. student a to that of his teacher; (iii) k_t is differentiable at $\bar{a}$, and (iv) v is differentiable on some interval $]\bar{a} - \delta, \bar{a}[$. Then for $N\theta \neq 1$,*

$$v'(a) = \frac{const}{|\bar{a} - a|^{\frac{\log N\theta}{\log N}}} - \frac{c\bar{b}'_E}{1 - \frac{1}{N\theta}} + o(1) \tag{28}$$

as $a \to \bar{a}$.

In Proposition 5.2 and Theorem 5.3 we have already seen that $N'\theta'$, $N\theta$ and $N\theta^2$ large enough guarantee (i) and (ii). We do not know conditions which guarantee (iii)-(iv). However (iii) follows from (25) if $b_E, u, v \in C^2$ near $\bar{a}$, so our theorem guarantees that some nearby singularity is produced when (i)–(ii) hold. The difference quotients whose limit defines $k'_t(\bar{a})$ are bounded under the conditions of Proposition 5.2(d), in which case we judge failure of (iii) to be less likely that the smooth gradient blow-up predicted by (28). We judge an accumulation (iv) of non-differentiabilities $A \setminus \mathrm{Dom}\, Du$ at $\bar{a}$ to be even more unlikely. On the other hand, the leading order behavior of (28) changes, depending on whether the influence $(N\theta)^d$ of a given teacher grows or decays as one moves through successive generations $d \geq 1$ of their academic descendants. This strongly suggests a sharp transition from bounded to unbounded wage gradients, at the critical value $N\theta = 1$ where this influence remains constant from generation to generation. Note however that the singularity (28) in the gradient is integrable, so that even when it is present the wages $v(k)$ tend to a finite limit as $k \to \bar{k}$. This answers the question raised at the outset: at least in the context of the present model, the maximum wage tends to a finite multiple of the average wage in the large population limit; its sensitivity to skill level, however, can be bounded or unbounded, depending on the effectiveness of education.

References

[1] G.S. Becker. A theory of marriage. Part I. *J. Political Econom.* **81** (1973) 813–846.

[2] G.S. Becker and K.M. Murphy. The division of labor, coordination costs, and knowledge. *Quart. J. Econom.* **107** (1992) 1137–1160.

[3] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.* **17** (2013) 477–501.

[4] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris Sér. I Math.* **305** (1987) 805–808.

[5] L. Caffarelli. Allocation maps with general cost functions. In P. Marcellini et al, editors, *Partial Differential Equations and Applications*, Lecture Notes in Pure and Appl. Math. number 177, pages 29–35. Dekker, New York, 1996.

[6] G. Carlier. Duality and existence for a class of mass transportation problems and economic applications. *Adv. Math. Econom.* **5** (2003) 1–21.

[7] P.-A. Chiappori, R.J. McCann, and L. Nesheim. Hedonic price equilibria, stable matching and optimal transport: equivalence, topology and uniqueness. *Econom. Theory* **42** (2010) 317–354.

[8] M.J.P Cullen and R.J. Purser. An extended Lagrangian model of semi-geostrophic frontogenesis. *J. Atmos. Sci.*, 41:1477–1497, 1984.

[9] I. Ekeland. Existence, uniqueness and efficiency of equilibrium in hedonic markets with multidimensional types. *Econom. Theory* **42** (2010) 275–315.

[10] A. Erlinger, R.J. McCann, A. Siow, X. Shi, and R. Wolthoff. Academic wages and pyramid schemes: a mathematical model. In preparation.

[11] A. Figalli, Y.-H. Kim, and R.J. McCann. When is multidimensional screening a convex program? *J. Econom Theory* **146** (2011) 454–478.

[12] X. Gabaix and A. Landier. Why has CEO compensation increased so much? *Quart. J. Econom.* **123** (2008) 49–100.

[13] D. Gale and L.S. Shapley. College admissions and the stability of marriage. *Amer. Math. Monthly* **69** (1962) 9–15.

[14] A. Galichon and B. Salanié. Cupid's invisible hand: social surplus and identification in matching models. *Working paper*, 2011.

[15] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.* **177** (1996) 113–161.

[16] L. Garicano. Hierarchies and the organization of knowledge in production. *J. Political Econom.* **108** (2000) 874–904.

[17] L. Garicano and E. Rossi-Hansberg. Organization and inequality in a knowledge economy. *Quarterly J. Econom.* **121** (2006) 1383–1435.

[18] N.E. Gretsky, J.M. Ostroy, and W.R. Zame. The nonatomic assignment model. *Econom. Theory* **2** (1992) 103–127.

[19] R. Jordan, D. Kinderlehrer and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** (1998) 1–17.

[20] G.G. Lorentz. An inequality for rearrangements. *Amer. Math. Monthly* **60** (1953) 176–179.

[21] R.J. McCann. Exact solutions to the transportation problem on the line. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, *455* (1999) 1341–1380.

[22] R.J. McCann, X. Shi, A. Siow, and R. Wolthoff. The organization of the labor market with communication and cognitive skills. *Preprint at* `www.math.toronto.edu/mccann/publications`.

[23] R.J. McCann and M. Trokhimtchouk. Optimal partition of a large labor force into working pairs. *Econom. Theory* **42** (2010) 375–395.

[24] J.A. Mirrlees. An exploration in the theory of optimum income taxation. *Rev. Econom. Stud.* **38** (1971) 175–208.

[25] J.-C. Rochet and P. Choné. Ironing, sweeping and multidimensional screening. *Econometrica* **66** (1998) 783–826.

[26] S. Rosen. The economics of superstars. *Amer. Econom. Rev.* **71** (1982) 845–858.

[27] L.S. Shapley and M. Shubik. The assignment game I: The core. *Internat. J. Game Theory* **1** (1972) 111–130.

[28] H. Sonnenschein. Price dynamics based on the adjustment of firms. *Amer. Econom. Rev.* **72** (1982) 1088–1096.

[29] M. Spence. Job market signaling. *Quarterly J. Econom.* **87** (1973) 355–374.

Department of Mathematics, University of Toronto, Toronto ON Canada M5S 2E4
E-mail: mccann@math.toronto.edu