Academic wages and pyramid schemes: a mathematical model[*]

Alice Erlinger[†][‡], Robert J. McCann[‡], Xianwen Shi[§],
Aloysius Siow[§] and Ronald Wolthoff[§]

August 10, 2015

Abstract

This paper analyzes a steady state matching model interrelating the education and labor sectors. In this model, a heterogeneous population of students match with teachers to enhance their cognitive skills. As adults, they then choose to become workers, managers, or teachers, who match in the labor or educational market to earn wages by producing output. We study the competitive equilibrium which results from the steady state requirement that the educational process replicate the same endogenous distribution of cognitive skills among adults

[†]Laboratoire Dieudonné, Université de Nice, Parc Valrose 06108 Nice Cedex 2 France `alice.erlinger@gmail.com`

[‡]Department of Mathematics, University of Toronto, Toronto Ontario M5S 2E4 Canada, `mccann@math.toronto.edu`

[§]Department of Economics, University of Toronto, `xianwen.shi@utoronto.ca`, `siow@chass.utoronto.ca`, `ronald.wolthoff@utoronto.ca`

in each generation (assuming the same distribution of student skills). We show such an equilibrium can be found by solving an infinite-dimensional linear program and its dual. We analyze the structure of our solutions, and give sufficient conditions for them to be unique. Whether or not the educational matching is positive assortative turns out to depend on convexity of the equilibrium wages as a function of ability, suitably parameterized; we identity conditions which imply this convexity. Moreover, due to the recursive nature of the education market, it is a priori conceivable that a pyramid scheme leads to greater and greater discrepancies in the wages of the most talented teachers at the top of the market. Assuming each teacher teaches N students, and contributes a fraction $\theta \in]0, 1[$ to their cognitive skill, we show a phase transition occurs at $N\theta = 1$, which determines whether or not the wage gradients of these teachers remain bounded as market size grows, and make a quantitative prediction for their asymptotic behaviour in both regimes: $N\theta \geq 1$ and $N\theta < 1$.

1   Introduction

It is an economic truism that prices are determined primarily by what the market will bear. For example, executive compensations in large firms may appear excessive when measured against average employee wages, but are often justified by arguing that they are determined competitively by the market. To understand what levels of compensation a large market will or won't bear, it is therefore tempting to ask questions such as: Can the ratio of the CEO's wages over the average wage in a firm be expected to tend to infinity or a finite limit, as the size of the firm grows without bound? The answer to such a question may be expected to depend on various aspects of the organization of the firm, such as the number of levels of management separating the CEO from the average worker, and the number of managers at each level. This organizational structure may itself be determined by market pressures — within the constraints of feasible technology.

In this paper we investigate an analogous question set in the context of the education market, rather than that of a firm. That is, we investigate how the wages of the most sought after gurus relate to those of the average teacher. The education market is special in various ways. It is stratified into many different levels or streams which interact with each other, with a

range of qualities available in every stream. Moreover, what it produces is human capital, the value of which is determined by the broader market for skills of which the education market is itself a small part. Thus there is a feedback mechanism in the education market, owing to the fact that those individuals who choose to become teachers participate at least twice in the market: first as consumers and later as producers, putting to work the skills previously acquired in this market to generate human capital for the next generation. It is this feedback mechanism which is responsible for many of the results we describe; it leads to the formation of an educational analog for a pyramid scheme, in which teachers at each level of the pyramid attempt to extract as much as they can from their students future earnings, in the form of tuition. The question this time is whether or not the large market limit leads to wages which display singularities at the apex of the pyramid.

We address this question using a variant of a steady state overlapping-generations matching model introduced by four of us to analyze the coupling of the education and labor markets [17]. We proposed this model not only to provide a microeconomic foundation which allows to compare and contrast different sectors, but to examine interdependencies and the different roles played by communication and cognitive skills in each of them. In this sense, our model improves on the work by Rosen [19] who analyzes "superstars" in the labor market in isolation. By allowing individuals to invest in their ability before entering the labor market, our work is also related to the literature on pre-matching investments (see e.g. [18] and the references there). We contribute by proposing an explicit model of the education market and the role of teachers rather than specifying a reduced-form investment function.

An unexpected conclusion in our companion paper [17] was that — as in much simpler (single stage, single sector) models [20] [11] [5], competitive equilibrium matching patterns for a heterogeneous steady state population can be found as the optimal solution to a stationary planner's problem taking the form of a linear program; see also [6]. The questions raised in the present manuscript will be addressed through a rigorous analysis of the resulting linear program and its solutions, including criteria for existence, uniqueness, singularities, and a detailed description of the matching patterns which can arise. A remarkable feature is that this simple model leads to the emergence of a hierarchical structure in the education sector, with fewer and fewer individuals at the top of the market earning higher and higher wages. A detailed exploration of this structure proves necessary to resolve the question of under what conditions these wages turn out to display singularities. An

3

analogous hierarchy was explored by Becker and Murphy in the context of a steady growth model [2, §VII] quite different from ours.

The education market is also unusual in many ways that our model does not capture. For example, non-pecuniary considerations are important for both teachers and students, and schools are often not operated on a for-profit basis; however, in our model we assume all participants maximize their expected monetary payoff. In addition, education markets (tuitions, for example) are heavily regulated, but here we abstract away all regulation restrictions. The goal of this paper, therefore, is not to provide a realistic account of how teachers' compensations are determined in the market, but rather to elucidate a feedback mechanism that is potentially important in determining wage compensation in education and other markets, and to provide a tool to solve matching models that incorporate this feedback mechanism with the potential to encompass multi-dimensional individual attributes.

In the present model, we assume that the communication skills are homogeneous over the entire population, hence deal with a population having a single dimension of heterogeneity plus parameters, rather than the multiple dimensions of heterogeneity in [17]. Hence, the model here can be viewed as a limiting case of multidimensional models in which the range of heterogeneities becomes narrow in all dimensions but one. There are two benefits from this simplifying assumption. First, it greatly simplifies our analysis. Second, the resulting model is a minimal departure from the classical matching model of one dimension of heterogeneity. We will show that this small departure actually generates results very different from the standard one-dimensional models of e.g. Lucas [13] or Garicano [9].

As in [17], we model communication skills as the number of students a teacher can teach or the number of workers a manager can manage, which is often referred to as "span of control". In particular, we assume that each teacher can teach $N > 1$ students. We use $\theta \in]0, 1[$ to represent the extent to which a teacher's cognitive skills get transmitted to each of their students. Similarly, $N' > 0$ and $\theta' \in]0, 1[$ represent the number of workers each manager can manage in the labor market, and the extent to which a manager's cognitive skill enhances the productivity of his or her workers. All market participants have the same N in the education market and the same N' in the labor market, but they differ in the cognitive skills k which are assumed to be continuously distributed over the interval $\bar{K} := [\underline{k}, \bar{k}] \subset \mathbf{R}$. As a result, the linear program is infinite-dimensional, and the analysis is complicated by a lack of a priori bounds which could be used to show that equilibrium

4

wages or payoffs exist for the model.[1] Moreover, a pyramid can form in the education sector, enhancing the wages of the most skilled teachers. It is not obvious whether or not this pyramid structure can lead to unbounded wage behavior. Our analysis suggests it does not, but leads to unbounded wage gradients instead.

We begin by elucidating a convexity property which allows us to derive the existence of equilibrium wages as solutions to an (infinite-dimensional) linear program. This convexity is reminiscent of that discussed by Rosen in his investigation of superstars [19]. More surprisingly, after addressing uniqueness and properties of these wages and the matches they induce, we go on to show that the model exhibits a phase transition, depending on the product of each teacher's capacity N for students times their teaching effectiveness θ: the wage gradients diverge at the highest skill type if and only if $N\theta \geq 1$. When $N\theta > 1$, the divergence is proportional to $|\bar{k} - k|^{-\frac{\log \theta}{\log N} - 1}$ as $k \to \bar{k}$. Only by integrating this divergence can we conditionally show wages tend to a finite limit at $\bar{k}$ which — in the large market limit — becomes independent of the size of the population being modelled.[2]

Although wage singularies for teachers may appear counter-factual, or at least modest compared to wage singularities for managers in the real world, this discrepancy between prediction and observation is easily explained by the fact that our model allows for only one layer of managers but a potentially unbounded number of layers of teachers. Thus a top teacher improves the cognitive skills of each of their N students who go on to be top teachers or managers. A good manager improves the productivity of each of their N' supervised workers. Thus, already in a two-layer hierarchy, a top teacher indirectly makes a large number $N \times N'$ of workers more productive. Since the number of layers of the educational hierarchy is endogenous to the model and can be very large, the impact of gurus on the productivity of their direct

---

[1]Although one can also formulate analogous models involving only finitely many types, an important advantage of the infinite-dimensional approach employed here is that it avoids binding integrality constraints which can potentially lead to non-existence of the equilibrium, as discussed e.g. in [4].

[2]One can also use our model to explore how changes in the span of control (i.e. a larger N and N', for example due to technological progress such as the advent of massively online open courses which have the potential to transform the education sector) impact the income inequality predicted by the model. In our companion paper [17], we show through simulation that as the span of control increases, the equilibrium number of managers and teachers falls, and the skewness of log wages increases.

and indirect students and workers can accumulate very substantially.

The term *phase transition* is borrowed from statistical physics, where it refers to a sharp threshold in parameters (such as temperature) separating qualitatively different behavior (such as liquid from solid). In that context, the non-smoothness arises from a continuum limit which admits approximation by finite dimensional models depending smoothly on the same parameter(s). By analogy, if our continuum of agent types could be approximated using finitely many agent types, we would expect to restore smooth dependence on the parameters N and θ, but this smoothness (i.e. the wage gradients) would not admit control uniform in the number of types. In statistical physics, it is often the case that the critical exponents of the singularities (such as $\frac{\log \theta}{\log N}$ above) do not vary over a wide class of models, a phenomenon known as universality. In the present context, we observe that the exponent $\frac{\log \theta}{\log N}$ governing growth of the wage gradients is universal in the sense that it does not depend on various details of the model, such as the exact form of the production functions, or the input distribution of student skills, at least within the classes of such data considered hereafter.

The remainder of this manuscript is organized as follows. In the first section and subsections we lay out the model, and its variational reformulation in terms of a planner's problem and its dual. We have argued in [17] that solutions to these infinite-dimensional linear programs represent competitive equilibria; the announcement [16] goes further by establishing equivalence between stationary competitive equilibria and solutions to such planner's problems, which can also be interpreted as intergenerational optima in the context of an overlapping-generations model.[3] In a second section and subsections we address the existence, uniqueness and properties of these solutions. Even the existence of equilibrium wages in this model is rather non-trivial, and goes beyond the range of validity of any statement of the second welfare theorem that we know. Standard arguments concerning existence of an optimal matching and absence of a duality gap are relegated to an appendix, which is logically independent of the rest of the analysis. Lemma 14 is also logically independent of the remaining analysis, and its first assertion is actually required at some earlier points in the text.

---

[3]Throughout this paper we impose the stationarity which implies this equivalence. Non-stationary equilibria and/or optima may exist and may differ from each other and the stationary solutions studied here (see e.g. the discussion in chapter 17 of [21]). We leave a detailed analysis of such possibilities for future research.

1.1 The model: competitive equilibria

Let us begin by describing our unidimensional variant of the model first introduced by [17]. Consider an economy populated by risk-neutral individuals who each lives for two periods. Individuals, when they are young, enter the education market as students. In the subsequent period as adults, they enter the labor market to become teachers in schools, or workers or managers in firms. Both the education market and the labor market are competitive. There is free entry for both schools and firms. Hence, the tuition fees a school collects from students are just enough to cover the wage of its teacher, and a firm's output exactly covers the wages of its employees (workers and managers). All individuals do not discount. The lifetime net payoffs of individuals are equal to the sum of their labor market plus non-labor market earnings minus tuition costs. Individuals choose what occupation to pursue and who to match with in each of the two markets to maximize their net payoffs.

Each individual is endowed with two kinds of skills, a communication skill ($N > 1$ or $N' > 1$) which is fixed throughout their lifetime, and an initial cognitive skill a which can be augmented through education. As in [17], we assume that individuals differ in their initial cognitive skills a. In contrast to [17], we assume that individuals share the same communication skills. By attending schools in the first period, individuals can augment their initial cognitive skills a to their adult cognitive skill k. Let $A = [\underline{a}, \bar{a}[$ with $-\infty < \underline{a} < \bar{a} < +\infty$ denote the range of students' initial cognitive skills a, and $K = [\underline{k}, \bar{k}[$ or rather its closure $\bar{K}$ the range of adult human capital k. Ability or human capital refers to cognitive skill in both cases, and we occasionally use the variable names a and k interchangeably for convenience. For the model discussed here, taking $K = A$ will not cost any generality, nor will the normalization $\underline{a} = \underline{k} = 0$.

The production functions in the education market and in the labor market are described as follows. We assume the cognitive skill $z(a, k)$ acquired by a student of ability $a \in A$ who studies with a teacher of ability $k \in K$ is given by the weighted average $z(a, k) = (1 - \theta)a + \theta k$ of their abilities, with weight $\theta \in]0, 1[$. We also assume the productivity $b_L((1 - \theta')a + \theta'k)$ of a worker with adult cognitive skill a supervised by a manager of skill k is given by a convex increasing function $b_L \in C^1(\bar{K})$ of another such average, this time with weight $\theta' \in]0, 1[$. Notice that abilities a and k here are measured on a logarithmic scale relative to the conventions of [17], a reparameterization

which is crucial for exposing the sense in which the equilibrium wages may turn out to be convex.

We allow for the possibility that cognitive skill z attained through education has value $cb_E(z)$ in addition to the wage earning potential it confers, where $c \geq 0$ is a dimensionless parameter and $b_E \in C^1(\overline{A})$ is another convex increasing function. The choice $b_E(k) = e^k = b_L(k)$ with $\theta = \frac{1}{2} = \theta'$ corresponds to the motivating example from [17]; more generally we assume b_E and b_L and their first two derivatives have positive lower bounds

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

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

$$0 < \underline{b}''_{E/L} \quad = \quad \inf_k b''_{E/L}(k), \tag{3}$$

where $\underline{b}''_{E/L}$ is defined as the largest constant for which $b_{E/L}(k) - \underline{b}''_{E/L}|k|^2/2$ is convex on $\overline{K}$. We hope strict positivity of the analogous quantities will be inherited by the equilibrium payoffs u and v.

Notice that what is being produced in each sector is different: in the labor and non-labor sectors we have not specified the service or goods which are being produced, except that they take adult cognitive skills as their input (communication skills entering through possible dependence of c on parameters such as N and θ); in the education sector it is adult cognitive skills which are being produced, taking student and teacher cognitive skills as their inputs. The dimensionless constant $c \geq 0$ measures the non-labor utility, if any, of individual attainment of cognitive skills relative to labor productivity; it replaces the marital utility used in early drafts of [17].

Let a probability measure $\alpha \geq 0$ on $\overline{A}$ represent the exogenous distribution of student abilities, and let $\mathrm{spt}\,\alpha$ denote the smallest closed subset of $\overline{A}$ carrying the full mass of α. Taking A smaller if necessary ensures $\mathrm{spt}\,\alpha$ contains both $\underline{a}$ and $\overline{a}$. Our problem is to find a pair Borel measures $\epsilon \geq 0$ on $\overline{A} \times \overline{K}$ and $\lambda \geq 0$ on $\overline{K} \times \overline{K}$, such that ϵ represents the *educational* pairing of students with teachers, and λ represents the *labor* pairing of workers with managers, along with a pair of payoffs or wage functions $u, v : \overline{K} \longrightarrow [0, \infty]$ representing the net lifetime expected utility $u(a)$ of a student with ability a, and the wage $v(k)$ paid to an adult of ability k, which together constitute a competitive equilibrium $(\epsilon, \lambda, u, v)$. Roughly speaking, this means the matchings ϵ, λ must clear the market at each generation in a steady-state, and the payoffs u and v must be large enough to be stable, yet small enough that in combination with (ϵ, λ) they satisfy a budget constraint.

Since we are interested in a steady state model, we assume the distribution of student abilities α on $\bar{A}$ is the same at each generation, and coincides with the left marginal

$$\epsilon^1 = \alpha \tag{4}$$

of the educational pairing $\epsilon \geq 0$ of student and teacher abilities. Here $\epsilon^1 = \pi^1_\# \epsilon$ and $\epsilon^2 = \pi^2_\# \epsilon$ denote the left and right projections of ϵ through $\pi^1(a, k) = a$ and $\pi^2(a, k) = k$, representing the respective distributions of student and teacher abilities. Similarly λ^1 and λ^2 will denote the left and right marginals of the labor pairing λ, representing the distribution of worker and manager skills. The steady state constraint requires that the educational pairing ϵ of students with adults reproduce the current distribution of adult skills at the next generation:

$$\lambda^1 + \frac{1}{N'}\lambda^2 + \frac{1}{N}\epsilon^2 = z_\# \epsilon, \tag{5}$$

where the expression on the left represents the sum of the current distributions of worker, manager and teacher skills; the latter have been scaled by N' and N respectively, to reflect the fact that each manager manages N' workers, and each teacher teaches N students, so comparatively fewer managers and teachers are required. The symbol $\kappa := z_\# \epsilon$ on the right represents the distribution of future adult skills resulting from the educational pairing ϵ; it is given by the push-forward of ϵ through the map $z : \bar{A} \times \bar{K} \longrightarrow \bar{K}$ representing the educational technology, and assigns mass $\kappa[B] := \epsilon[z^{-1}(B)]$ to each set $B \subset \bar{K}$.

The marginal constraint (4) forces ϵ and hence $\kappa = z_\# \epsilon$ to be probability measures, like α. The workers form a fraction $(1 - \frac{1}{N})/(1 + \frac{1}{N'})$ of the population, coinciding with the total mass of λ. The restriction $K = A$ costs no generality, since we are in a steady state, and since our education technology satisfies $z(a, a) = a$, whence $z(\underline{a}, \underline{k}) = \underline{k}$ and $z(\bar{a}, \bar{k}) = \bar{k}$.

Letting $v(k)$ denote the wage commanded by an adult of skill k, and $u(a)$ the net lifetime utility of a student of ability a, both must satisfy the stability conditions

$$u(a) + \frac{1}{N}v(k) \geq cb_E(z(a, k)) + v(z(a, k)) \quad \text{and} \tag{6}$$

$$v(a) + \frac{1}{N'}v(k) \geq b_L((1 - \theta')a + \theta'k) \qquad \text{on } \bar{A} \times \bar{K}. \tag{7}$$

The constraint (7) enforces stability of matchings in the labor sector. If the reverse inequality held, N' adults with skills a and one with skill k would

9

abandon their occupations to form N' worker-manager pairs each producing enough output b_L to improve all $N' + 1$ adults' wages. Similarly (6) is a stable matching condition for the education sector. The lifetime net utility of a student with cognitive skill a plus the tuition $v(k)/N$ paid by each student of a teacher with skill k must exceed a's lifetime earnings plus any other benefits derived from cognitive skills which would have resulted had he (and $N - 1$ of his clones) chosen to study with k. We can also regard the stability constraints (6)–(7) as combining to ensure each adult of type k in the population chooses the profession (worker, manager, or teacher) and partners (manager, workers, or students, respectively) which maximize their wage $v(k)$ on the labor market.

Finally, the budget constraint asserts that equality holds ϵ-a.e. in (6), and λ-a.e. in (7). In other words, the productivity $b_L((1 - \theta')a + \theta'k)$ of λ-a.e. manager-worker pair (a, k) which actually forms is sufficient to pay the worker's wage plus a fraction $1/N'$ of the manager's salary. Similarly, ϵ-a.e. student-teacher pairing (a, k) which forms must produce an adult whose earnings $v(z(a, k))$, supplemented by any additional utility $cb_E(z(a, k))$ derived from the skill $z(a, k)$ he acquires, must add up to the net lifetime utility which remains to the student after paying tuition equal to his share $v(k)/N$ of his teacher's earnings.

To complete the specification of the model, we need to say in what class of functions the payoffs u, v must lie. Since we wish to allow for the possibility that the payoffs $u, v : K \longrightarrow [0, \infty]$ become unbounded at the upper end $\bar{k}$ of the skill range, it is convenient to define $A = K = [0, \bar{k}[$ as a half open interval. We shall consider payoffs from the feasible set F_0 consisting of pairs $(u, v) = (u_0 + u_1, v_0 + v_1)$ satisfying (6)–(7) which differ from bounded continuous functions $u_0, v_0 \in C(\bar{A})$ by non-decreasing functions $u_1, v_1 : \bar{A} \longrightarrow [0, \infty]$. If v takes extended real values, we also require

$$\tfrac{N}{N-1}(u(k) - cb_E(k)) \geq v(k) \geq \tfrac{N'}{N'+1}b_L(k) > 0 \qquad \text{on } \bar{K}, \qquad (8)$$

which otherwise follows from $a = k$ in (6)–(7). We often require u and v to be *proper*, meaning lower semicontinuous and not identically infinite. This costs little generality, since when (6)–(8) hold for non-negative functions (u, v), they continue to if u and v are replaced by their lower semicontinuous hulls.

A *competitive equilibrium* refers to a pair of measures $\epsilon, \lambda \geq 0$ and func-

tions $(u, v) \in F_0$ satisfying (4)–(8) plus the budget constraint

$$\text{equality holds } \epsilon\text{-a.e. in (6), and } \lambda\text{-a.e. in (7)} \tag{9}$$

relating (ϵ, λ) to (u, v). The economic idea behind this definition is that no individual agent (nor any group of agents which is small relative to the size of the market) can improve their outcome by choosing to match otherwise than as prescribed by ϵ and λ. Here ϵ represents an assignment of N students to each teacher, and reproduces the current distribution of adult skills in the next generation, starting from the given distribution α of student skills and educational technology $z(a, k) = (1 - \theta)a + \theta k$; the future earnings plus any non-labor utility received by the N students exactly add up to their net lifetime utilities, plus the salary of the teacher. Similarly, λ represents an assignment of N' workers to each manager, the productivity of these worker-manager teams exactly sufficing to pay the respective wages of each team member. Both the educational and the labor markets clear, and the stability constraints guarantee no adult would prefer an occupation other than the one he or she has been assigned, nor to work with anyone other than the partners prescribed by $(a, k) \in \text{spt }\lambda$ in the case of workers or managers, or by ϵ in the case of teachers. Similarly, each pair $(a, k) \in \text{spt }\epsilon$ represents a student of ability a, who cannot improve his net lifetime payoff by training with any teacher other than the one of skill k that he is paired with under ϵ.

1.2   The planner's problem and its dual

Shapley and Shubik's basic insight is that stable matching problems with transferable utility have a variational reformulation using linear programs and their duals. In [17] we observe that this insight extends from the familiar single-stage, single-sector setting of [20] [11] and [5], to steady-state multi-sector models such as the one introduced above; this is fleshed out in [16]. Denoting our education and labor market technologies by $b_\theta(a, k) = b_E((1 - \theta)a + \theta k)$ and $b_{\theta'}^L(a, k) = b_L((1 - \theta')a + \theta' k)$, the quartuple $(\epsilon, \lambda, u, v)$ forms a competitive equilibrium if and only if (u, v) attain the infimum

$$LP_* := \inf_{(u,v) \in F_0} \int_{[0,\bar{a}]} u(a)\alpha(da) \tag{10}$$

over (6)–(8), while (ϵ, λ) attain the supremum

$$LP^* := \max_{\substack{\epsilon \geq 0 \text{ and } \lambda \geq 0 \text{ on } [0,\bar{a}]^2 \\ \text{satisfying } (4)-(5)}} \int_{[0,\bar{a}] \times [0,\bar{k}]} [cb_\theta(a, k)\epsilon(da, dk) + b_{\theta'}^L(a, k)\lambda(da, dk)]. \tag{11}$$

11

We shall henceforth refer to $(u, v) \in F_0$ as *optimal* if it attains the infimum (10), and to (ϵ, λ) as *optimal* if it attains the supremum (11). Whereas the notion of competitive equilibrium relates (u, v) to (ϵ, λ) through (9), one can discuss *optimality* of (u, v) without referring to (ϵ, λ), and vice-versa. This is the first of many advantages conferred by our Shapley-Shubik-like reformation of the problem at hand.

We often use $\alpha(u)$ as a shorthand notation to denote the integral appearing in (10), which represents the average student's net lifetime utility. Similarly, $c\epsilon(b_\theta) + \lambda(b_{\theta'}^L)$ denotes the argument appearing in the supremum (11), and represents the total (non-labor + labor) utility produced by the pairings ϵ and λ. Thus if equilibrium wages $(u, v) \in F_0$ exist, they minimize the expected lifetime utility of students subject to the stability constraints. Similarly, any equilibrium matches maximize the utility $c\epsilon(b_\theta) + \lambda(b_{\theta'}^L)$ being produced our model's two sectors in each generation, subject to the market-clearing constraints (4)–(5) in steady-state. The latter can be interpreted as a social planner's problem; it is also the linear program dual to (10). Satisfaction of the budget constraint (9) follows from the absence of a duality gap: the fact $LP_* = LP^*$, which is established below under the technical hypothesis that α satisfy a doubling condition at the top skill type $\bar{a}$, meaning there exists $C < \infty$ such that

$$\int_{[\bar{a}-2\Delta a, \bar{a}]} \alpha(da) \leq C \int_{[\bar{a}-\Delta a, \bar{a}]} \alpha(da) \tag{12}$$

for all $\Delta a > 0$. A surprisingly delicate part of the proof is the inequality $LP^* \leq LP_*$ shown in Proposition 8, where (12) provides sufficient control to bound the expectation of the wages against the unknown distribution of adult skills near the top skill type; the rest of the duality argument reproduced in Appendix A is quite standard.

The variational characterization given by (10)–(11) is our starting point for the further analysis for the payoffs (u, v) and matchings (ϵ, λ) we seek. To show such competitive equilibria exist, it is enough to establish the infimum and supremum are attained. Attainment of the planner's supremum is standard, as recalled in Appendix A. It is less straightforward to show that the infimum (10) is attained, and to elucidate the properties of the extremizers for either problem. A continuity and compactness argument is complicated by the fact that the wage function v appears on *both* sides of the education sector stability constraint, and has no obvious upper bound except in $L^1(\bar{A}, \alpha)$; c.f. (8).

When minimizers (u, v) exist, it is useful to know as much structural information as we can about them, in order to analyze the properties of the corresponding equilibrium matches. In the cases for which we have been able to deduce the existence of minimizers, they turn out to be non-negative, non-decreasing, *convex* functions of $a \in [0, \bar{a}]$. The fact that the monotonicity and convexity of u and v survive limits is crucial to the analysis. Indeed, our existence strategy is to first show (10) is minimized under the additional assumption of convexity and monotonicity for u and v, and then to show that this additional constraint does not bind for the minimizing (u, v), which must therefore optimize the original problem of interest. In the absence of an atom at the top skill type, $\alpha[\{\bar{k}\}] = 0$, it seems possible a priori that both $u(k)$ and $v(k)$ diverge to $+\infty$ as $k \to \bar{k}$, without violating boundedness of the expected value $LP_* = \alpha(u)$. Although Theorem 16 tends to rule out this possibility, giving conditions instead for the gradients $u'(a)$ and $v'(k)$ to diverge, for the intermediate analysis it is useful to let $A = [0, \bar{k}[= K$ denote a half-open interval where we can assume u and v are real valued.

In addition to (N, θ) and (N', θ'), dimensionless parameters such as $\bar{b}'_L / \underline{b}'_L \geq 1$ and $c \geq 0$ govern the behavior displayed by the model. Here $\underline{b}'_L$ is from (2) and

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

so $\bar{b}'_L / \underline{b}'_L$ indexes the relative impact of an increase in skill on labor productivity at the top versus the bottom of the skills market, while c measures the relative importance of any other satisfactions derived from cognitive skills apart from the returns to labor which they help to enhance. Such satisfactions could be intrinsic, or they could represent externalities that cognitive skills and education provide, such as social status or — as in early drafts of [17] — marital prospects. We can also remove this effect from the model by setting $c = 0$. However, to implement the existence strategy outlined above, it turns out to be technically easier to analyze the case $c > 0$ first, and then take the limit $c \to 0$ if desired. Many but not all of our structural results such as uniqueness, specialization, and positive assortativity also survive this limit; see Proposition 7 and Theorem 15 for example.

We shall also investigate occupational specialization by cognitive skill, showing $\min\{N'\theta', N\theta\} \geq \bar{b}'_L / \underline{b}'_L$ implies that the highest types become teachers, while the lowest types become either workers or teachers, but not managers. More refined statements appear in Proposition 7. For continuously

distributed skill types, we show that a pyramid can form in the education sector, sometimes leading to divergence of wage gradients at the highest skill type when $N\theta \geq 1$, meaning the span of control at each node in the pyramid is large enough. More explicitly, under suitable conditions Theorem 16 asserts that as $k \to \bar{k}$,

$$v'(k) \sim \begin{cases} const|\bar{k} - k|^{-1 - \frac{\log \theta}{\log N}} & \text{for} \quad N\theta > 1, \\ c\bar{b}'_E/(\frac{1}{N\theta} - 1) & \text{for} \quad N\theta < 1, \end{cases}$$

so a phase transition occurs at $N\theta = 1$. A less involved investigation of an analogous pyramid structure was given by Becker and Murphy [2, §VII], in a different production model incorporating the cost of acquiring knowledge and assuming steady-growth as opposed to steady-state. To produce a similar pyramid in the labor sector, our model would need to be modified to permit managers to manage other managers — as in Garicano [9] with Rossi-Hansberg [10] — instead of being forced to manage only workers whose productivity is inherently limited. If such a modification to our model could be achieved, it would have the potential to complement existing models for executive compensation such as Gabaix and Landier's [7], which rely instead on comparing given tail behaviors of the distributions of company size and managerial talent.

Finally, Corollary 9 characterizes the optimizers in the primal and dual problems (10)–(11). Theorem 15 provides sufficient conditions for uniqueness of (ϵ, λ), and discusses in what sense (u, v) are also unique. It gives conditions guaranteeing the optimal pairings λ of workers with managers and ϵ of teachers with students are positive assortative in cognitive skills, meaning spt λ and spt ϵ are non-decreasing subsets of the plane. This monotonicity is intimately tied to the convexity of wages v as a function of $k \in [0, \bar{k}]$ asserted above.

2 Analysis

2.1 Terminology and notation

In this section, we introduce terminology and notation that will be useful for dealing with functions which need neither be smooth nor bounded, and with the measures which arise naturally as their duals.

Given any convex set $B \subset \mathbf{R}^n$, a function $u : B \longrightarrow \mathbf{R} \cup \{+\infty\}$ is said to be continuous if it is upper and lower semicontinuous. It is said to be *Lipschitz* with Lipschitz constant L if either u is identically infinity or else if

$$L := \sup_{B \ni x \neq y \in B} \frac{|u(x) - u(y)|}{|x - y|}$$

is finite. It is said to be *semiconvex* with semiconvexity constant C if the function $x \in B \longmapsto u(x) + C|x|^2/2$ is convex. It is said to be *locally* Lipschitz (respectively semiconvex) on B, if u is Lipschitz (respectively semiconvex) on every compact convex subset of B. Locally Lipschitz (respectively semiconvex) functions are once (respectively twice) differentiable Lebesgue a.e. In addition, locally semiconvex functions fail to be once differentiable on a set of Hausdorff dimension at most $n - 1$.

By *support* of a Borel measure $\alpha \geq 0$ on $\mathbf{R}^m$, we mean the smallest closed subset $\operatorname{spt} \alpha \subset \mathbf{R}^m$ of full mass: $\alpha[\mathbf{R}^m \setminus \operatorname{spt} \alpha] = 0$. The *push-forward* $f_\#\alpha$ of α through a Borel map $f : \mathbf{R}^m \longrightarrow \mathbf{R}^n$ is a Borel measure defined by $(f_\#\alpha)[Z] = \alpha[f^{-1}(Z)]$ for each $Z \subset \mathbf{R}^n$. We say α has no *atoms* if $\alpha[\{x\}] = 0$ for each $x \in \mathbf{R}^m$. A measure ϵ on $\mathbf{R}^2$ is said to be *positive assortative* if $\operatorname{spt} \epsilon$ forms a non-decreasing subset in the plane: i.e. if $(a' - a)(k' - k) \geq 0$ for all $(a, k), (a', k') \in \operatorname{spt} \epsilon$. We use $\alpha|_B$ to denote the restriction $\alpha|_B(Z) = \alpha[Z \cap B]$ of α to $B \subset \mathbf{R}^m$, and H^n to denote Lebesgue measure on $\mathbf{R}^n$.

2.2   The educational pyramid

In this section, we discuss the extent to which we can expect optimizers (u, v) to the minimization (10) to be smooth, at least away from the top skill type $\bar{k}$. We then apply these results to elucidate the nature of the pyramid structure which can form in the education sector.

Given $(u, v) \in F_0$ feasible for the infimum (10), use $b_{\theta'}^L(k', k) := b_L((1 - \theta')k' + \theta'k)$ and $z(a, k) = (1 - \theta)a + \theta k$ to define the wages implicitly available to an individual of cognitive skill k employed as a worker, manager, or teacher, respectively:

$$
\begin{align}
v_w(k) \quad &:= \quad \sup_{k' \in \bar{K}} b_{\theta'}^L(k, k') - \tfrac{1}{N'}v(k'), \tag{13}\\
v_m(k) \quad &:= \quad N' \sup_{k' \in \bar{K}} b_{\theta'}^L(k', k) - v(k'), \qquad \text{and} \tag{14}\\
v_t(k) \quad &:= \quad N \sup_{a \in \bar{A}} cb_E(z(a, k)) + v(z(a, k)) - u(a), \tag{15}
\end{align}
$$

15

where we complete definition (15), and later (18), with the convention

$$\infty - \infty := \infty. \tag{16}$$

The suprema (13)–(14) are attained when u and v are proper (hence lower semicontinuous), and the same holds true for (15) if, in addition, v is convex non-decreasing (hence continuous).

Clearly feasibility (6)–(8) implies $v \geq \bar{v} := \max\{v_w, v_m, v_t\}$. When equality holds — as we shall see that it does (Theorem 13) for some v minimizing (10) — this implies strong conclusions. For example, v_w and v_m inherit Lipschitz and convexity properties from b_L by an envelope argument (Lemma 2), which v also inherits wherever it coincides with v_w or v_m. Something similar is true but more subtle to verify for v_t (and hence for $\bar{v}$) — because of the recursive structure built into the educational pyramid; in (15), as opposed to (13)–(14), this is manifested in the fact that the k dependence in the argument of the supremum involves the unknown function v. As another example, when $N'\theta'$ and $cN\theta$ are large enough, Proposition 7 derives complete specialization of types into low (workers), medium (managers), and high (teachers). This at least tells us the role of κ-a.e. agent, leaving the distribution $\kappa = \kappa_w + \kappa_m + \kappa_t$ of adults as the only unknown. Here $\kappa_w = \lambda^1$, $\kappa_m = \lambda^2/N'$ and $\kappa_t = \epsilon^2/N$ are measures representing the distribution of worker, manager, and teacher types, and have respective masses $\kappa_w[\bar{K}] = \frac{(N-1)N'}{N(N'+1)}$, $\kappa_m[\bar{K}] = \frac{N-1}{N(N'+1)}$ and $\kappa_t[\bar{K}] = \frac{1}{N}$. If $c = 0$ but $\min\{N'\theta', N\theta\} \geq \bar{b}'_L/\underline{b}'_L$, the same proposition yields more subtle conclusions.

A first insight into the educational pyramid is provided by the following example.

Example 1 (Gurus) *Fix the number of students each teacher can teach or the number of workers each manager can manage to be $N = N' = 10$. If our probability measure κ represents the skill distribution for a population of 110 adults, 90 of them will be workers, managed by 9 managers, and 11 of them will be teachers. Nine of these $11 = 9 + 1 + 1$ will specialize in teaching workers, one in teaching teachers, and one in teaching a combination of 9 managers and 1 teacher. We may remember this with the mnemonic $110 = 90 + 9 + (9 + 1 + 1)$. On the other hand, if κ represents the skill distribution for a population of $11000 = 9000 + 900 + (900 + 90 + (90 + 9 + (9 + 1 + 1)))$ adults, 9000 of them will be workers, managed by 900 managers, while 1100 of them will be teachers. Of these, 900 will teach workers, 90 will teach managers,*

16

*and 110 will teach teachers. Within these 110, there is further specialization
as before: 90 will teach teachers who teach workers, 9 will teach teachers who
teach managers, and 11 will teach teachers who teach teachers. Within this
11, 9 teach worker-teacher-teachers, 1 teaches manager-teacher-teachers, and
1 teaches only teacher-teacher-teachers. These last two may be thought of as
'gurus'. One of the questions at stake is whether the salaries of these gurus
can grow without bounds as the population size grows.*

Next we recall without proof a well-known result which can be proved as
in [8]:

Lemma 2 (Upper envelopes inherit derivative bounds) *If $f : A \times K \longrightarrow$
$\mathbf{R}$ is locally Lipschitz in $a \in A$, uniformly in $k \in K$, then $g(a) = \sup_{k \in K} f(a, k)$
is locally Lipschitz and for each $\delta > 0$ we have the bounds*

$$\inf_{k \in K, |a'-a|<\delta} f_a(a', k) \leq g'(a) \leq \sup_{k \in K, |a'-a|<\delta} f_a(a', k)$$

*in the pointwise a.e. senses. Similarly, if f is locally semiconvex in $a \in A$,
uniformly in $k \in K$, then $g(a)$ is locally semiconvex and obeys the bound*

$$g''(a) \geq \inf_{k \in K, |a'-a|<\delta} f_{aa}(a', k)$$

in the same senses. Here $f_a := \frac{\partial f}{\partial a}$ and $f_{aa} := \frac{\partial^2 f}{\partial a^2}$.
*If $f(a', \cdot)$ extends upper semicontinuously to $\bar{k}$ for some $a' \in A$, al-
lowing $f(a', \bar{k}) = -\infty$ as a possible value, there exists $k' \in \bar{K}$ such that
$g(a') = f(a', k')$; if $g(a)$ is differentiable at $a' \in]\underline{a}, \bar{a}[$, the envelope theorem
then yields $g'(a') = f_a(a', k')$ provided $f(\cdot, k')$ is locally semiconvex near a';
similarly, $g''(a') \geq f_{aa}(a', k')$ provided both functions admit a second order
Taylor expansion with respect to a at a'.*

Definition 3 (Supermodular) *Given intervals $I, J \subset \mathbf{R}$, a function $f :
I \times J \longrightarrow \mathbf{R}$ is* weakly supermodular *if*

$$f(a, k) + f(a', k') \geq f(a, k') + f(a', k) \tag{17}$$

for all $1 \leq a < a' \in I$ and $1 \leq k < k' \in J$. It is strictly *supermodular if, on
the same domain, the inequality (17) remains strict.*

Remark 4 (Supermodular extensions) *It is elementary to check that a function f which is weakly (or strictly) supermodular on $A \times K$ and has an upper semicontinuous extension to $\bar{A} \times \bar{K}$ that is continuous and real-valued except perhaps at $(\bar{a}, \bar{k})$, is weakly (respectively strictly) supermodular on $\bar{A} \times \bar{K}$.*

Throughout we assume θ, θ', N, N' and $\bar{a} = \bar{k}$ are positive parameters with $\max\{\theta, \theta'\} < 1 \leq N$, and set $c \geq 0$ and $A = [0, \bar{a}[= K$. Unless otherwise noted, the utilities $b_E, b_L \in C^1(\bar{K})$ of education and labor have positive lower bounds $\underline{b}'_{E/L}$ and $\underline{b}''_{E/L}$ on their first two derivatives (1)–(3), hence are strictly convex and increasing.

Lemma 5 (Structure of wage functions) *Let $v : K \longrightarrow \mathbf{R}$ be convex non-decreasing, with $v(\bar{k}) \geq \limsup_{k \to \bar{k}} v(k)$. Then $f(a, k) = v(z(a, k))$ will be weakly supermodular on $\bar{A} \times \bar{K}$, and strictly supermodular unless the convexity of v fails to be strict.*

Set $z(a, k) = (1 - \theta)a + \theta k$, $b_\theta = b_E \circ z$ and $b^L_{\theta'}(k', k) = b_L((1 - \theta')k' + \theta'k)$ where $b_{E/L} \in C^1(\bar{K})$ satisfy (1)–(3). Then the student payoff defined by

$$u(a) = \sup_{k \in \bar{K}} cb_E(z(a, k)) + v(z(a, k)) - \tfrac{1}{N}v(k) \tag{18}$$

is also convex non-decreasing on K and satisfies $\frac{u'}{1-\theta} \geq c\underline{b}'_E + \inf_k v'(k)$ and $\frac{u''}{(1-\theta)^2} \geq c\underline{b}''_E + \inf_k v''(k)$ pointwise a.e.

The worker, manager, and teacher wage functions $v_{w/m/t}$ defined by (13)–(16) and their maximum $\bar{v} := \max\{v_w, v_m, v_t\}$ are then monotone and convex on $\bar{K}$, real-valued on K, and satisfy $\bar{v}' \geq \min\{(1 - \theta')\underline{b}'_L, N'\theta'\underline{b}'_L, N\theta(c\underline{b}'_E + \inf_k v'(k))\}$ and $\bar{v}'' \geq \min\{(1-\theta')^2\underline{b}''_L, (\theta')^2 N'\underline{b}''_L, N\theta^2(c\underline{b}''_E + \inf_k v''(k))\}$ pointwise a.e.

Proof. First note that convexity and monotonicity imply v is continuous throughout $K = [0, \bar{k}[$. Any convex $v \notin C^2$ can be approximated by convex $v_i \in C^2$ locally uniformly on $]0, \bar{k}[$, with $v'_i \to v'$ pointwise a.e. (and $v''_i \to v''$ weakly).

Now let $f(a, k) = cb_E(z(a, k)) + v(z(a, k))$. For each fixed $\tilde{k}$, we see f is convex non-decreasing as a function of $a \in \bar{A}$, so the same must be true of the supremum $u(a) = \sup_{k \in \bar{K}} f(a, k) - v(k)/N$. Supposing for simplicity that v and b_E are $C^2(\bar{A})$, from

$$f_a(a, k) = (cb'_E(z(a, k)) + v'(z(a, k))) \, z_a(a, k)$$

18

and $0 \leq z(a, k) = (1 - \theta)a + \theta k$ we compute bounds

$$c\underline{b}'_E + \inf v' \leq \frac{f_a(a, k)}{1 - \theta} \leq cb'_E(z(a, \bar{k})) + v'(z(a, \bar{k}))$$

and

$$\frac{f_{aa}(a, k)}{(1 - \theta)^2} = cb''_E(z(a, k)) + v''(z(a, k))$$

$$\geq c\underline{b}''_E + \inf v''$$

which are uniform in $k \in \bar{K}$. The analogous bounds for u follow from Lemma 2.

So far, we have been working under the assumption that v and b_E are $C^2(\bar{K})$. More generally, v and b_E can be approximated uniformly on compact subsets of K by C^2 functions v^i and b^i_E satisfying the same hypotheses as v and b_E. As a result, $f^i(a, k) := b^i_E(z(a, k)) + v^i(z(a, k))$ converges to f uniformly on compact subsets of $\bar{A}^2 \setminus \{(\bar{a}, \bar{k})\}$, and $u^i(a) := \sup_{k \in \bar{K}} f^i(a, k) - \frac{1}{N}v(k)$ converges uniformly to u on compact subsets of A. Thus u inherits the same Lipschitz and local semiconvexity bounds as u^i in the distributional (and hence pointwise a.e.) sense. See (34) for the distributional definition of the inequality $v''_i \geq g$.

On the other hand, $f(a, k; \theta) = f(k, a; 1 - \theta)$ is symmetrical, and $v_t(k)/N$ is defined by essentially the same formula as $u(a)$, but with the roles of $a \leftrightarrow k$ and $\theta \leftrightarrow 1 - \theta$ interchanged. Thus v_t is also locally Lipschitz and convex on K, and satisfies $v'_t \geq N\theta(c\underline{b}'_E + \inf_b v'(b))$ and $v''_t \geq N\theta^2(c\underline{b}''_E + \inf_b v''(b))$.

Turning to v_w and v_m, we apply Lemma 2 but with $f(a, k) := b^L_{\theta'}(a, k) = b_L((1 - \theta')a + \theta'k)$, which is jointly convex and increasing in each variable. Approximating b_L by $C^2(\bar{K})$ functions if necessary, shows bounds

$$\underline{b}'_L = b'_L(0^+) \leq \frac{f_a(a, k)}{1 - \theta'} = b'_L((1 - \theta')a + \theta'k) \leq b_L((1 - \theta')a^- + \theta'\bar{k}) \leq \bar{b}'_L$$

and

$$\frac{f_{aa}(a, k)}{(1 - \theta')^2} = b''_L(z'(a, k)) \geq \underline{b}''_L$$

are inherited by the convex increasing functions v_w and $\frac{1}{N'}v_m$ on K. Thus $\bar{v} = \max\{v_w, v_m, v_t\}$ is convex, non-decreasing, locally Lipschitz and inherits the bounds $\bar{v}' \geq \underline{\min}\{(1 - \theta')\underline{b}'_L, N'\theta'\underline{b}'_L, N\theta(c\underline{b}'_E + \inf_b v'(b))\}$ and $\bar{v}'' \geq \min\{(1 - \theta')^2\underline{b}''_L, N'(\theta')^2\underline{b}''_L, N\theta^2(c\underline{b}''_E + \inf_b v''(b))\}$ on K.

Finally, setting $f(a, k) = v(z(a, k))$, using convexity of $v \in C^2(\bar{K})$ we compute

$$
\begin{aligned}
f(a_0, k_0) \; + \;\; & f(a_1, k_1) - f(a_0, k_1) - f(a_1, k_0) \\
= \;\; & (1 - \theta)\theta \int_{a_0}^{a_1} \int_{k_0}^{k_1} v''((1 - \theta)a + \theta k) da dk \\
\geq \;\; & 0
\end{aligned}
$$

for $a_0 < a_1$ and $k_0 < k_1$. For $v \notin C^2$, the same formulas hold by smooth approximation of $v = \lim v^i$. Strict inequality holds unless $v'' = 0$ throughout $]z(a_0, k_0), z(a_1, k_1)[$. This yields the (strict) supermodularity (17) asserted. ∎

We are now in a position to prove our first main result, which describes how occupations are allocated according to cognitive skill. It depends on the relative size of various parameters: the teaching capacity N (resp. N') and effectiveness θ (resp. θ') of teachers (resp. managers) in the population in question, the range $\bar{k}$ of cognitive skills, and the relative utility $c \geq 0$ of cognitive achievement compared to wages. When $N'\theta'$ and $cN\theta$ are large enough it turns out that there is a complete ordering (a)-(b) of skill types between workers, managers, and teachers in a steady-state economy. However $N\theta \geq 1$ is enough to ensure that no student studies with a teacher whose cognitive skills are inferior to their own (d), while $N'\theta'$ and $N\theta$ large enough guarantee that the most cognitively skilled types all become teachers (c) (though not that all teachers have high cognitive skills). This conclusion will help us to establish the phase transition from bounded to unbounded wage gradients that these teachers enjoy as $N\theta$ passes through 1 (in section 2.6). The possibility (f) that the number $d(k)$ of types of academic descendants a teacher can have may grow without bound as $k \to \bar{k}$ foreshadows the analysis there.

Remark 6 *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 7 (Specialization by type; the educational pyramid) *Fix $K = [0, \bar{k}[$ with $\bar{k} > 0$, and $c \geq 0$. Suppose $u, v : K \longrightarrow \mathbf{R}$ are convex, non-decreasing, and satisfy $v = \max\{v_w, v_m, v_t\}$.*

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) $N'\theta' > (1 - \theta') \sup_{k\in K} b'_L(1 - \theta')k + \theta'\bar{k}^-)/b'_L(\theta'k^+)$ then all of the worker types lie weakly below all of the manager types.

If (c) $N\theta \geq \sup_{0\leq z\leq k} b'_L((1 - \theta')z^- + \theta'\bar{k})/(b'_L(\theta'z^+) + \frac{c}{N'\theta'}b'_E(z^+))$ 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 $(a,k) \in K \times K$ where $v(z(a,k)) = v_m(z(a,k))$, then $v > v_m$ on $]k,\bar{k}]$. 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 $a \in K$ 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 $k \in K$ will display one of at most finitely many $d = d(k)$ distinct skill types, unless differentiability of v fails at k. However, $d(k)$ and $v'(k) \geq (N\theta)^{d(k)}v'(0^+)$ may diverge as $k \to \bar{k}$.

Proof. Lemma 5 asserts convexity of $v_{w/m/t}$, hence one-sided differentiability everywhere, and two sided-differentiability except perhaps at countably many points. At points $k \in]0,\bar{k}[$ of differentiability, Lemma 2 (the envelope theorem) allow us to estimate the wage gradients

$$v'_w(k) = \qquad (1 - \theta')b'_L(z'(k,k'_m)) \qquad \in (1 - \theta')]\underline{b}'_L, \bar{b}'_L[\qquad (19)$$

$$v'_m(k) = \qquad N'\theta'b'_L(z'(k'_w,k)) \qquad \in \quad N'\theta']\underline{b}'_L, \bar{b}'_L[, \qquad (20)$$

$$v'_t(k) = \quad N\theta(cb'_E(z(a,k)) + v'(z(a,k))) \quad \geq \quad N\theta cb'_E(\theta k) \qquad (21)$$

where k'_m, k'_w and a are the respective points at which the suprema (13)–(15) (or their extension to $\bar{K}$) are attained. Such points exist in $\bar{K}$ according to the same lemma; we can extend $v(\bar{k}) = v(\bar{k}^-) := \lim_{k\uparrow\bar{k}} v(k)$ and $u(\bar{k}) \in \mathbf{R} \cup \{+\infty\}$ similarly without changing $v_{w/m/t}$. Consideration of the worst-case scenario $k'_w = 0$ and $k'_m = \bar{k}$ in (19)–(20) shows if (b) holds that $v'_m(k) < v'_w(k)$ at each point k where both derivatives are defined. Then the locally Lipschitz function $v_m - v_w$ is strictly increasing. Since this function is non-positive on $\{k \mid \bar{v} = v_w\}$ and non-negative on $\{k' \mid \bar{v} = v_m\}$, the first set must lie entirely to the left of the second, as desired.

Estimating the wage gradient for a teacher of type $k_0 = k \in K$ is more subtle, due to the recursive nature of formula (21). Since the student of ability $a_1 = a$ taught by k_0 winds up with cognitive skill $k_1 = (1-\theta)a_1 + \theta k_0 =$

$z(a_1, k_0)$, we find

$$v'_t(k_i) \quad = \quad N\theta(cb'_E(k_{i+1}) + v'(k_{i+1})) \tag{22}$$

for $i = 0$, assuming differentiability of v_t at k_0. Differentiability of v and b_E at k_1 (and also of $v_t \leq v$) follows from convexity, since replacing k by k_0 produces equality in $u(a_1) + \frac{1}{N}v_t(k) - v(z(a_1, k)) - cb_E(z(k, a_1)) \geq 0$: the first-order condition

$$(v' + cb'_E)(z(a_1, k_0)^-) \geq \quad \frac{1}{N}v'_t(k_0)/z_k(a_1, k_0) \quad \geq (v' + cb'_E)(z(a_1, k_0)^+)$$

forces the one-sided derivatives $(v' + cb'_E)(k_1^-) \leq (v' + cb'_E)(k_1^+)$ to agree. From (22) we have $v'_t(k_0) \geq N\theta c\underline{b}'_E$, which dominates $(1 - \theta')\bar{b}'_L$ and $N'\theta'\bar{b}'_L$ in case (a). Since $v_{w/m/t}$ are locally Lipschitz, monotonicity of $v'(k)$ then combines with the estimates (19)–(20) already established to show all teacher types k_0 are at least as high as the highest worker and manager types.

From (d) $N\theta \geq 1$ and (22) we conclude $v'(k_1) \leq v'_t(k_0)$, and this inequality is strict if (e) also holds, in which case every student studies with a teacher more skilled than himself, or — what is equivalent in our model — no student (except the very top type $a = \bar{a}$) becomes as skilled as his teacher.

Next, assume as in case (c), that a teacher of type $k \in K$ teaches a student of type a who becomes a manager of type $z = z(a, k)$. Since $v \geq v_m$ with equality at z, we have $v'(z^+) \geq v'_m(z^+)$. Analogously to (20)–(21) we find

$$\begin{aligned} \frac{1}{N\theta}v'_t(k^+) \quad &\geq \quad cb'_E(z^+) + v'_m(z^+) \\ &\geq \quad cb'_E(z^+) + N'\theta'b'_L((1 - \theta')k_w + \theta'z^+), \\ &\geq \quad cb'_E(z^+) + N'\theta'b'_L(\theta'z^+) \end{aligned}$$

with equality holding in the first two estimates if all the derivatives in question exist. On the other hand,

$$v'_m(\bar{k}) \quad \leq \quad N'\theta'b'_L((1 - \theta')z^- + \theta'\bar{k})$$

since (b) implies the worker types all lie below the manager type z. Hypothesis (c) now yields $v'_t(k^+) \geq v'_m(\bar{k})$, and the convexity of v_t and strict convexity of v_m shown as in Lemma 5 then imply $v'_t > v'_m$ on $]k, \bar{k}[$. Vanishing of the non-negative function f at (a, k) implies $v_t(k) = v(k) \geq v_m(k)$, whence the desired conclusion $v_t > v_m$ follows on $]k, \bar{k}]$ by integration.

Case (f) is more delicate, and our conclusions for it are more involved. If the student a_1 above elects to become a worker or manager, we can estimate (22) using (19)–(20). However, if the student becomes a teacher whose students' innate ability a_2 allows them to acquire human capital $k_2 = z(a_2, k_1)$, we must iterate (22). And if these students in turn become teachers teaching students of ability a_3 to acquire human capital $k_3 = z(a_3, k_2)$, we must iterate again, and continue iterating until the student of ability a_d who acquires human capital $k_d = z(a_d, k_{d-1})$ elects to become a worker or manager instead of another teacher. Assuming (d)–(f), we claim this occurs for some finite d: otherwise the skills $k_{i+1} < k_i$ converge to some $k_\infty \in K$, for which the limit of (22) yields an identity $(\frac{1}{N\theta} - 1)v'(k_\infty^+) = cb'_E(k_\infty^+)$ equating quantities with different signs. Recalling $v'(k_\infty^+) \geq 0$ and $c \geq 0$, hypothesis (f) asserts at least one of these inequalities is strict, while (d) asserts $N\theta \geq 1$. Unless $N\theta = 1$ and $c = 0$, this contradicts the limiting identity. But $N\theta = 1$ and $c = 0$ contradicts (e). Thus the sequence k_i terminates at some finite d (which depends on k_0).

At this point we have

$$
\begin{aligned}
v'_t(k) &= N\theta\left(cb'_E(k_1) + N\theta\left(cb'_E(k_2) + N\theta\left(\ldots + N\theta\left(cb'_E(k_d) + v'(k_d)\right)\right)\right)\right) \\
&\geq \begin{cases} \dfrac{1 - (N\theta)^d}{1 - N\theta}N\theta cb'_E(\theta^d k) + (N\theta)^d v'(\theta^d k) & \text{if } N\theta \neq 1 \\ dcb'_E(\theta^d k) + v'(\theta^d k) & \text{if } N\theta = 1, \end{cases}
\end{aligned}
\tag{23}
$$

where we have summed the geometric series and estimated $k_d \geq \theta^d k_0$. ∎

2.3 Characterization of optimality

When we turn to the question of existence of optimal payoffs (u, v) for the linear program (10), our strategy will be to perform the minimization under the additional assumption that u and v are convex non-decreasing, and then to show these additional constraints are non-binding at the optimum, thus have no effect on the outcome. Convexity and monotonicity provide the requisite compactness for extracting limits from minimizing sequences. In order to show these constraints are non-binding however, it is necessary to control the payoff $u(a)$ on the full interval $A = [0, \bar{a}[$, and not only on $\operatorname{spt}\alpha$. Similarly it is necessary to control v on the full interval $K = A$, and not only on the support of the unknown distribution κ of adult skills. Since the

original problem is largely insensitive to the values of u and v outside $\operatorname{spt}\alpha$ and $\operatorname{spt}\kappa$, we introduce a perturbed version of the problem to provide this control: for each $\delta > 0$ set

$$LP(\delta)_* := \inf_{(u,v)\in F_\delta} \delta\langle u+v\rangle_A + \int_{[0,\bar a]} u(a)\alpha(da) \qquad (24)$$

where $\langle v\rangle_A := \frac{1}{H^1(A)}\int_A vdH^1$ denotes the Lebesgue average of v over A. Here $F_\delta = F_0$ denotes the same feasible set as before, with a subscript denoting only the possible dependence of the constant $c = c_\delta$ in (6) on $\delta > 0$, (allowing us to approximate either limit $c > 0$ or $c = 0$ by choosing the sequence $c_\delta = c$ or $c_\delta = \delta$.) Also u (and hence v) $\in L^1(\bar A, \alpha)$, and if $\delta > 0$ then $u, v \in L^1(A, H^1)$. We must first solve the perturbed problem (24) and then extract the $\delta \to 0$ limit. For the latter endeavor and to characterize the optimizers, it will be crucial to know $LP(\delta)_*$ is in fact dual to

$$LP(\delta)^* := \max_{\epsilon,\lambda\geq 0 \text{ on } \bar A\times\bar K \text{ satisfying } (26)-(27)} c_\delta\epsilon(b_E \circ z) + \lambda(b_L \circ z') \qquad (25)$$

where

$$\epsilon^1 = \alpha + \frac{\delta}{|A|}H^1|_A \qquad (26)$$

and

$$\lambda^1 + \frac{1}{N'}\lambda^2 + \frac{1}{N}\epsilon^2 = z_\#\epsilon + \frac{\delta}{|K|}H^1|_K. \qquad (27)$$

Let us begin by verifying $LP(\delta)^* \leq LP(\delta)_*$. This would be standard if the primal infimum were restricted to continuous bounded functions $u, v \in C(\bar A)$, as in Appendix A where the reverse inequality and attainment of the dual maximum are verified. However, a priori we know only that u, v differ from continuous bounded functions by non-decreasing functions, and even a posteriori we do not know whether or not minimizers of (10) or (24) are bounded at $\bar k$. We have only the conditional result of Theorem 16 to suggest that they are. Thus we are forced to work in a space which includes unbounded functions, and to check their inclusion does not spoil the otherwise elementary duality inequality $LP(\delta)^* \leq LP(\delta)_*$.

Proposition 8 (Easy direction of duality for unbounded functions)
Fix δ, c_δ non-negative and $\theta, \theta', N, N', \bar a = \bar k$ positive with $\max\{\theta, \theta'\} < 1 \leq N$. Let α be a Borel probability measure on $\bar A$, where $A = [0, \bar a[= K$, and

define $z(a,k) = (1-\theta)a + \theta k$, $b_\theta = b_E \circ z$ and $b_{\theta'}^L(a,k) = b_L((1-\theta')a + \theta'k)$ where $b_{E/L} \in C(\bar{K})$. If Borel measures (ϵ, λ) and Borel functions $(u,v) \in F_\delta$ are feasible for the primal and dual problems (24)–(25), with $u \in L^1(\bar{A}, \alpha)$ and $u\delta, v\delta \in L^1(A, H^1)$, then $\alpha(u) + \delta\langle u+v \rangle_A \geq c_\delta \epsilon(b_\theta) + \lambda(b_{\theta'}^L)$ provided $v \in L^1(\bar{A}, z_\#\epsilon)$. If α satisfies the doubling condition (12), then $v \in L^1(\bar{A}, z_\#\epsilon)$.

Proof. Taking feasible pairs (ϵ, λ) of measures and $(u,v) \in F_\delta$ of functions with $u \in L^1(\bar{A}, \alpha)$ and $u\delta, v\delta \in L^1(A, H^1)$, the stability constraint for the education sector implies

$$u(a) - c_\delta b_E(z(a,k)) \geq v(z(a,k)) - \tfrac{1}{N}v(k), \qquad (28)$$

on $\bar{A} \times \bar{K}$, and the left hand side is in $L^1(\bar{A}^2, \epsilon)$. Thus

$$+\infty \;>\; \alpha(u) - c_\delta \epsilon(b_\theta) + \delta\langle u+v \rangle_A$$
$$\geq\; \langle \delta v \rangle_K + \int_{\bar{A} \times \bar{K}} [v(z(a,k)) - \tfrac{1}{N}v(k)]\epsilon(da, dk) \qquad (29)$$

since $\epsilon^1 = \alpha + \frac{\delta}{|A|}H^1|_A$. On the other hand, the steady state constraint $z_\#\epsilon + \frac{\delta}{|K|}H^1|_K = \lambda^1 + \frac{1}{N'}\lambda^2 + \frac{1}{N}\epsilon^2$ combines with the stability constraint $v(a) + \frac{1}{N'}v(k) \geq b_{\theta'}^L(a,k)$ for the labor sector to imply

$$\langle \delta v \rangle_K + \int_{\bar{K}} vd(z_\#\epsilon - \tfrac{1}{N}\epsilon^2) \;=\; \int_{\bar{K}} vd(\lambda^1 + \tfrac{1}{N'}\lambda^2) \qquad (30)$$
$$\geq\; \int_{\bar{A} \times \bar{K}} b_{\theta'}^L d\lambda \qquad (31)$$
$$>\; 0.$$

Now if $v \in L^1(\bar{A}, z_\#\epsilon)$ we can equate the right hand side of (29) with the left hand side of (30) to obtain the first stated conclusion.

We must still show that the doubling (12) of α at $\bar{a}$ implies $0 \leq v \in L^1(\bar{A}, z_\#\epsilon)$. Recall that $(u,v) = (u_0 + u_1, v_0 + v_1)$ with $u_0, v_0 \in C(\bar{A})$ and $u_1, v_1 : \bar{A} \longrightarrow [0, \infty]$ non-decreasing (in fact strictly increasing without loss of generality). Since v_0 is bounded there is no question about its integrability. We shall use $v \leq u$ from (8) and $u \in L^1(\bar{A}, \alpha)$ to deduce $v_1 \in L^1(\bar{A}, \kappa)$ for $\kappa := z_\#\epsilon$. Since v_1 is strictly increasing, $v_1^{-1}(y) \in \mathbf{R} \cup \{\pm\infty\}$ can be defined unambiguously. Lemma 14, the doubling condition (12), and the layer-cake representation [12] of the Lebesgue integral imply

$$\int_{\bar{K}} v_1(k)\kappa(dk) \;=\; \int_0^\infty \kappa[v_1^{-1}[y, \infty]]dy$$

$$= \int_0^\infty \kappa[\bar{a} - (\bar{a} - v_1^{-1}(y)), \bar{a}]dy$$

$$\leq \int_0^\infty \alpha[\bar{k} - \frac{1}{1-\theta}(\bar{a} - v_1^{-1}(y)), \bar{a}]dy$$

$$\leq C^{\frac{1}{\theta}-1} \int_0^\infty \alpha[v_1^{-1}(y), \bar{a}]dy \tag{32}$$

for some $C < \infty$. On the other hand, $v_1 \leq u_0 + u_1 - v_0 \leq u_1 + const$ yields $u_1^{-1}(y - const) \leq v_1^{-1}(y)$, so

$$\int_0^\infty \alpha[u_1^{-1}(y), \bar{a}]dy = \int u_1(a)\alpha(da) < +\infty$$

implies finiteness of (32) and completes the proof that $v_1 \in L^1(\bar{K}, \kappa)$. ∎

Corollary 9 (Characterizations of optimality) *Fix $\delta, c_\delta, \theta, \theta', N, N', z, b_\theta, b_{\theta'}^L$ and $\bar{k} = \bar{a}$ as in Proposition 8, and a Borel probability measure α on $\bar{A}$ satisfying (12), where $A = [0, \bar{a}[= K$ and $\bar{a} \in \mathrm{spt}\,\alpha$. A pair of feasible measures $\epsilon, \lambda \geq 0$ on $\bar{A}^2$ maximizes the dual problem (25) if there exist feasible $(u, v) \in F_\delta$ such that $\alpha(u) + \delta\langle u + v\rangle_A = c_\delta\epsilon(b_\theta) + \lambda(b_{\theta'}^L)$.*

Conversely, $(u, v) \in F_\delta$ minimize the primal problem if and only if there exist $\epsilon, \lambda \geq 0$ feasible for the dual problem such that $\alpha(u) + \delta\langle u + v\rangle_A = c_\delta\epsilon(b_\theta) + \lambda(b_{\theta'}^L)$.

For feasible pairs in the given spaces, $\alpha(u) + \langle u + v\rangle_A = c_\delta\epsilon(b_\theta) + \lambda(b_{\theta'}^L)$ is equivalent to the assertions $\epsilon(f) = 0 = \lambda(g)$ where $f(a, k) = u(a) + \frac{v(k)}{N} - c_\delta b_\theta(a, k) - v(z(a, k)) \geq 0$ and $g(k', k) = v(k') + \frac{v(k)}{N} - b_{\theta'}^L(k', k) \geq 0$ on $\bar{A} \times \bar{K}$.

Proof. Let (ϵ, λ) be a pair of feasible measures for the dual problem, and $(u, v) \in F_\delta$ so that $u \in L^1(\bar{A}, \alpha)$ and $u\delta, v\delta \in L^1(A, H^1)$ and $f, g \geq 0$ when defined as above. Then Proposition 8 asserts $v \in L^1(\bar{A}, z_{\#}\epsilon)$ and $c_\delta\epsilon(b_\theta) + \lambda(b_{\theta'}^L) \leq LP(\delta)^* \leq LP(\delta)_* \leq \alpha(u) + \delta\langle u + v\rangle_A$. If $c_\delta\epsilon(b_\theta) + \lambda(b_{\theta'}^L) = \alpha(u) + \delta\langle u + v\rangle_A$ this forces this chain of inequalities to become equalities, showing (ϵ, λ) and (u, v) to optimize their respective problems.

The converse is proved using the result $LP(\delta)^* = LP(\delta)_*$, which follows by combining the same proposition with Theorem 17. Suppose $\alpha(u) + \delta\langle u + v\rangle_A = LP(\delta)_*$, meaning (u, v) is optimal. Theorem 17 also provides (ϵ, λ) such that $c_\delta\epsilon(b_\theta) + \lambda(b_{\theta'}^L) = LP(\delta)^*$.

Finally, we claim that $c_\delta\epsilon(b_\theta) + \lambda(b_{\theta'}^L) = \alpha(u) + \delta\langle u + v\rangle_A$ is equivalent to $\epsilon(f) = 0 = \lambda(g)$. This follows from the chain of inequalities which establish

26

$c\epsilon(b_\theta) + \lambda(b_{\theta'}^L) \ge \alpha(u)$ in Proposition 8: $\epsilon(f) = 0$ is equivalent to equality in (29), $\lambda(g) = 0$ is equivalent to equality in (31), and when both of these hold then $c_\delta \epsilon(b_\theta) + \lambda(b_{\theta'}^L) = \alpha(u) + \delta\langle u + v\rangle_A$. ∎

Remark 10 (Converse) *According to Theorems 13 and 17, the sufficient condition for optimality of (ϵ, λ) given by Corollary 9 is also necessary.*

2.4 Optimal wages for the primal problem

Using the foundations laid in the previous sections, we are ready to demonstrate the existence of optimal wages $v(k)$ and payoffs $u(a)$ for the primal problem (10). This is done using a compactness and (lower semi-)continuity argument for the perturbed problem (24), and then taking the limit $\delta \to 0$. For $\delta > 0$, we assume v is convex nondecreasing, and then use Lemma 5 and the characterization $v = \max\{v_w, v_w, v_t\}$ — which identifies the wage of an ability k adult with the maximum he can earn as a worker, manager or teacher — to show the convexity and monotonicity assumptions on v do not bind, so play no role in the outcome of our (infinite-dimensional) linear program. Thus convexity of the wages in our model emerges for reasons which manifest rather differently than in Rosen's investigation of superstars [19].

Compactness for convex non-decreasing v is asserted in the following lemma. Some delicacy is required to show that if v or u diverges to $+\infty$, then both do so on the same half-open interval, and at a uniform rate.

Lemma 11 (Compactness for wage functions) *Fix $K = [0, \bar{k}[$ and $g \in L^1_{loc}(K)$. A sequence $v_i : K \longrightarrow [0, \infty[$ of convex non-decreasing functions satisfying $v_i''(k) \ge g(k)$ a.e., admits a subsequence which converges pointwise to a limit $v_0 : K \longrightarrow [0, \infty]$ which is real valued on $[0, k_0[$, and infinite on $]k_0, \bar{k}[$, for some $k_0 \in [0, \bar{k}]$. The convergence is uniform on compact subsets of $[0, k_0[$, and the analogous bound $v_0''(k) \ge g(k)$ holds a.e. on its interior. Furthermore, for $a > k_0$*

$$u_i(a) := \max_{k \in [0, \bar{k}[} cb_E(z(a, k)) + v_i(z(a, k)) - \tfrac{1}{N} v_i(k)$$

diverges to $u_0(a) = +\infty$ as $i \to \infty$ along the subsequence described above, where $b_E \in C^1(\bar{K})$ satisfies (1)-(3), $c \ge 0$ and $z(a, k) = (1 - \theta)a + \theta k$.

Proof. The fundamental theorem of calculus yields

$$v_i(k') = v_i(0) + \int_0^{k'} v_i'(k)dk. \tag{33}$$

Since $0 \le v_i'(k)$ is non-decreasing for each i, Helly's selection theorem provides a subsequence converging to a non-decreasing limit $v_0'(k)$ on K, except possibly at discontinuities of v_0' in $]0, \bar{k}[$. Choose a further subsequence for which $v_0(0) := \lim_{j \to \infty} v_{i(j)}(0)$ converges; unless such a sequence exists, $v_0(0) = +\infty$ and the lemma follows immediately with $k_0 = 0$. Therefore assume $v_0(0) < \infty$ and choose $k_0 \in [0, \bar{k}]$ so that $v_0'(k) < \infty$ for $k < k_0$ and $v_0'(k) = \infty$ for $k > k_0$. For $k' < k_0$, Lebesgue's dominated convergence theorem allows us to take $i(j) \to \infty$ in (33), to obtain a continuous limit $v_0(k')$ on $[0, k_0[$. It follows that $v_{i(j)} \to v_0$ uniformly on compact subsets of $[0, k_0[$. Monotonicity of v_i' ensures $v_{i(j)}(k) \to \infty$ for each $k > k_0$. For $v_i, g \in L^1_{loc}$, the inequality $v_i'' \ge g$ holds in the distributional sense — meaning

$$\int_0^{\bar{k}} [f''(k)v_i(k) - f(k)g(k)]dk \ge 0 \tag{34}$$

for each smooth compactly supported test function $0 \le f \in C_c^\infty(]0, \bar{k}[)$ — if and only if it holds in the a.e. sense. Thus $v_i'' \ge g$ distributionally, and the bound $v_0'' \ge g$ follows on $]0, k_0[$, using Lebesgue's dominated convergence theorem again. Taking $g = 0$ shows v_0 is convex on $]0, k_0[$, for example.

Now if $a > k_0$, taking $k = k_0$ implies $k_0 < z(a, k) = (1 - \theta)a + \theta k$, thus $u_{i(j)}(a) \ge c v_{i(j)}(z(a, k_0))$ diverges to $+\infty$ as $j \to \infty$. ∎

Corollary 12 (Convergence uniform from below) *Suppose a sequence $v_i : [0, \bar{k}[\longrightarrow [0, \infty[$ of functions satisfying the hypotheses of Lemma 11 converges pointwise to $v_0 : [0, \bar{k}[\longrightarrow [0, \infty]$ which is real valued on $[0, k_0[$, and infinite on $]k_0, \bar{k}[$ for some $k_0 \in [0, \bar{k}]$. If $v_0(k_0^-) := \lim_{k \uparrow k_0} v_0(k) < +\infty$ then*

$$0 \le \liminf_{i \to \infty} \inf_{k \in [0, k_0[} v_i(k) - v_0(k). \tag{35}$$

On the other hand, if $v_0(k_0^-) = +\infty$ then the sequence grows uniformly in the sense that for each $c < \infty$ taking $i' < \infty$ large enough implies $v_i(k) \ge c$ for all $k > k_0 - 1/i'$ and $i > i'$.

28

Proof. Given $\delta > 0$, taking $k_1 < k_0$ sufficiently large makes $v_0(k_1) > v_0(k_0^-) - \delta/2$. Taking i sufficiently large then ensures $v_i(k_1) > v_0(k_0^-) - \delta$, whence for all $k \in [k_1, k_0[$ monotonicity yields $v_i(k) > v_0(k) - \delta$. Since the convergence $v_i \to v_0$ is uniform on $[0, k_1]$, this concludes the corollary in case $v_0(k_0^-) < +\infty$ is finite.

If $v_0(k_0^-) = +\infty$, given $c < \infty$ take i' sufficiently large that $v_0(k_0 - 1/i') > c$ and then larger still to ensure $v_i(k_0 - 1/i') > c$ for all $i > i'$. Monotonicity again concludes the proof. ■

Theorem 13 (Existence of minimizing wages) *Fix $c \geq 0$ and positive θ, θ', N, N' and $\bar{a} = \bar{k}$ with $\max\{\theta, \theta'\} < 1 \leq N$. Set $A = [0, \bar{a}[= K$ and let α be a Borel probability measure on $\bar{A}$ satisfying the doubling condition (12) at $\bar{a} \in \mathrm{spt}\,\alpha$. Define $z(a, k) = (1 - \theta)a + \theta k$, $b_\theta = b_E \circ z$ and $b_{\theta'}^L(a, k) = b_L((1 - \theta')a + \theta'k)$, where $b_{E/L} \in C^1(\bar{K})$ satisfy (1)–(3). Then infimum (10) is attained by functions (u, v) satisfying $v = \max\{v_w, v_m, v_t\}$ on $\bar{K} = [0, \bar{k}]$ and*

$$u(a) = \sup_{k \in \bar{K}} cb_E(z(a, k)) + v(z(a, k)) - \tfrac{1}{N}v(k) \tag{36}$$

on $\bar{A}$, where the $v_{w/m/t}$ are defined by (13)–(16); 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 $d^j v/dk^j \geq \underline{b}_L^{(j)} \min\{(1 - \theta')^j, (\theta')^j N'\}$.

Proof. Fix $0 < \delta < 1$ and $c_\delta := c > 0$ positive; if we prefer $c = 0$ set $c_\delta = \delta$ in the $\delta \to 0$ limit procedure which follows. We are going to study the perturbed primal problem (24) under the same feasibility constraints (6)–(8) as (10) — which include $u \in L^1(\bar{A}, \alpha)$ — plus the artificial constraint that v be convex nondecreasing. From (8), both u and $v \in L^1(\bar{A}, \alpha)$ have positive lower bounds. For $\delta > 0$ we assume $u, v \in L^1(A, H^1)$ without loss of generality, since otherwise the term $\langle u + v \rangle_A = +\infty$ makes the objective diverge. Feasibility of the pair $(u, v) = (1 + c_\delta \bar{b}_E/\bar{b}_L, 1)\bar{b}_L$ yields an upper bound $(1 + 2\delta)(c_\delta \bar{b}_E + \bar{b}_L)$ for the infimum (24). As remarked after (8), we may always replace u and v by their lower semi-continuous hulls without violating feasibility. Since $\alpha \geq 0$, this only improves the objective (24); for the same reason, it costs no generality to henceforth suppose u to be related to v by (36). Lemma 5 then implies both v and u are convex and non-decreasing, hence continuous as extended real-valued functions.

Lemma 11 allows us to extract a subsequential limit (u_δ, v_δ) satisfying the same constraints from any sequence of approximate minimizers for (24).

Fatou's lemma ensures the limit (u_δ, v_δ) minimizes the objective subject to these constraints. Replacing the monotone convex functions u_δ and v_δ again by their lower-semicontinuous hulls ensures both are continuous. Since $\bar{a} \in$ spt α, our a priori bound $(1 + 2\delta)(c_\delta \bar{b}_E + \bar{b}_L)$ on the objective implies the non-decreasing functions $u_\delta(a)$ and $v_\delta(k)$ are real-valued, except possibly at $\bar{a}$ and $\bar{k}$, and

$$\int_{\bar{A}} u_\delta(a)\alpha(da) \le (1 + 2\delta)(c_\delta \bar{b}_E + \bar{b}_L). \tag{37}$$

Notice equality must hold in

$$u_\delta(a) \ge \sup_{k \in [0,\bar{k}]} c_\delta b_E(z(a,k)) + v_\delta(z(a,k)) - \tfrac{1}{N} v_\delta(k) \tag{38}$$

since otherwise replacing u_δ by the right-hand side of (38) yields a feasible pair which lowers the objective functional, contradicting the asserted optimality. Use $(u,v) = (u_\delta, v_\delta)$ to define $(v_\delta^w, v_\delta^m, v_\delta^t) := (v_w, v_m, v_t)$ and $\bar{v}_\delta := \max\{v_w, v_m, v_t\}$.

Feasibility implies $v_\delta \ge \bar{v}_\delta$, and Lemma 5 implies $\bar{v}_\delta$ is continuous on K, convex increasing on $\bar{K}$, and satisfies

$$\bar{v}_\delta' \ge \min\{(1 - \theta')\underline{b}_L', N'\theta'\underline{b}_L', (c_\delta b_E' + \inf v_\delta'(k))N\theta\} \quad \text{and} \tag{39}$$

$$\bar{v}_\delta'' \ge \min\{(1 - \theta')^2 \underline{b}_L'', (\theta')^2 N' \underline{b}_L'', (c_\delta \underline{b}_E'' + \inf v_\delta'')N\theta^2\} \tag{40}$$

on $]0, \bar{k}[$. If $\eta := v_\delta - \bar{v}_\delta$ is positive somewhere, it is positive on an interval where the only binding constraints can be $v_\delta' = 0$ or $v_\delta'' = 0$. For small $\lambda > 0$, the perturbation $v^\lambda := (1 - \lambda)v_\delta + \lambda \bar{v}_\delta$ respects these differential constraints. We will now show the pair (u_δ, v^λ) respects the other constraints as well; unless the continuous function $\eta = 0$ throughout K, this pair lowers the objective functional, a contradiction forcing $v_\delta = \bar{v}_\delta$.

Since $v^\lambda = v_\delta - \lambda\eta = \bar{v}_\delta + (1 - \lambda)\eta$, for $k', k \in \bar{K}$ we find

$$v^\lambda(k') + \tfrac{v^\lambda(k)}{N'} - b_{\theta'}^L(k',k) = \bar{v}_\delta(k') + \tfrac{v_\delta(k)}{N'} - b_{\theta'}^L(k',k) + (1 - \lambda)\eta(k') - \tfrac{\lambda}{N}\eta(k)$$

$$\ge \eta(k')[1 - \lambda(1 + \tfrac{1}{N}\tfrac{\eta(k)}{\eta(k')})], \tag{41}$$

and also

$$v^\lambda(k') + \tfrac{v^\lambda(k)}{N'} - b_{\theta'}^L(k',k) = v_\delta(k') + \tfrac{\bar{v}_\delta(k)}{N'} - b_{\theta'}^L(k',k) - \lambda\eta(k') + \tfrac{1-\lambda}{N}\eta(k)$$

$$\ge \tfrac{\eta(k)}{N}[1 - \lambda(1 + \tfrac{N\eta(k')}{\eta(k)})]. \tag{42}$$

30

If both $\eta(k') \geq 0$ and $\eta(k) \geq 0$ are non-zero, then taking $\lambda < 1/2$ ensures either (41) or (42) is positive. The same conclusion remains true if one of $\eta(k')$ or $\eta(k)$ vanishes. If both vanish, there is nothing to prove.

On the other hand, adding $u_\delta(a) - c_\delta b_E(z(a,k))$ to

$$\tfrac{v^\lambda(k)}{N} - v^\lambda(z(a,k)) = \tfrac{\bar{v}_\delta(k)}{N} - v_\delta(z(a,k)) + \tfrac{1-\lambda}{N}\eta(k) + \lambda\eta(z(a,k))$$

shows

$$u_\delta(a) + \tfrac{v^\lambda(k)}{N} - c_\delta b_E(z(a,k)) - v^\lambda(z(a,k)) \geq \tfrac{1-\lambda}{N}\eta(k) + \lambda\eta(z(a,k)) \geq 0$$

as desired, since $\bar{v}_\delta \geq v_\delta^t$. This establishes $v_\delta = \bar{v}_\delta$ on K. At $\bar{k}$, convexity implies upper semicontinuity of $\bar{v}_\delta$ and it is dominated by the continuous function v_δ, so identity $v_\delta = \bar{v}_\delta$ extends to $\bar{K}$.

As a consequence of (39)–(40), for $c_\delta > 0$ both v_δ' and v_δ'' are bounded away from zero so the constraints $\min\{v', v''\} \geq 0$ are not binding. We claim (u_δ, v_δ) must also minimize the linear program (24) even among feasible pairs which do not satisfy these additional constraints. To see this, we'll suppose the objective was lower at some other feasible pair $(u, v) \in F_0$ and derive a contradiction. If $u, v \in C^2(\bar{A})$, then the pair $(1 - s)(u_\delta, v_\delta) + s(u, v) \in F_0$ also lowers the objective for $s > 0$ sufficiently small, and inherits the strict convexity and monotonicity of (u_δ, v_δ) to produce the desired contradiction. If $u, v \notin C^2(\bar{A})$, the same contradiction will be obtained after approximating (u, v) by a smooth feasible pair. We can at least assume u and v are continuous and bounded according to the proof of Theorem 17. The Stone-Weierstrauss theorem then shows u and v can be approximated uniformly by smooth functions $(\tilde{u}_\sigma, \tilde{v}_\sigma)$ such that $u + \sigma \leq \tilde{u}_\sigma \leq u + 2\sigma$ and $v \leq \tilde{v}_\sigma \leq v + \sigma$ as $\sigma \to 0^+$. In this case, $(\tilde{u}_\sigma, \tilde{v}_\sigma) \in F_0$ follows from $(u, v) \in F_0$. Convergence of the objective function to its limiting value as $\sigma \to 0$ is readily verified. This establishes the desired contradiction, hence the minimality of (u_δ, v_δ) in F_0.

Now Corollary 9 asserts there are non-negative measures $\epsilon_\delta \geq 0$ and $\lambda_\delta \geq 0$ satisfying the perturbed feasibility constraints (25) such that

$$\alpha(u_\delta) + \delta\langle u_\delta + v_\delta\rangle_A = c_\delta\epsilon_\delta(b_\theta) + \lambda_\delta(b_{\theta'}^L). \tag{43}$$

Lemma 11 yields a subsequential limit $(u_{\delta_i}, v_{\delta_i}) \to (u_0, v_0)$ pointwise on $\bar{A} \times \bar{K}$ and uniformly on compact subsets of $[0, a_0[\times [0, k_0[$, with $u_0(a) = +\infty$ for $a > a_0 \in [0, \bar{a}]$ and $v_0(k) = +\infty$ for $k > k_0 \in [0, \bar{k}]$ and $a_0 \leq k_0$. We claim $a_0 = \bar{a}$.

Recalling the monotonicity of u_δ, if $a_0 < \bar{a}$ we have $u_{\delta_i}(k) \to +\infty$ uniformly on $a \in [(a_0 + \bar{a})/2, \bar{a}]$. Since $\bar{a} \in \operatorname{spt} \alpha$, Fatou's lemma will contradict the bound (37) unless $a_0 = \bar{a}$. This also forces equality in $\bar{k} = \bar{a} \leq k_0 \leq \bar{k}$. Thus (u_0, v_0) are feasible for the original problem (10).

Extracting a further subsequence if necessary, we may also assume $(\epsilon_{\delta_i}, \lambda_{\delta_i}) \to (\epsilon_0, \lambda_0)$ weak-$*$ in $C(\bar{A} \times \bar{K})^*$ as $\delta_i \to 0$ to feasible measures for the dual problem (11). Taking the $\delta \to 0$ limit of (43), Fatou's lemma combines with the weak-$*$ convergence to give

$$\alpha(u_0) \leq c_0 \epsilon_0(b_\theta) + \lambda_0(b_{\theta'}^L) \in \mathbf{R}.$$

Proposition 8 yields the opposite inequality, and its corollary then confirms the desired optimality of (u_0, v_0) (and of (ϵ_0, λ_0)).

Noting $v_\delta = \bar{v}_\delta$, the inequalities (39)–(40) survive passage to the $\delta_i \to 0$ limit in both the distributional (34) and a.e. senses. For $j = 1$ or $j = 2$, when $N\theta^j \geq 1$, these inequalities imply $d^j v_\delta / dk^j \geq \underline{b}_L^{(j)} \min\{(1 - \theta')^j, (\theta')^j N'\}$ throughout K before and hence after the limit. It remains to show the identity $v_\delta = \bar{v}_\delta$ survives the $\delta_i \to 0$ limit first on K, and eventually on $\bar{K}$.

Although we have only subsequential convergence of (u_δ, v_δ), we abuse notation by writing $\delta \to 0$ to denote this subsequence hereafter. Taking $\delta \to 0$ in the remaining identity of interest $v_\delta = \bar{v}_\delta$ yields

$$v_0 := \lim_{\delta \to 0} v_\delta = \max\{\limsup_{\delta \to 0} v_\delta^w, \limsup_{\delta \to 0} v_\delta^m, \limsup_{\delta \to 0} v_\delta^t\}. \tag{44}$$

Using $\bar{k}^-$ to denote the limit $k \uparrow \bar{k}$, we claim $u_0(\bar{k}^-) < \infty$ if $v_0(\bar{k}^-) < \infty$, and $u_0(\bar{k}^-) = \infty$ if $v_0(\bar{k}^-) = \infty$. The second claim follows from (8), which gives $u_0(a) \geq \frac{N-1}{N} v_0(a)$; the first claim is more subtle unless α has a Dirac mass at $\bar{a}$, but follows from the boundedness of v_0 in the supremum (36) due to the following parenthetical paragraph.

(To see that (36) continues to hold when $\delta = 0$ assuming $\alpha[\{\bar{a}\}] = 0$, consider the continuous function $f_\delta(a, k) := u_\delta(a) + \frac{1}{N} v_\delta(k) - c_\delta b_E(z(a, k)) - v_\delta(z(a, k)) \geq 0$ on $A \times \bar{K}$. The zero set Z_δ of f_δ is relatively closed in $A \cap \bar{K}$; it is non-decreasing by the strict submodularity shown in Lemma 5, and contains $(A \times \bar{K}) \cap \operatorname{spt} \epsilon_\delta$ according to Corollary 9. For each $(a_\delta, k_\delta) \in Z_\delta$ this monotonicity implies

$$\int_{]a_\delta, \bar{a}] \times \bar{K}} \epsilon_\delta(da, dk) \leq \int_{\bar{A} \times [k_\delta, \bar{k}]} \epsilon_\delta(da, dk). \tag{45}$$

32

Fixing $a_\delta = a$, to establish the limiting case of (36) it is enough to show $\limsup_{\delta \to 0} k_\delta < \bar{k}$. Recalling that the left and right marginals of ϵ_δ are given by (25), setting $\Delta a = \bar{a} - a$ and $\Delta k_\delta = \bar{k} - k_\delta$, from (45) we deduce

$$
\begin{aligned}
\frac{1}{N}(\bar{a}\alpha(]\bar{a} - \Delta a, \bar{a}]) + \delta \Delta a) &\leq \bar{a}(z_\# \epsilon_\delta)([\bar{k} - \Delta k_\delta, \bar{k}]) + \delta \Delta k_\delta \\
&\leq \delta \Delta k_\delta + (\bar{a}\alpha + \delta H^1|_A)([\bar{a} - \frac{1}{1-\theta}\Delta k_\delta, \bar{a}])
\end{aligned}
$$

where the second inequality follows from (47). Since $\bar{a} \in \mathrm{spt}\,\alpha$ but $\alpha[\{\bar{a}\}] = 0$, the left hand side remains bounded away from zero in the limit $\delta \to 0$, whence we conclude $\liminf_{\delta \to 0} \Delta k_\delta > 0$ also. Thus (36) holds for $a \in A$ with $\delta = 0$.)

Now if $v_0(\bar{k}^-) < \infty$ then Corollary 12 allows us to deduce $\limsup_{\delta \to 0} v_\delta^t \leq v_0^t$ for $k \in [0, \bar{k}[$ from

$$
v_\delta^t(k) = N \sup_{a \in [0, \bar{a}[} c_\delta b_E(z(a, k)) + v_\delta(z(a, k)) - u_\delta(a), \tag{46}
$$

noting $u_0(a) \leq \liminf_{\delta \to 0} u_\delta(a)$ uniformly on $[0, \bar{a}[$ and $z(a, k)$ is constrained to the range where the convergence $v_\delta \to v_0$ is uniform. Showing $\limsup_{\delta \to 0} v_\delta^w \leq v_0^w$ and $\limsup_{\delta \to 0} v_\delta^m \leq v_0^m$ is similar but simpler, whence $v_0 \leq \max\{v_0^w, v_0^m, v_0^t\}$. The opposite inequality follows from the constraints satisfied by (u_0, v_0).

If $v_0(\bar{k}^-) = +\infty$ on the other hand, then for fixed $k \in [0, \bar{k}[$ let C_δ denote the supremum of $c_\delta b_E(z(a, k)) + v_\delta(z(a, k))$ over $a \in [0, \bar{a}[$ and observe $C_\delta \to C_0 < \infty$ as $\delta \to 0$. Take $\delta_0 > 0$ sufficiently small that $C_{\delta_0} < 2C_0$, and smaller if necessary using Corollary 12 so that $u_\delta(\bar{a} - \delta_0) > 2C_0$ for all $\delta < \delta_0$. For $\delta < \delta_0$, the supremum (46) is unchanged if we restrict its domain $a \in [0, \bar{a} - \delta_0]$ to an interval where convergence $(u_\delta, v_\delta) \to (u_0, v_0)$ is uniform. Thus taking $\delta \to 0$ in (46) yields $\lim_{\delta \to 0} v_\delta^t(k) = v_0^t(k)$. A similar but simpler argument yields $v_0^w(k) = \lim_{\delta \to 0} v_\delta^w(k)$ and $v_0^m(k) = \lim_{\delta \to 0} v_\delta^m(k)$, whence the desired identity follows from (44).

It costs no generality to replace u_0 by the right hand side of (38) with $\delta = 0$, which is feasible and no larger than u_0 in any case. (In fact, they coincide throughout A by the parenthetical paragraph above.) Let us now argue that we may take v_0 to be continuous, or equivalently take equality to hold in $v_0(\bar{k}^-) \leq v_0(\bar{k})$. If $v_0(\bar{k}^-) < v_0(\bar{k})$, replacing $v_0(\bar{k})$ with $v_0(\bar{k}^-)$ does not violate any of the feasibility constraints. Nor does it affect the values of v_w, v_m, v_t or u_0 — except to remedy any discontinuity in v_t or u_0 by reducing

$v_t(\bar{k})$ and $u_0(\bar{a})$. This can only improve the objective, and by continuity of all of the resulting functions extends the identity $v_0 = \bar{v}_0$ from K — where it was already established — to $\bar{K}$, to complete the proof. ∎

2.5   Uniqueness and properties of optimal matchings

Finally, we are ready to tackle the structure of optimal matchings in the education and labor sectors, and to give conditions guaranteeing uniqueness of optimizers for both the primal and dual problems (10)–(11).

The structure our education sector often leads to positive assortative matching ϵ of students with teachers. (Our labor sector always leads to positive assortative matching of workers to managers.) However, since distribution κ of cognitive skills acquired by adults in our population is endogenous, it might not be unique. The following theorem specifies conditions for uniqueness. These require, in particular, that κ as well as the exogenous distribution of student skills α be atom free. The following lemma details how κ inherits this and other useful properties from the distribution α of student skills input. Even without positive assortativity, unless the (exogenous) probability measure α concentrates positive mass at the top skill type $\alpha[\{\bar{a}\}] > 0$, it follows that κ concentrates no mass at the upper endpoint of $K = [0, \bar{k}[$. Then any matching $\epsilon \geq 0$ on $\bar{A} \times \bar{K}$ which satisfies the steady-state constraint $\frac{1}{N}\epsilon^1 \leq z_{\#}\epsilon$ must concentrate all of its mass on $A \times K$.

Lemma 14 (Endogenous distribution of adult skills) *Fix $\theta \in]0, 1[$ and a Borel measure $\alpha \geq 0$ on $\bar{A}$ with $\alpha[\bar{A}] < \infty$ for $A = [0, \bar{a}[$ with $\bar{a} > 0$. Set $K = [0, \bar{k}[= A$ and $z(a, k) = (1 - \theta)a + \theta k$. If $\epsilon \geq 0$ on $\bar{A} \times \bar{K}$ has $\alpha = \epsilon^1$ as its left marginal, then for each $\bar{k} - \Delta k \in K$ the corresponding distribution $\kappa = z_{\#}\epsilon$ of adult skills satisfies*

$$\int_{[\bar{k}-\Delta k,\bar{k}]} \kappa(dk) \leq \int_{[\bar{a}-\frac{1}{1-\theta}\Delta k,\bar{a}]} \alpha(da). \tag{47}$$

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

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 $k_t : \bar{A} \longrightarrow \bar{K}$. uniquely determined α-a.e. by κ. Moreover, if $\alpha(da) = \alpha^{ac}(a)da$ is given by a density $\alpha^{ac} \in L^1(A)$, then $\kappa(dk) = \kappa^{ac}(k)dk$ is given by a related density $\kappa^{ac} \in L^1(K)$ satisfying

$$\alpha^{ac}(a) = (1 + \theta(k_t'(a) - 1)) \, \kappa^{ac}(z(a, k_t(a))) \tag{48}$$

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

Proof. The definition $\kappa = z_\# \epsilon$ yields $\kappa([\bar{k} - \Delta k, \bar{k}]) = \epsilon[z^{-1}([\bar{k} - \Delta k, \bar{k}])]$. Now $\bar{k} - \Delta k \leq z(a, k) \leq (1 - \theta)a + \theta\bar{k}$ implies $a \geq \bar{k} - \frac{1}{1-\theta}\Delta k$. Thus

$$\kappa([\bar{k} - \Delta k, \bar{k}]) \leq \epsilon\left([\bar{a} - \tfrac{1}{1-\theta}\Delta k, \bar{a}] \times \bar{K}\right) = \alpha([\bar{a} - \tfrac{1}{1-\theta}\Delta k, \bar{a}])$$

which is the desired bound (47).

For the measure ϵ to be positive assortative means its support spt ϵ is non-decreasing. Except possibly for a countable number of jump discontinuities, this support is then contained in the graph of some non-decreasing map $k_t : \bar{A} \to \bar{K}$. If α is free of atoms, the countable set of a where the jumps occur is a set of measure zero. Then the formula $\epsilon = (id \times k_t)_\# \alpha$ and uniqueness of k_t are well-known facts, established e.g. in Lemma 3.1 of [1] and the main theorem of [14]. It follows that $f(a) = z(a, k_t(a))$ is non-decreasing, and pushes α forward to κ. By Lebesgue's theorem, $f'(a) = 1 - \theta + \theta k_t'(a)$ exists H^1-a.e. and enjoys the positive lower bound $f'(a) \geq 1 - \theta$. Thus f is one-to-one and there is an inverse function $g : \bar{K} \longrightarrow \bar{A}$ with Lipschitz constant at most $\frac{1}{1-\theta}$ such that $g(\bar{f}(k)) = k$ for any non-decreasing extension $\bar{f} : \bar{K} \longrightarrow \bar{A}$ of f (to points where $k_t(a)$ may not be differentiable). For $K' \subset \bar{K}$ we have $\kappa[K'] = \alpha[f^{-1}(K')] = \alpha[g(K')]$. Taking K' to consist of any single point shows κ has no atoms if α has no atoms. Taking K' to be an arbitrary set of Lebesgue measure zero shows κ absolutely continuous with respect to Lebesgue if α is absolutely continuous with respect to Lebesgue, noting $H^1[g(K')] \leq \frac{1}{1-\theta}H^1[K']$. The formula $\alpha^{ac}(a) = f'(a)\kappa^{ac}(f(a))$ then follows essentially from the fundamental theorem of calculus, and is argued rigorously in [15]. The bound $\|\kappa^{ac}\|_{L^\infty(K)} \leq \frac{1}{1-\theta}\|\alpha^{ac}\|_{L^\infty(A)}$ is a consequence, so the proof is complete. ∎

Theorem 15 (Positive assortative and unique optimizers) *Fix $c \geq 0$ and positive θ, θ', N, N' and $\bar{a}$ with $\max\{\theta, \theta'\} < 1 \leq N$. Set $A = [0, \bar{a}[$ and let α be a Borel probability measure on $\bar{A}$ satisfying the doubling condition (12) at $\bar{a} \in$ spt α. Define $z(a, k) = (1 - \theta)a + \theta k$, $b_\theta = b_E \circ z$ and $b_{\theta'}^L(a, k) = b_L((1 - \theta')a + \theta'k)$ where $b_{E/L} \in C^1(\bar{K})$ satisfy (1)–(3). If $\epsilon, \lambda \geq 0$ on $\bar{A}^2$ maximize the dual problem (11), 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 dual problem (10) which are non-decreasing and strictly convex, (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 7 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.

Proof. Set $K = [0, \bar{k}[= A$. Existence of a maximizing pair (ϵ, λ) is asserted by Theorem 17. Let us begin by showing that they are positive assortative under the extra condition that minimizing payoffs (u, v) exist for (10) which are strictly convex. Lemma 5 asserts $v(z(a, k))$ is then strictly supermodular.

Set $f(a, k) = u(a) + \frac{v(k)}{N} - cb_E(z(a, k)) - v(z(a, k)) \geq 0$ and $g(k', k) = v(k') + \frac{v(k)}{N} - b_{\theta'}^L(k', k)) \geq 0$ on $\bar{A} \times \bar{K}$, with the convention $f(\bar{a}, \bar{k}) \leq 0$ if $v(\bar{k}) = +\infty$, and vanishing if and only if $u(\bar{a}) = +\infty$ in addition. Corollary 9 asserts $\epsilon(f) = 0$ and $\lambda(g) = 0$ for any dual maximizers (ϵ, λ). Thus ϵ and λ must vanish outside the respective zero sets $F \subset \bar{A} \times \bar{K}$ of f and $G \subset \bar{K}^2$ of g.

When f and g are strictly submodular, then F and G are non-decreasing in the plane, meaning λ and ϵ are positive assortative. This strict submodularity follows from that of $-b_E(z(a, k))$ and $-v(z(a, k))$.

Finally, assume in addition that α is atom free. If (ϵ_i, λ_i) are dual maximizers, for $i = 0, 1$, then so is their average $(\epsilon_2, \lambda_2) := (\epsilon_0 + \epsilon_1, \lambda_0 + \lambda_1)/2$. Thus ϵ_2 vanishes outside the non-decreasing set F, as do $\epsilon_{0/1}$. Similarly λ_i all vanish outside the same non-decreasing set G for $i = 0, 1, 2$. This strongly suggests the asserted uniqueness, an intuition we now make precise. Except perhaps for a countable number of vertical segments, the non-decreasing set F is contained in the graph of a non-decreasing map $k_t : \bar{A} \longrightarrow \bar{K}$. Any joint measure ϵ with $\epsilon^1 = \alpha$ cannot charge these vertical segments, since this would imply α has atoms. Since our maximizers ϵ_i vanishes outside the graph of k_t, we conclude they must coincide with the measure $(id \times k_t)_{\#}\alpha$ by Lemma 3.1 of [1]. This identification shows $\epsilon_0 = \epsilon_1$. The associated distributions $\kappa = z_{\#}\epsilon_0$ and $\kappa_t = (\epsilon_0)^2/N$ of adult and teacher skills are therefore also unique. Moreover, κ is free from atoms, according to Lemma 14.

Let λ_i^1 and λ_i^2 be the left and right marginals of each maximizer $\lambda_i \geq 0$ for the labor sector, whose feasibility implies $\lambda_i^1 + \lambda_i^2/N' = \kappa - \kappa_t$ is also atom-free. Let $\Delta\lambda = \lambda_0 - \lambda_1$ denote the difference of the two maximizers.

Recall that both λ_i — and hence $\Delta\lambda$ — must vanish outside the same non-decreasing set G. Just as before, the non-decreasing set G has at most countably many horizontal and vertical segments, which λ_i cannot charge since its marginals are free from atoms. Now the positive marginals $\Delta\lambda_+^1 := ((\Delta\lambda)_+)^1 = ((\Delta\lambda)^1)_+$ and $\Delta\lambda_+^2$ of the difference must have the same mass, since the atom-free condition precludes cancellations. On the other hand, feasibility implies $N\Delta\lambda_+^1 - N\Delta\lambda_-^1 + \Delta\lambda_+^2 - \Delta\lambda_-^2 = 0$, which forces $N\Delta\lambda_+^1 = \Delta\lambda_-^2$ (and $N\Delta\lambda_-^1 = \Delta\lambda_+^2$). Since these two measures have the same mass, $N \neq 1$ produces a contradiction unless $\Delta\lambda = 0$. If $N = 1$ so that all adults are teachers, then $\lambda_i = 0$. This establishes the uniqueness asserted for the dual problem.

Having established the existence of positive assortative maximizers when v is strictly convex, we now turn to the case that strict convexity fails. According to Theorem 13, this happens only when $c = 0$ and $N\theta^2 < 1$, so we can approximate this situation as a $c \to 0$ limit. Let (ϵ_c, λ_c) and (u_c, v_c) be the (non-negative) optimizers described above for the problem with $c > 0$, so that $c\epsilon_c(b_\theta) + \lambda_c(b_{\theta'}^L) = \alpha(u_c)$ according to Remark 10. Using the Banach-Alaoglu theorem and the compactness results of Lemma 11, we extracting a subsequential limit $(\epsilon_c, \lambda_c) \to (\epsilon, \lambda)$ in the weak-$*$ topology on $C(\bar{A} \times \bar{K})^*$ and $(u_c, v_c) \to (u, v)$ locally uniformly on $[0, a_0[$, with $u(a) = +\infty = v(a)$ for all $a > a_0$. The limiting pairs are feasible for the primal and dual problems respectively, and positive assortativity survives the limiting process [14]. Fatou's lemma allows us to take the subsequential limit of $c\epsilon_c(b_\theta) + \lambda_c(b_{\theta'}^L) = \alpha(u_c)$ to arrive at $\lambda(b_{\theta'}^L) \geq \alpha(u)$. The reverse inequality is asserted by Proposition 8, and confirms optimality of (ϵ, λ) by Corollary 9.

We now address uniqueness of the primal minimizers. Since u and v are strictly convex, both are continuous functions with one-sided derivatives throughout K, and two-sided derivatives except perhaps at countably many points. Define $u(\bar{a}) = \lim_{a \to \bar{a}} u(a)$ and $v(\bar{k})$ similarly. Since the measures α and $z_\#\epsilon$ have no atoms, the asserted derivatives of u and v exist. Denote the distribution of workers and managers by $\kappa_w := \pi_\#^1\lambda$ and $\kappa_m := \pi_\#^2\lambda/N'$. The projections of $\mathrm{spt}\,\epsilon$ through $\pi^1(a, k) = a$ and $\pi^2(a, k) = k$ are compact sets of full measure for κ_w and κ_m respectively. Take $\mathrm{Dom}\,v' \subset]0, \bar{k}[$ by convention. For each $k' \in \pi^1(\mathrm{spt}\,\lambda) \cap \mathrm{Dom}\,v'$, there is a unique $k \in \bar{K}$ with $(k', k) \in \mathrm{spt}\,\lambda \subset G$. The first-order condition $g_{k'}(k', k) = 0$ then gives $v'(k') = (1 - \theta')b_L'((1 - \theta)k' + \theta k)$; by strict convexity of b_L there cannot be two such k without differentiability of v failing at k'. This shows v' to be uniquely determined by λ throughout $\pi^1(\mathrm{spt}\,\lambda) \cap \mathrm{Dom}\,v'$ — a set of full

κ_w measure. A similar argument with the roles of k' and k interchanged shows $v'(k) = N'\theta'b'_L((1 - \theta')k' + \theta'k)$ is uniquely determined by λ on the set $\pi^2(\mathrm{spt}\,\lambda) \cap \mathrm{Dom}\,v'$ containing κ_m-a.e. manager type k.

To address $v'(k)$ for the teacher types k, assume hypotheses (d)-(f) of Proposition 7. For $k_1 \in \mathrm{spt}\,\kappa_t \cap \mathrm{Dom}\,v'$, that proposition provides a recursive formula (22) asserting $k_2 \in \mathrm{Dom}\,v'$, and relating $v'(k_1)$ to $v'(k_2)$, where $(a_1, k_1) \in \mathrm{spt}\,\epsilon$ and $k_2 = z(a_1, k_1)$ is the skill of those adults who were trained by type k_1 teachers. The strict monotonicity of $v'(k)$ we have assumed implies a_1 and k_2 are unique. The proposition also asserts that after a finite number d of iterations, this recursion terminates with an adult of skill k_d who is willing to become a worker or a manager, and whose wage gradient $v'(k_d)$ is therefore determined by the considerations above. Thus $v'(k_1)$ is uniquely determined by ϵ, λ, and (23). This establishes the κ-a.e. uniqueness of the wage gradient v'.

Finally, we turn to the net lifetime surplus $u(a)$ of student type $a \in]0, \bar{a}[$. For $a \in \pi^1(\mathrm{spt}\,\epsilon) \cap \mathrm{Dom}\,u'$, there exists $k \in \bar{K}$ (which we'll show to be unique) such that $(a, k) \in \mathrm{spt}\,\epsilon \subset F$. The first-order conditions for one-sided derivatives $\pm f_a(a^\pm, k) \geq 0$ give

$$v'(z(a, k)^-) + cb'_E(z(a, k)^-) \geq \frac{u'(a)}{1 - \theta} \geq v'(z(a, k)^+) + cb'_E(z(a, k)^+).$$

However, the convexity of v on $]0, \bar{k}[$ assert $v'(z^-) \leq v'(z^+)$ and similarly for b_E, so both v and b_E must be differentiable at $z(a, k)$ and equalities hold throughout. Thus

$$v'(z) + cb'_E(z) = \frac{1}{1 - \theta}u'(a).$$

Since the left hand side is strictly increasing in z, we find $z(a, k)$ and hence k is unique. Since v' was uniquely determined for $z_{\#}\epsilon$ adult type, it follows that u' is uniquely determined for α-a.e. student type. If α dominates some absolutely continuous measure whose support fills $\bar{A}$, this shows u is unique up to an additive constant. Given another feasible minimizer (u_0, v_0) with v_0 locally Lipschitz, we see u_0 must produce equality α-a.e. in the inequality (38); otherwise replacing u_0 by the right-hand side would remain feasible and lower the objective (10). On the other hand, the right hand side is locally Lipschitz, according to Lemma 2. The arguments above then yield $u_0 = u + const$. But the constant must vanish since both minimizers yield the same value for the objective functional, showing u_0 is unique in $L^1(\bar{A}, \alpha)$.

∎

2.6 Phase transition to unbounded wage gradients

Having come this far, one may wonder whether establishing the existence of competitive equilibria need be so involved. If we had been content to find optimizing wages u and v which are merely non-decreasing, an argument based on Helly's selection theorem might have sufficed. However, we would not then know the convexity of the wages (used to prove their uniqueness), nor positive assortativity of the education sector.

In this section, we explore the actual behavior of $v(k)$ near the top skill type $\bar{k}$, assuming the distribution of student types is given by a continuous density $\alpha(da) = \alpha^{ac}(a)da$ on $A = K = [0, \bar{k}[$. Under mild differentiability hypotheses, our next theorem establishes the existence of a phase transition separating bounded from unbounded wage gradients. For $N\theta > 1$, it shows the education sector may form into a pyramid scheme in which the marginal wage $v'(k)$ diverges to infinity as $k \to \bar{k}$, even though the absolute wage $v(k)$ remains bounded. For $N\theta \neq 1$, it gives precise asymptotics (49) for the wage function $v(k)$ and the endogenous distribution $\kappa^{ac}(k)$ of adult skills near $\bar{k}$. Notice this formula makes an explicit quantitative prediction for the dependence of the rate of divergence on the teaching capacity N and effectiveness θ assumed in the model. In all cases this divergence is integrable, so the wages tend to a finite limit. For $N\theta < 1$ it predicts a specific limiting slope $v'(k) \to c/(\frac{1}{N\theta} - 1)$ as $k \to \bar{k}$, while for $N\theta > 1$ it predicts $v'(k) \to \infty$ at a specific rate. Thus the differences in marginal wages amongst the very top echelons of teachers ('gurus') is negligible in a thin (or equivalently, vertical) pyramid $N\theta < 1$, but becomes more and more exaggerated if $N\theta > 1$, and at a rate which increases with $N\theta$, corresponding to a fatter and fatter (or equivalently, more and more horizontal) organizational structure with wider effective span of control. When the theorem applies, it also predicts that the density of adults (= teachers) at the highest skill level $\bar{k} = \bar{a}$ tends to a constant multiple $\frac{1-\theta/N}{1-\theta}$ of the density of students.

Theorem 16 (Wage behavior and density of top-skilled adults) *Fix $c \geq 0$ and positive θ, θ', N, N' and $\bar{a} = \bar{k}$ with $\max\{\theta, \theta'\} < 1 \leq N$. 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 $A = [0, \bar{a}[$. Set $z(a, k) = (1 - \theta)a + \theta k$, $b_\theta = b_E \circ z$ and $b^L_{\theta'}(a, k) = b_L((1 - \theta')a + \theta'k)$, where $b_{E/L} \in C^1(\bar{K})$ satisfy (1)–(3). Suppose (ϵ, λ) and convex $(u, v) \in F_0$ optimize the primal and dual problems (10)–(11), and (i) $\bar{k} \in (\mathrm{spt}\, \epsilon^2) \setminus \mathrm{spt}(\lambda^1 + \lambda^2)$, 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 $]\bar{k} - \delta, \bar{k}[$. Then for $N\theta \neq 1$,

$$v'(k) = \frac{const}{|\bar{k} - k|^{\frac{\log N\theta}{\log N}}} - \frac{c\bar{b}'_E}{1 - \frac{1}{N\theta}} + o(1) \tag{49}$$

as $k \to \bar{k}$, and the steady state distribution $\kappa = z_{\#}\epsilon$ of adult skills satisfies

$$\kappa^{ac}(\bar{k}) := \lim_{\delta \to 0} \frac{1}{\delta} \int_{\bar{k}-\delta}^{\bar{k}} \kappa(dk) = \frac{1 - \theta/N}{1 - \theta} \alpha^{ac}(\bar{a}). \tag{50}$$

Proof. As in Lemma 14, hypothesis (ii) implies some non-decreasing function $k_t : A \to K$ gives the equilibrium matching of students with teachers, so that $k_g(a) = (1 - \theta)a + \theta k_t(a)$ gives the matching of student ability with human capital acquired when the student grows up. Then $(k_g)_{\#}\alpha = \kappa$ and $(k_t)_{\#}\alpha = N\kappa_t$, where $\kappa = \kappa_m + \kappa_m/N' + \kappa_t/N$ gives the distribution of adult skill types on K, as a sum of the distributions of worker, manager and teacher skill types. Now

$$Nk'_t(a)\kappa_t^{ac}(k_t(a)) = \alpha^{ac}(a), \tag{51}$$
$$\text{and} \quad k'_g(a)\kappa^{ac}(k_g(a)) = \alpha^{ac}(a) \tag{52}$$

is known to hold for a.e. $a \in \bar{A}$. In particular, techniques of [15] can be used to show it holds at $a = \bar{a}$ provided $k'_t(\bar{a})$ (and hence $k'_g(\bar{a})$) exists (iii) and are non-vanishing. On the other hand, the upper bound $\|\kappa^{ac}\|_{L^\infty} < \infty$ from Lemma 14 gives a positive lower bound for $k'_t(a)$ near $\bar{a}$ a.e. in (51), which precludes the possibility that $k'_t(\bar{a}) = 0$.

From (i) and the steady state constraint $\kappa = \lambda^1 + \frac{1}{N'}\lambda^2 + \frac{1}{N}\epsilon^2$ we have $k_t(\bar{a}) = \bar{k} = k_g(\bar{a})$ and $\kappa^{ac}(\bar{a}) = \kappa_t^{ac}(\bar{a})$. From (51)–(52) we conclude $Nk'_t(\bar{a}) = k'_g(\bar{a})$. On the other hand, differentiating $k_g(a) = (1 - \theta)a + \theta k_t(a)$ yields $k'_g(\bar{a}) = 1 + \theta(k'_t(\bar{a}) - 1) \geq 1 - \theta$. Solving this linear system of two equations in two unknowns gives $k'_t(\bar{a}) = \frac{1-\theta}{N-\theta}$ and

$$k'_g(\bar{a}) = \frac{1 - \theta}{1 - \theta/N}; \tag{53}$$

(52) now implies (50).

Next we consider the equilibrium wage $v(k)$ of each type of adult and payoff $u(a)$ to each type of student. The stability constraint asserts $u(a) + \frac{1}{N}v(k) - v(z(a,k)) - cb_E(z(a,k)) \geq 0$ for all a and k, with equality holding when $k = k_t(a) = \theta^{-1}k_g(a) + (1 - \frac{1}{\theta})a$. The first-order condition in k for this non-negative function to attain its minimum gives

$$v'\left(\frac{k_g(a) - (1-\theta)a}{\theta}\right) = (v'(k_g(a)) + cb'_E(k_g(a)))N\theta.$$

Taylor expanding $k_g(\bar{a} - \Delta a) = \bar{k} - k'_g(\bar{a})\Delta a + o(\Delta a)$ using $k'_g(\bar{a})$ from (53), we find a recursive relation for $v'(k)$ near $\bar{k}$:

$$v'\left(\bar{k} - \frac{1-\theta}{N-\theta}\Delta a + o(\Delta a)\right) = N\theta[v' + cb'_E]_{k=\bar{k} - \frac{1-\theta}{1-\theta/N}\Delta a + o(\Delta a)}.$$

Neglecting the $o(\Delta a)$ terms and setting $\bar{b}'_E f(x) := v'(\bar{k}-x)/c + (1 - \frac{1}{N\theta})^{-1}\bar{b}'_E - (1 - \frac{1}{N^2\theta})^{-1}b''_E(\bar{k})x$, the recursion simplifies to $f(\frac{x}{N}) = N\theta f(x)$ which is solved by constant multiples of $f(x) = x^{-\log(N\theta)/\log N}$. Thus, to leading order

$$v'(\bar{k} - \Delta k) = const|\Delta k|^{-\frac{\log N\theta}{\log N}} - \frac{c\bar{b}'_E}{1 - \frac{1}{N\theta}} + \frac{cb''_E(\bar{k})}{1 - \frac{1}{N^2\theta}}\Delta k.$$

Either the first or the second summand dominates this expression as $\Delta k \to 0$, depending on the sign of $N\theta - 1$. One might worry that $const$ depends on the sequence along which the recursion is solved, but for $N\theta > 1$ the monotonicity of v' precludes this, to yield the desired identity (49). ∎

Some remarks concerning hypotheses (i)–(iv): Proposition 7 ensures (i) holds if $N'\theta'$ and $N\theta$ are large enough, while Theorem 15 ensures (ii) holds when $c > 0$, and can be selected otherwise. We do not know conditions which guarantee (iii)-(iv), since differentiability may fail for $k_t(a)$ on a set of zero measure, and for $v(k)$ at a countable number of points. We can however, ensure that k_t is bi-Lipschitz by combining the lower bound on its derivative from Lemma 14 with the upper bound provided by Proposition 7 in case $N\theta \geq 1$. This makes failure of (iii) seem unlikely, since the value of $k'_t(a)$ would have to oscillate between these positive bounds, producing a reciprocal oscillation in $\kappa(k)$ near $\bar{k}$. Similarly, the alternative to (iv) is that jump discontinuities in the monotone function $v'(k)$ accumulate at $\bar{k}$. At least one of the three types of singular behavior must occur, and (49) seems the most likely, especially given its consistency with the divergence (23) predicted by Proposition 7. To be absolutely correct, however, one should say Theorem 16

provides strong evidence in favor of a phase transition with wage gradients diverging if and only if $N\theta \geq 1$, where the leading order behavior of (49) changes. The theorem also provides concrete quantitative predictions which can be investigated numerically.

A Optimal plans and absence of a duality gap

This appendix establishes the existence of measures achieving the maxima LP^* and $LP^*(\delta)$ in the original (11) and δ-perturbed dual problems (25), and confirms the absence of a duality gap by verifying $LP^*(\delta) = LP_*(\delta)$. While such claims are natural analogs to duality results well-known in finite-dimensional linear programming, in our infinite-dimensional context they will remain true only if we are careful to choose the correct functional analytic setting. These choices involve Banach spaces of continuous bounded functions normed by the supremum, and their dual spaces (well-known to consist of signed measures of finite total variation by the Riesz-Markov theorem). The specific spaces are indicated in the proof of the following theorem.

Its proof is based on a version of the Fenchel-Rockafellar duality theorem found in Borwein and Zhu [3] (and pointed out to us by Yann Brenier). However, as $LP_*(\delta)$ is necessarily defined by an infimum over a class of functions F_δ which includes some unbounded ones, the Fenchel-Rockafellar result based on the Banach spaces $C(\overline{A})$ and $C(\overline{K})$ yields only an inequality $LP_*(\delta) \leq LP^*(\delta)$ and not the desired equality. Fortunately, the complementary inequality is established in Proposition 8.

Theorem 17 (Optimal measures exist; duality gaps don't) *Fix δ, c_δ non-negative and θ, θ', N, N' with $\max\{\theta, \theta'\} \leq 1 \leq N$. Let α be a Borel probability measure on $\overline{A}$ with $0 < \overline{a} = \overline{k} \in \mathrm{spt}\, \alpha$, and define $z(a, k) = (1 - \theta)a + \theta k$, $b_\theta = b_E \circ z$ and $b_{\theta'}^L(a, k) = b_L((1 - \theta')a + \theta' k)$ where $b_{E/L} \in C(\overline{K})$. Then there exist feasible measures $\epsilon_\delta \geq 0$ and $\lambda_\delta \geq 0$ on $\overline{A}^2$ attaining the maximum value $LP^*(\delta)$ in the dual problem (25). This value $LP^*(\delta) \geq LP_*(\delta)$ dominates that of the infimum (24); the reverse inequalty holds if α satisfies the doubling condition (12).*

Proof. Let $H : Z \longrightarrow W$ be a bounded linear transformation between Banach spaces Z and W, on which convex functions $\varphi : Z \longrightarrow \mathbf{R} \cup \{+\infty\}$ and $\phi : W \longrightarrow \mathbf{R} \cup \{+\infty\}$ are defined. Let $\mathrm{Dom}\, \varphi := \{z \in Z \mid \varphi(z) < \infty\}$.

Define the Legendre transform of φ on the dual space Z^* by

$$\varphi^*(z^*) := \sup_{z \in Z} \langle z, z^* \rangle - \phi(z), \tag{54}$$

and define ϕ^* on W^* analogously. Here $\langle z, z^* \rangle$ denotes the duality pairing. If ϕ is continuous and real-valued at some point in $H(\mathrm{Dom}\, \varphi)$, then pp. 135-137 of [3] asserts the identity

$$\inf_{z \in Z} \varphi(z) + \phi(Hz) = \max_{z^* \in Z^*} -\varphi^*(H^* z^*) - \phi^*(-z^*),$$

as well as the fact that the maximum is attained.

In our case

$$\varphi_\delta(u, v) = \delta \langle u + v \rangle_A + \int_{[0, \bar{a}]} u(a) \alpha(da)$$

so

$$\varphi_\delta^*(\mu, \nu) = \begin{cases} 0 & \text{if } (\mu, \nu) = (\alpha + \frac{\delta}{|A|} H^1|_A, \frac{\delta}{|K|} H^1|_K) \\ +\infty & \text{else,} \end{cases}$$

while

$$\phi_\delta(\tilde{u}, \tilde{v}) = \begin{cases} 0 & \text{if } \tilde{u} \geq c_\delta b_\theta \text{ and } \tilde{v} \geq b_{\theta'}^L \\ +\infty & \text{else;} \end{cases}$$

so

$$\phi_\delta^*(\epsilon, \lambda) = \begin{cases} c_\delta \epsilon(b_\theta) + \lambda(b_{\theta'}^L) & \text{if } \epsilon \leq 0 \text{ and } \lambda \leq 0 \\ +\infty & \text{else;} \end{cases}$$

and $H : C(\bar{A}) \oplus C(\bar{K}) \longrightarrow C(\bar{A} \times \bar{K}) \oplus C(\bar{K} \times \bar{K})$ is given by

$$H \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u(a) + \frac{1}{N} v(k) - v(z(a, k)) \\ v(k') + \frac{1}{N'} v(k) \end{pmatrix},$$

so that

$$H^* \begin{pmatrix} \epsilon \\ \lambda \end{pmatrix} = \begin{pmatrix} \epsilon^1 \\ \lambda^1 + \frac{1}{N'} \lambda^2 + \frac{1}{N} \epsilon^2 - z_{\#} \epsilon \end{pmatrix}.$$

Notice φ_δ is continuous, while taking u, v large and constant makes $\phi_\delta \circ H$ finite. With these definitions (54) therefore asserts:

$$\begin{aligned} LP_*(\delta) &\leq \inf_{\substack{u \in C(\bar{A}) \\ v \in C(\bar{K})}} \varphi_\delta(u, v) + \phi_\delta(H(u, v)) \\ &= \max_{\substack{\epsilon \geq 0 \text{ on } \bar{A} \times \bar{K} \\ \lambda \geq 0 \text{ on } \bar{K} \times \bar{K}}} -\varphi_\delta^*(H^*(\epsilon, \lambda)) - \phi_\delta^*(-\epsilon, -\lambda) \\ &= LP^*(\delta). \end{aligned}$$

Here we have an inequality rather than the desired equality because the definition of $LP_*(\delta)$ involves minimizing over a broader class of feasible functions (24) which need neither be continuous nor bounded. For such functions however, Proposition 8 asserts the opposite inequality under the doubling condition (12), to conclude the proof of the theorem. ∎

References

[1] N. Ahmad, H.K. Kim, and R.J. McCann. Optimal transportation, topology and uniqueness. *Bull. Math. Sci.* **1** (2011) 13–32.

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

[3] J.M. Borwein and Q.J. Zhu. *Techniques of Variational Analysis.* CMS Books in Mathematics, 20. Springer-Verlag, New York, 2005.

[4] P.-A. Chiappori, A. Galichon, and B. Salanié. The roommate problem is more stable than you think. *Working paper*, 2012.

[5] 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.

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

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

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

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

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

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

[12] E.H. Lieb and M. Loss. *Analysis*. American Mathematical Society, Providence, 1997.

[13] R.E. Lucas, Jr. On the size distribution of business firms. The Bell Journal of Economics **9** (1978) 508–523.

[14] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** (1995) 309–323.

[15] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.* **128** (1997) 153–179.

[16] R.J. McCann. Academic wages, singularities, phase transitions and pyramid schemes. *Proceedings of the International Congress of Mathematicians (Seoul 2014), vol III Invited Lectures.* S.Y. Jang et al, eds. Kyung Moon SA, Seoul (2014) 835–849.

[17] R.J. McCann, X. Shi, A. Siow, and R. Wolthoff. The organization of the labor market with communication and cognitive skills. *J. Law, Econom. Organization* (2015) doi: 10.1093/jleo/ewv002

[18] G. Nöldeke and L. Samuelson. Investment and competitive matching *Econometrica* **83** (2015) 835–896.

[19] S. Rosen. The economics of superstars. *American Economic Review* **71** (1981) 845–858.

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

[21] N. Stokey and R. Lucas, with E. Prescott. *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge MA, 1989.