Coupled Education and Labour Markets

by

Rosemonde Lareau-Dussault

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto

© Copyright 2017 by Rosemonde Lareau-Dussault

Abstract

Coupled Education and Labour Markets

Rosemonde Lareau-Dussault
Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto
2017

This thesis combines two major projects: a dynamic and a two-dimensional extension of an existing coupled labour and education matching market ([13] and [27]).

The first extension addresses the dynamics of the distribution of skills in a population over many generations. Two overlapping generation models are proposed: the first assumes complete information, which allows (and requires) all generations to be solved simultaneously, while the second assumes incomplete information, forcing the competitive equilibrium at each subsequent generation to be found iteratively. Both models combine a labour and an education matching problem. The skill distribution for each generation of adults is determined from that of the previous generation by the educational matching market.

We present conditions for the sequence of adult skills to converge. Next, we study the asymptotic which is specific to each model. For the incomplete information model we prove that, if the sequence of wage functions over the generations converges, the limiting steady state solves the steady state model of [13], which allows for an explicit formulation for solutions of [13]. To study the limiting society of the complete information model, we introduce a new steady state model, which includes a discounting factor to reduce the impact of future generations relative to how far in the future they are.

The second extension provides proofs for certain results stated in the study of the steady state model when individuals' skills are two dimensional [27] and presents more rigorous proofs for some of their results. In particular, we prove strong duality and existence of solutions, following [13].

Pour mes soeurs, mes meilleures amies

Acknowledgements

First, I'd like to thank my supervisor Robert McCann. Robert has been very present and helpful during my doctorate, helping me not only with my projects, but also to improve my skills as a mathematician. Chapter 3 of this thesis is the product of joint work with Robert.

I also wish to thank the members of my supervisory committee, Almut Burchard and Bob Jerrard and my external examiner Adrien Blanchet.

I want to thank the UofT math department and the staff who have been not only incredibly helpful, but also very friendly and welcoming. A particular thanks to Jemima!

Though it seems long ago now, the first year of my PhD was very intense. I'm grateful for my professors who challenged me and to the classmates I worked with: Asif, Bea, Ben, Jack, Jeremy, Kevin, Mark, Nik, and Tracey.

I'd like to thank my officemates. They made my time on campus really fun, either by geeking out with me, or by slowly learning way more then they ever wanted to about magic cards.

I'd like to thank my close friends and my family, particularly my sisters and my father, who have been very patient with me during this trying period of my life. Another special thanks to Andrew, Dave, James, Jenny, and Nick who proofread this thesis.

Finally, I want to thank James, who has been exceptional moral support.

Contents

List of Figures

Chapter 1

Introduction

Optimal transportation has so many applications to economic theory that Galichon devoted a book to the subject [15]. This work concerns an economic application of optimal transportation, namely two-sided matching. Two-sided matching was first formulated in terms of matching medical residents to hospitals in the first half of the 20th century [33]. A solution to this problem is to centralize the assignment. In the middle of the 20th century, scientists came up with algorithms to stabilize the matching [14]. Another important advancement in two-sided matching theory is the formulation of the problem as a linear problem by Shapley and Shubik [35]. This new formulation uses linear optimization theory to solve matching problems, including the dual formulation of the problem. Shapley and Shubik's model takes the shape of a transportation problem.

In this thesis, we study the coupling of two matching markets: one for education and one for labour production, following the models of McCann, Shi, Siow, and Wolthoff [27], and of the same group of authors plus Erlinger [13].

Coupling education and labour in the same model allows us to study economic growth. Economic growth quantifies the increase of the growth of production of a fixed economy over time. Before the 1980s, economic growth was typically modelled and explained using exogenous factors (see [6, Section I.4] for a historical review). A way to model growth endogenously is to add human capital to the model. Human capital can represent many different aspects of an individual (or a group of individuals) such as knowledge, acquired skills, and innate abilities. These characteristics influence the productivity of the workers.

Because the productivity of workers depends on their characteristics, individuals who wish to improve their productivity may be motivated to modify these characteristics. The concept of improvements of workers has been studied in the context of task specialization. A way to model improvements of workers is called *learning by doing* and is inspired by the theory that by repeating the same task over and over, a worker improves at doing that specific task. We instead represent the improvement of workers by coupling the labour market with an education market, as in Becker and Murphy [8].

In our models, matching in both sectors (educational and labour) is represented using linear optimization. One example of such a model with a labour market was proposed by McCann and Trokhimtchouk [28], following Hao and Suen [21]. Their model associates a manager of type $m \in M$ to one or many workers of type $w \in W$. Let $p : M \times W \to \mathbf{R}$ be a function that represents the production of a manager of type m teamed up with a worker of type w. The goal of the model in [28] is to optimize the total production for the society, given a distribution for the workforce population. Using duality theory, it is also possible to view the problem as a competitive equilibrium in which individuals choose the occupation that optimizes their wage.

McCann, Shi, Siow, and Wolthoff generalize the approach of [28] to include an education sector in [27]. Students from a fixed distribution have to choose a teacher and a future employment, i.e. teacher, worker, or manager. Students' future skills as adults will be a fixed convex combination of their abilities as students and their teachers' skills. The goal is to optimize the society's production, which is a combination of labour's production and other benefits from education. In [27], individuals' skills are two-dimensional: they are comprised of a cognitive skill and a communication skill. The authors study the education and labour matching and the profession choices for individuals. They also present numerical simulations.

Erlinger, McCann, Shi, Siow, and Wolthoff [13] studied the same model in one dimension, that is, with only a cognitive skill. They present a complete analysis that includes a duality theory result and conditions for the existence and uniqueness of an optimizer. They also found some general asymptotic results for the behaviour of adult wages near the highest possible skill. Finally, McCann [26] elucidates the concept of competitive equilibrium in the one-dimensional case.

In the models of both [13] and [27], the adults' distribution of skills is in a steady-state form. It is determined by the distribution of abilities of students and the matching of students to teachers. It would be more realistic to assume that the distributions of adults' skills evolve from one generation to another, allowing the current generation's distribution of adults' skills and education market to determine the distribution of adults' skills for the next generation. We introduce two such models and study whether for these models, the sequence of distributions of adults' skills will converge to a steady state distribution of adults' skills, like the one in [13] and [27]. Of course, this depends on initial conditions and hypotheses of the models.

In order to incorporate several generations in our models, we use overlapping generation models. An overlapping generation model is a type of model in economics that combines a finite number (N) of generations. Each individual will then interact N distinct times with the society. A classical example considers adults to be either working or retired (in which case $N = 2$).

Overlapping generation models have been introduced to study interest rates [34]. Overlapping generation models have also been used to study the economics of education, as they offer a way to incorporate aging. For example, they are used to compare different methods of funding in education ([41], [20], and [19]). Artige [5] studies the dynamics of a population with a model that combines education and production. In all of these models, students are assumed to be identical and adults make educational choices for their dependants. In our models, students' abilities are heterogeneous and students make their own choices regarding their education and future profession.

With this goal in mind, we adopt models which borrow their production technologies from [13] and [27], but in which the distribution of adults' skills in the present generation is induced by the educational matching of the previous generation. We fix an initial distribution of skills for the adults, κ_1, not necessarily related to the students' distribution of abilities. From it, our models create a sequence of distributions of adults' skills determined, at every step $i > 1$, by the students' distribution of abilities and educational choices of students from the previous generation found in step $i - 1$.

The educational choices for a fixed generation will generate a new distribution for the skills of the next generation of adults. The focus of the study is on the dynamics of the distributions of skills of the adult population over different generations.

Chapter 3 (joint work with Robert McCann) introduces and studies two dynamic versions of [13]. For the first model, we assume that the students have incomplete information. That is, students have to estimate their wage as adults. For the second model, we study a complete information model in which we study all generations together. This allows for students to rationally forecast their exact future wage. The main results of Chapter 3 are Theorems 3.4.4, 3.5.3 and 3.6.4. They give conditions which guarantee that any initial distribution of skills will evolve in the long time limit toward a unique steady state.

In the incomplete information case (introduced in Section 3.1), the limiting steady state comes from [13]. This result justifies the validity and interest of the steady state model of [13]. Although the steady state of Erlinger et al is in general specified as the solution to a variational optimization problem, under the conditions in question we show it can instead be constructed directly. In a concrete example, this explicit construction allows us to support a conjecture from [13] concerning the presence and form of wage singularities in Section 3.7.

For the complete information case (introduced in Section 3.2), we introduce a new steady state which includes discounting, i.e. such that the production of the current and near future generations is more important than production of future generations.

Another important extension to the one dimensional steady state model from [13] is to consider the case where the skill and ability space is two dimensional. As discussed in [27], cognitive skills are not enough to explain the professional success and estimated wage of an individual. Indeed, studies show that non-cognitive skills influence educational and professional opportunities.

In order to quantify non-cognitive skills, McCann, Shi, Siow, and Wolthoff decided in [27] to include a communication skill in their model. They assume that production in a company can be separated into two tasks; one of production from the worker and one of supervision from the manager. In their model, only the supervision task requires communication skill (as his or her communication skill increases, a manager can supervise more workers) and the efficiency of the team is a combination of the cognitive skills of the worker and of the manager.

In [27], the authors describe solutions to their model. For example, they show that individuals will be specialized. That is, if an individual does the manager's task in a team, then he or she won't do the worker's task in another team. They show that teamwork increases productivity. Chapter 4 extends duality and existence results from [13] to the case where individuals' skills are two-dimensional (see [27]) and provide rigorous proofs for some of the results claimed in [27].

This thesis is organized as follows: Chapter 2 introduces notation, mathematical concepts, and specific hypotheses for the following models and presents the steady-state model from [13]. Chapter 3 introduces and studies two dynamic versions of [13]. Chapter 4 extends duality and existence results from [13] to the case where individuals' skills are two-dimensional (see [27]) and provide rigorous proofs for some of the results claimed in [27].

Chapter 2

Notation and Preliminaries

In this chapter, we introduced notation and concepts used in this thesis.

2.1 Functions and measures

Definition 2.1.1 (Proper function)**.** *A function* $f : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ *is said to be* proper *if there exists an* $x \in \mathbf{R}^n$ *such that* $f(x) \in (-\infty, \infty)$*.*

Definition 2.1.2 (Lower semi-continuous function)**.** *A function* $f : [a, b] \to \mathbf{R} \cup \{\infty\}$ *is* lower semi-continuous *at a point* $x \in [a, b]$ *if for all* $\epsilon > 0$*,*

$$\liminf_{\epsilon \to 0} f(x + \epsilon) \geq f(x).$$

A function $f : [a, b] \to \mathbf{R} \cup \{\infty\}$ *is* lower semi-continuous *if it is lower semi-continuous at all points* $x \in [a, b]$*.*

In words, a function is lower semi-continuous if its limits all dominate its values (see figure 2.1).

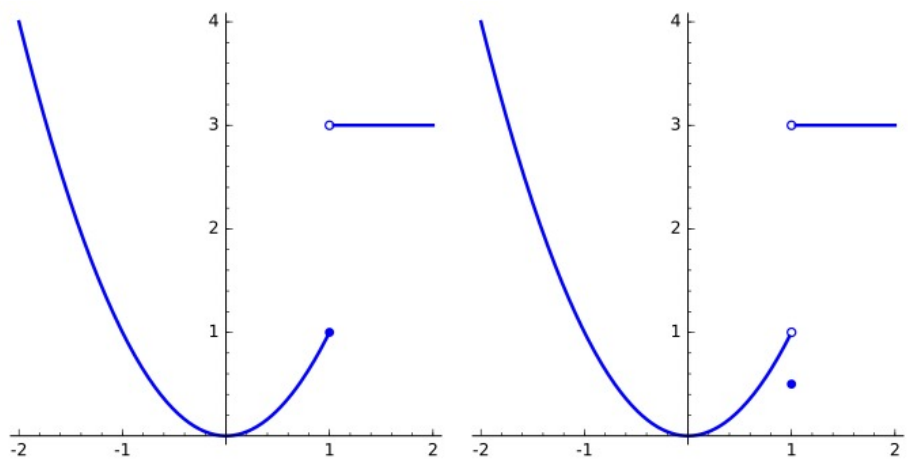

Figure 2.1: Two examples of lower semi-continuous function.

Definition 2.1.3 (Lower semi-continuous hull)**.** *Let* $f : [a, b] \to \mathbf{R}$*. The* lower semi-continuous hull *of* f *is a function lsc* $f : [a, b] \to \mathbf{R}$*, such that lsc* $f(x) = \liminf_{\epsilon \to 0} f(x + \epsilon)$*.*

Note that lsc f coincides with f at points where f is continuous.

For a function

$$f : (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n \mapsto f(x_1, x_2, \ldots, x_n) \in \mathbf{R},$$

we'll use the following notation for the partial derivative with respect to x_i

$$f_{x_i} = \frac{\partial}{\partial x_i} f.$$

We will use π_i to denote the projection on the ith coordinate of an element in a product space:

$$\pi_i : (x_1, x_2) \in \mathbf{R}^2 \mapsto x_i \in \mathbf{R}.$$

More generally, $\pi^{i_1, i_2, \ldots, i_n}$ is the projection on the $i_1, i_2, \ldots, i_n$ coordinates.

Let $f : X_1 \to X_2$ be a measurable map and let μ be a measure on X_1. The push forward of μ by f is a measure on X_2 defined by

$$f_{\#}\mu(A) = \mu\left(f^{-1}(A)\right)$$

for each $A \subset X_2$. A measurable function $g : X_2 \to \mathbf{R}$ is integrable with respect to $f_{\#}\mu$, if and only if $g \circ f : X_1 \to \mathbf{R}$ is integrable with respect to μ and in that case,

$$\int_{X_2} g d(f_{\#}\mu) = \int_{X_1} g \circ f d\mu.$$

Let $\mathcal{P}(X)$ denote the set of (Borel) probability measures on X. A measure $\mu \in \mathcal{P}(\mathbf{R})$ is said to be absolutely continuous if for every measurable subset $A \subset \mathbf{R}$ with zero Lebesgue measure $\mathcal{L}(A) = 0$, we also have $\mu(A) = 0$. In that case, we can write $d\mu(x) = \mu^{ac}(x)dx$, where μ^{ac} is a Lebesgue integrable function. In this thesis, for an absolutely continuous measure μ, we sometimes abuse notation by using μ to represent both the measure and its density μ^{ac}. The set of absolutely continuous measures on X will be denoted $\mathcal{P}^{ac}(X)$.

The support of a positive measure $\gamma \in \mathcal{P}(\mathbf{R}^2)$ is smallest closed set S such that γ vanished outside of S:

$$\text{spt }(\gamma) = \left\{(x, y) \in \mathbf{R}^2 | \gamma(U) > 0 \text{ for every open neighborhood } U \text{ of } (x, y)\right\}.$$

Definition 2.1.4 (Positive Assortative Measure)**.** *A measure* $\lambda \in \mathcal{P}(\mathbf{R}^2)$ *is said to be* positive assortative *if for all* $(x_1, y_1), (x_2, y_2) \in spt\ \lambda$*,*

$$(x_1 - x_2)(y_1 - y_2) \geq 0;$$

that is, λ *is supported on a non-decreasing subset of* $\mathbf{R}^2$*.*

Lemma 2.1.5 ([1, Lemma 3.1])**.** *For a measure* $\gamma \in \mathcal{P}(\mathbf{R}^2)$ *with support on the graph of a function* $T : \mathbf{R} \to \mathbf{R}$*, then* T *is* $\pi_{1\#}\gamma$ *measurable and* $\gamma = (\mathbb{1} \times T)_{\#}(\pi_{1\#}\gamma)$*, where*

$$\mathbb{1} \times T : x \in \mathbf{R} \mapsto \left(x, T(x)\right) \in \mathbf{R}^2.$$

The map $T : \mathbf{R} \longrightarrow \mathbf{R}$ is called a matching function (or Monge map) from spt $\pi_{1\#}\gamma$ to spt $\pi_{2\#}\gamma$. If $\pi_{1\#}\gamma$ is absolutely continuous, then γ is positive assortative if and only if T is a *non-decreasing* matching function.

A function $f : \mathbf{R}^2 \to \mathbf{R}$ is said to be *submodular* if $a < a'$ and $k < k'$ implies $f(a,k) + f(a',k') \leq f(a,k') + f(a',k)$ and strictly submodular if the inequality is strict.

The following lemma will allow us to quantify derivatives of wage functions. The proof is postponed to Appendix A.

Lemma 2.1.6. *If $f : \bar{A} \times \bar{K} \to \mathbf{R}$ is locally Lipschitz in a, uniformly in k, then $g(a) = \sup_{k \in \bar{K}} f(a,k)$ is locally Lipschitz and for each $a \in \bar{A}$ for every neighbourhood $a \in U_a \subset \bar{A}$ the following bounds hold:*

$$\inf_{k \in \bar{K},\ a' \in U_a} f_a(a',k) \leq g'(a) \leq \sup_{k \in \bar{K},\ a' \in U_a} f_a(a',k)$$

in the pointwise a.e. senses, where $f_a(a',k) = \frac{\partial}{\partial a} f(a,k)\big|_{a=a'}$. Similarly, if f is locally semi-convex in $a \in \bar{A}$, uniformly in $k \in \bar{K}$, then $g(a)$ is locally semi-convex and for all $a \in \bar{A}$ and for every neighbourhood $a \in U_a \subset \bar{A}$

$$g''(a) \geq \inf_{k \in \bar{K},\ a' \in U_a} f_{aa}(a',k)$$

in the pointwise a.e. senses, where $f_{aa}(a',k) = \frac{\partial^2}{\partial a^2} f(a,k)\big|_{a=a'}$.

2.2 Topological Spaces

We topologize measures using *weak-$*$ convergence*, or narrow convergence, which refers to convergence against bounded continuous test functions. We denote weak-$*$ convergence of μ_i to μ by $\mu_i \xrightarrow{w^*} \mu$. For *probability* measures, weak-$*$ convergence can be metrized by the Prokhorov metric [30].

The set of wage functions of interest are real functions of $K = [0, \bar{k})$ that are bounded, increasing, and convex. Being bounded, increasing and convex, these functions are continuous, except maybe at $\bar{k}$ and by the Arzelà-Ascoli theorem, for any sequence of such functions $\{ f_j \}$ there exists a subsequence that converges uniformly on compact subsets. This means that equicontinuity can only fail at the upper end point. Therefore, we need a metric that doesn't weight the right end point $\bar{k}$. The distance function we use to metrize uniform convergence on compact subsets of $K = [0, \bar{k})$ is

$$d_K(f,g) = \sum_{i=1}^{\infty} 2^i \left\| f - g \right\|_{L^\infty\left([0, \bar{k} - \frac{1}{i}]\right)}. \tag{2.1}$$

The upper endpoint is excluded from K to permit wage singularities for superstars; since the analysis of [13] suggests these singularities remain in $C^0 \setminus C^1$, we often make the simplifying hypothesis that in any given model, all wages are uniformly bounded.

The next lemma presents compactness properties of the distance function d_K. The proof is postponed to Appendix A.

Lemma 2.2.1 (Properties of d_K)**.** *Let $\{ f_j \}$ be a sequence of real functions of K.*

(a) We have that $f_j \to f$ in the sense of uniform convergence on compact subsets of K if and only if $d_K(f_j, f) \to 0$.

(b) Consider a sequence of absolutely continuous measures with respect to the Lebesgue measure $\{ \mu_i \}$ *such that* $\mu_i \to \mu$ *in the weak-$*$ topology, where* $\mu \ll \mathcal{L}$*. Suppose*

$$\lim_{\delta \to 0} \left(\sup_i \mu_i \left([\bar{k} - \delta, \bar{k}] \right) \right) \to 0^1. \tag{2.2}$$

Let $\{ f_j \}$ *be a uniformly bounded set of continuous functions on* K *such that* $f_i \overset{d_K}{\to} f$ *which is also continuous on* K*. In this case, we have*

$$\mu_i(f_i) \to \mu(f).$$

(c) The non-decreasing convex functions on K *taking values in* $[-b, b] \subset \mathbf{R}$ *form a compact set when metrized by* d_K*.*

2.3 Γ-Convergence

In order to prove that sequences of optimizers over generations converge to an optimizer, we use the notion of Γ-convergence. We use definitions and theorems from [11].

Definition 2.3.1 ([11, Definition 1.5, Γ-convergence])**.** *Let* (X, d) *be a metric space. For all* i *let* $f_i : X \to \mathbf{R}$*. We say that the sequence* $\{f_i\}$ *Γ-converges in* X *to* f_∞ *if for all* $x \in X$ *we have:*

1. *(lim inf inequality) for every sequence* $\{ x_i \}$ *converging to* x

$$f_\infty(x) \leq \varliminf_{i \to \infty} f_i(x_i);$$

2. *(lim sup inequality) there exists a sequence* $\{ x_i \}$ *converging to* x *such that*

$$f_\infty(x) \geq \varlimsup_{i \to \infty} f_i(x_i).$$

The function f_∞ *is called the Γ-limit of* $\{ f_i \}$*, and we write* $f_\infty = \Gamma - \lim_i f_i$*.*

In order to have sequences of minimizers converge to a minimizer, we need to add a coerciveness condition to Γ-convergence.

Definition 2.3.2 ([11, Definition 1.19, coerciveness conditions])**.** *A function* $f : X \to \bar{\mathbf{R}}$ *is coercive if for all* $t \in \mathbf{R}$ *the set* $\{ f \leq t \}$ *is precompact. A function* $f : X \to \bar{\mathbf{R}}$ *is mildly coercive if there exists a non-empty compact set* $K \subset X$ *such that* $\inf_X f = \inf_K f$*. A sequence* $\{ f_i \}$ *is equi-mildly coercive if there exists a non-empty compact set* $K \subset X$ *such that* $\inf_X f_i = \inf_K f_i$ *for all* i*.*

We are now ready to state the conditions under which sequences of minimizers converge to a minimizer.

Theorem 2.3.3 ([11, Theorem 1.21])**.** *Let* (X, d) *be a metric space, let* $\{ f_i \}$ *be a sequence of equi-mildly coercive functions on* X*, and let* $f_\infty = \Gamma - \lim_i f_i$*; then* $\min_X f_\infty$ *is attained and*

$$\min_X f_\infty = \lim_i \inf_X f_i.$$

[1] Note that if we have a uniform L^∞ bound on μ_i, hypothesis (2.2) is satisfied.

Moreover, if $\{ x_i \}$ *is a precompact sequence such that* $\lim_i f_i(x_i) = \lim_i \inf_X f_i$*, then every limit of a subsequence of* $\{ x_i \}$ *is a minimum point for* f_∞*.*

2.4 Duality

We include here a general duality theorem that we'll use to prove duality of many model.

Lemma 2.4.1 ([10, Theorem 4.4.3] Fenchel-Rockafellar Duality)**.** *Let* A *and* B *be Banach spaces, and*

$$\varphi : A \to \mathbf{R} \cup \{ \infty \}$$
$$\phi : B \to \mathbf{R} \cup \{ \infty \}$$

be convex functions with associated Legendre transforms

$$\begin{aligned} \varphi^* : A^* &\to \mathbf{R} \cup \{ \infty \} \\ x^* &\mapsto \sup_{x \in A} \langle x, x^* \rangle - \varphi(x) \qquad \text{and} \\ \phi^* : B^* &\to \mathbf{R} \cup \{ \infty \} \\ y^* &\mapsto \sup_{y \in B} \langle y, y^* \rangle - \phi(y). \end{aligned}$$

Let $H : A \to B$ *and let* $H^* : B^* \to A^*$ *be its adjoint operator, i.e.*

$$\forall x \in A, y \in B, \qquad \langle H(x), y^* \rangle = \langle x, H^*(y^*) \rangle .$$

Let Dom $\varphi := \{ x \in A \mid \varphi(x) < \infty \}$*. Then if* ϕ *is continuous and real-valued at some point in* $H(\text{Dom } \varphi)$ *(condition* (4.3.2) *in [10]),*

$$\inf_{a \in A} \varphi(a) + \phi(Ha) = \max_{y^* \in B^*} -\varphi^*(H^* y^*) - \phi^*(-y^*).$$

2.5 Hypotheses

This subsection introduces notation and definitions needed for the proposed models. For more detail on the choices of notation and hypotheses, see [27].

Let $A = [0, \bar{k}) = K$ be the space of abilities of students and skills of adults respectively. A student's ability, combined with the skill or his or her teacher, will impact his or her skill when becoming an adult. For the adults in the labour market, their skills quantify how much they produce.

Fix $\theta, \theta' \in (0, 1)$ and let

$$z_E : A \times K \to K : (s, t) \mapsto (1 - \theta)s + \theta t \tag{2.3}$$

be the future skill of a student of skill s who chooses a teacher of skill t and

$$z_L : K \times K \to \mathbf{R} : (w, m) \mapsto (1 - \theta')w + \theta' m$$

be the effectiveness of a team with a worker of skill w and a manager of skill m.

With these functions, we can estimate the general benefit of pairing a student of type s with a teacher of type t to be $cb_E(z_E(s, t))$ where $c \geq 0$ is fixed. Similarly, the productivity of a worker of skill w with a manager of skill m is $b_L(z_L(w, m))$.

The functions b_E and b_L can theoretically be as general as we wish them to be. In order to ensure that the wage functions are positive, non-decreasing and convex, we restrict b_E and b_L. We assume b_E and b_L are both strictly positive, increasing, convex, and twice differentiable. More precisely, we suppose both functions and their derivatives satisfy the following positive lower bounds:

$$\begin{array}{rcccl} 0 & < & \underline{b}_{E/L} & = & b_{E/L}(0) \\ 0 & < & \underline{b}'_{E/L} & = & b'_{E/L}(0) \\ 0 & < & \underline{b}''_{E/L} & = & \inf_{k \in [0, \bar{k})} b''_{E/L}(k). \end{array} \tag{2.4}$$

A case of interest is when $b_{E/L}$ is an exponential function, because it means that the production function is a transformation of the Cobb-Douglas production function with constant returns to scale. The logarithmic transformation allows the objective function to be convex and for $b_{E/L}$ to satisfy (2.4).

The distribution of students is assumed to be known and constant. We will denote it by $\alpha \in \mathcal{P}(A)$.

2.6 Steady State Model

In this model, the goal is to find an optimal measure, ϵ, correlating students to teachers and an optimal measure, λ, correlating workers to managers. Let N be the number of students a teacher can teach, and N' be the number of workers a manager can supervise.

The distribution of adults will be induced by z_E and the pairing of students/teachers ϵ through the formula: $\kappa = z_{E\#}\epsilon$. These measures are found solving the following linear problem (see [13]):

$$\max_{0\leq\epsilon,\lambda} \int_{A\times K} cb_E\big(z_E(s,t)\big)d\epsilon(s,t) + \int_{K\times K} b_L\big(z_L(w,m)\big)d\lambda(w,m) \tag{2.5}$$

s.t. $\alpha = \pi_{1\#}\epsilon$ i.e. the distribution of students is known and (2.6)

$z_{E\#}\epsilon = \frac{\pi_{2\#}\epsilon}{N} + \pi_{1\#}\lambda + \frac{\pi_{2\#}\lambda}{N'}$ i.e. the distribution of adults is induced by z_E and ϵ. (2.7)

The measure of adults' skills κ is the sum of the measures of teachers' skills $\left(\kappa^t = \frac{\pi_{2\#}\epsilon}{N}\right)$, workers' skills $\big(\kappa^w = \pi_{1\#}\lambda\big)$, and managers' skills $\left(\kappa^m = \frac{\pi_{2\#}\lambda}{N'}\right)$.

Let $v : K \to \mathbf{R}$ be the wage function for adults and $u : A \to \mathbf{R}$ be an indirect utility for students, then the dual problem of (2.5) is:

$$\inf_{u,v} \int_A u(s)\alpha(s)$$

s.t. $cb_E(z_E(s,t)) + v(z_E(s,t)) \leq u(s) + \frac{v(t)}{N}$ i.e. the education market is stable and (2.8)

$b_L(z_L(w,m)) \leq v(w) + \frac{v(m)}{N'}$ i.e. the labour market is stable.

Chapter 3

Dynamics of a coupled education and labour matching over generation

In this chapter, we use overlapping generation models with two overlapping generations to study educational choices. Students have heterogeneous abilities that influence their personal educational choices and choices of their future profession. The educational choices for a fixed generation will generate a new distribution for the skills of the adult population of the next generation. The focus of the study is on the dynamic of the distributions of skills of the adult population over different generations.

We fix an initial distribution of skills for the adults, κ_1, a measure on K, which isn't necessarily related to the students' distribution of abilities. We create a sequence of distributions of adults' skills determined, at every step $i > 1$, by the students' distribution of abilities and educational choices of students from the previous generation found in step $i - 1$. We study the dynamic of the sequence of measures representing the adults' distribution of skills.

There are many ways to model the overlapping generation problem we want to study. We will present two of them. Both will use the same production function and make the same hypotheses as [27] and [13]. For the first model, we assume that the students have incomplete information. That is, students have to estimate their wage as adults based only on the wage functions of previous generations. For the second model, we study a complete information model in which we study all generations together. This allows for students to rationally forecast their exact future wage.

Overall this chapter introduces two new models describing a society with an education and a labour market that evolves from one generation to another according to individuals' educational choices. By deriving a closed form representation of the dynamics in terms of the inverses of the cumulative distribution functions of the adults' skill distributions, we are able to solve these models explicitly. In the incomplete information model we prove, under certain conditions, that the sequence of adults' skill distributions converge to an adults' skill distribution associated to a solution for the steady state model studied in [13]. When this is the case, we can use these solutions to find explicit solutions for the steady state model by studying the limiting case.

To formulate the complete information model with an infinite time horizon we use discounting to guarantee convergence. We characterize the limit of this model as the solution of a steady problem with discounting, proposed here for the first time. Unlike the other models discussed, this model is not purely variational, but rather takes the form of a family of variational problems depending on parameters plus a self-consistency condition relating the parameters to the solution. We sketch a duality theory for this family and employ the techniques described above to deduce that for a suitable choice of parameters, a self-consistent solution exists and is the unique limit of the time dynamics.

The main results of this chapter are Theorems 3.4.4, 3.5.3, and 3.6.4. Theorem 3.4.4 proves that for any initial distribution of skills, the sequence of distributions over infinite generations will converge to a predetermined distribution under suitable conditions. Theorem 3.5.3 shows that, if the sequence of wage functions evolving from one generation to another is convergent, sequences of solutions converge to a solution to the steady state model from [13]. Note that we give conditions for the sequence of wage functions to be convergent in Proposition 3.5.2. Finally, Theorem 3.6.4 gives conditions under which solutions of the complete information model converge to solutions of a new steady state model which include a discounting factor.

This chapter is organized as follows: Section 3.1 introduces the model with incomplete information and Section 3.2 introduces the model with complete information proposed in this chapter. Section 3.3 extends results from [13] to our models and gives preliminary results. Section 3.4 studies the dynamic of sequences of adults' distribution of skills and Sections 3.5 and 3.6 separate the study of the limit of a sequence of solutions for the incomplete and the complete information models. Using results from Sections 3.4 and 3.5 we construct explicit solutions for the steady state model from [13] in Section 3.7.

To complement this section, a finite time horizon version of the complete information model is introduced and used to prove some facts about the complete information model in Appendix B. In this chapter, we will use constants and functions defined in subsection 2.5. This chapter is joint work with Robert McCann.

3.1 Incomplete Information Model

We propose two models. For the first, suppose that students don't have all the needed information to optimize their educational choices. That is, they can't forecast their wage as adults and they have to estimate it. There is an important decision to be made in this case. That is, how to approximate a student's future wage? In this chapter, we assume every student will use the same approximation: they will use the wage function for the previous generation of adults to represent the wage function when they become adults. This allows the model to stay simple and maintain some semblance of realism. One could also use the wage function for the current generation of adults or any other wage function such as a combination of wage functions from previous generations.

We model this problem with a sequence of infinite dimensional linear problems. Each of these problems optimize the total production for one generation. For the first generation, we begin with an initial distribution for adults κ_1 and an estimated future wage for students $\tilde{v}_1$ assumed to be positive, convex and non-decreasing.The goal is to identify a matching $\epsilon_1 \in \mathcal{P}(A \times K)$ between students and teachers and a matching $\lambda_1 \in \frac{N'}{N} \frac{N-1}{N'+1} \mathcal{P}(K \times K)$ between workers and managers for the first generation. To do so, we solve an optimization problem. The dual of this optimization problem seeks competitive wage functions $v_1 : K \to \mathbf{R}$ and $u_1 : A \to \mathbf{R}$.

We will present both the primal incomplete information model (3.1) and the dual incomplete information model (3.4). Their optimal solutions, (ϵ, λ) and (u, v) respectively, form a competitive equilibrium [26]. That is, solving the primal problem (3.1) is equivalent to solving the dual problem (3.4). The primal problem optimizes the estimated production of the entire society by allocating a role and a matching partner to each individual. This means that, given a distribution of abilities for students and a distribution of skills for adults, it finds matching measures ϵ and λ. The dual problem finds optimal wages that minimize the sum of all the wages individuals will get. That is, it takes the point of view of someone who owns the entire society and hires every individual. This fictitious owner wants to pay as little as possible, but has to pay at least as much as the individuals are producing in order to retain them; otherwise his workforce would be unstable. The fact that the primal and dual problems are equivalent, means that the optimal labour matching λ produces enough to pay the workers and the managers according to the optimal wage function v and the optimal educational matching ϵ produces enough to pay the teachers their expected salaries and to satisfy the benefit from education u. In other words, the matchings (ϵ, λ) and wages (u, v) combine to satisfy the estimated budget constraint.

Once we have a competitive equilibrium for the first generation, we set $\kappa_2 = z_{E\#}\epsilon_1$ to be the distribution of adults for the next generation. We now consider the same optimization problem, but with the distribution of adults being κ_2 instead of κ_1 and the estimated future wage function for students to be $\tilde{v}_2 = v_1$. We iterate this process, always setting $\kappa_{i+1} = z_{E\#}\epsilon_i$ and $\tilde{v}_{i+1} = v_i$.

At every step i, we fix $\kappa = \kappa_i \in \mathcal{P}(K)$ to be the current distribution of adults, and $\tilde{v} = \tilde{v}_i$ to be the estimated future wage function for students at step i. At each step, the linear optimization problem to solve is:

$$\max_{0\leq \epsilon,\lambda} \int_{A\times K} \Big(cb_E\big(z_E(s,t)\big) + \tilde{v}\big(z_E(s,t)\big)\Big) d\epsilon(s,t) + \int_{K\times K} b_L\big(z_L(w,m)\big) d\lambda(w,m) \quad \text{s.t.} \tag{3.1}$$

$$\alpha = \pi_{1\#}\epsilon \qquad \text{i.e. the distribution of students' skills is known and} \tag{3.2}$$

$$\kappa = \frac{\pi_{2\#}\epsilon}{N} + \pi_{1\#}\lambda + \frac{\pi_{2\#}\lambda}{N'}. \tag{3.3}$$

The second constraint (3.3) represents the fact that the distributions of teacher, worker and manager skills sum to the known distribution of adult skills. The objective function represents the external benefits from education plus the expected future wages of the current students, plus the total production of the labour market. The constraints represent the fact that the distributions of student skills and adults skills are known.

For fixed κ and $\tilde{v}$, we can associate to (3.1) the following dual problem:

$$\inf_{u,v} \int_A u(s)\alpha(s) + \int_K v(k)\kappa(k) \tag{3.4}$$

$$\text{s.t.} \quad cb_E(z_E(s,t)) + \tilde{v}(z_E(s,t)) \leq u(s) + \frac{v(t)}{N} \qquad \text{i.e. the education market is stable and} \tag{3.5}$$

$$b_L(z_L(w,m)) \leq v(w) + \frac{v(m)}{N'} \qquad \text{i.e. the labour market is stable.} \tag{3.6}$$

The proof that the values of (3.1) and (3.4) coincide, and that both are attained (called strong duality) parallels the proof of duality in [13] and uses the doubling condition

$$\int_{[\bar{a}-2\Delta a,\bar{a}]} \alpha(da) \leq C \int_{[\bar{a}-\Delta a,\bar{a}]} \alpha(da). \tag{3.7}$$

(see also Section B.1). More precisely the infimum (3.4) is attained by a pair of continuous convex functions u, v on K which formal asymptotics suggest remain bounded near $\bar{k}$ [13]; we'll generally assume this boundedness for simplicity.

The dual problem represents the perspective of someone that has to pay wages to the adults and give students their benefit for education. It minimizes the wages to pay, under the constraint that the individuals receive enough to be retained.

The wage function for generation i is $v_i = v$, where v is an optimizer of (3.4) with $\kappa = \kappa_i$ and $\tilde{v} = \tilde{v}_i$. The function u represents the personal benefit of education. By assumption, $\tilde{v}_{i+1} = v_i$.

One goal of this chapter is to study the evolution of the sequence of measures $\{ \kappa_i \}$ obtained by solving a version of (3.1)–(3.3) and (3.4)–(3.6) for every generation, to obtain $(\epsilon_i, \lambda_i; u_i, v_i) = (\epsilon, \lambda; u, v)$ from $(\kappa, \tilde{v}) = (\kappa_i, \tilde{v}_i)$ at the ith iteration. In other words, setting

$$G^I : (\kappa_i, \tilde{v}_i) \in \mathcal{P}(K) \times C(K) \mapsto (\kappa_{i+1}, \tilde{v}_{i+1}) = (z_{E\#}\epsilon_i, v_i) \in \mathcal{P}(K) \times C(K) \tag{3.8}$$

it is equivalent to study the dynamics of iterating G^I. Note that G^I depends on α, b_E, b_L, N, N', θ and θ'.

Before we study (3.1) and the dynamics of G^I, we introduce another model. The general idea is the same, but we assume students can forecast their future wages with complete accuracy.

3.2 Complete Information Model

This section proposes a model for a problem that is similar to the one presented in Section 3.1. The only difference is that students know what their salary will be once they become adults. To ensure knowledge of future wages, we need a model that will solve all generations at the same time.

Let κ_1 be an absolutely continuous Borel probability measure that represents the initial distribution for adults. The goal is to optimize the society's total production over all future generations, by separating adults by profession, i.e. worker, manager and teacher, and by matching workers to managers and students to teachers. That is, we are seeking sequences of measures

$$\{ \epsilon_i \}_{i=1}^{\infty} \qquad \text{and} \qquad \{ \lambda_i \}_{i=1}^{\infty},$$

where ϵ_i represents the matching between students and teachers for generation i and λ_i represents the labour matching for generation i. These measures have to satisfy the following population (or market clearing) constraints:

$$\pi_{1\#}\epsilon_i = \alpha \qquad i = 1, 2, \ldots$$

i.e. the distribution of students' skills is known,

$$\pi_{1\#}\lambda_1 + \frac{1}{N'}\pi_{2\#}\lambda_1 + \frac{1}{N}\pi_{2\#}\epsilon_1 = \kappa_1$$

i.e. the first distribution of adults' skills is known, and

$$\pi_{1\#}\lambda_i + \frac{1}{N'}\pi_{2\#}\lambda_i + \frac{1}{N}\pi_{2\#}\epsilon_i = z_{E\#}\epsilon_{i-1} \qquad i = 2, 3, \ldots$$

i.e. distributions of adults' skills are induced.

The last set of constraints represents the fact that the sum of the teacher, worker and manager skills at generation i agrees with the distribution of adult skills at generation i induced by the education matching at generation $i - 1$.

The goal is to optimize the total productivity of the society. We introduce a discount factor $e^{-\beta}$ (take $\beta = 0$ to remove it) to represent the fact that immediate gain is more valuable than future gain. The proposed complete information model then is the maximization:

$$C(\kappa_1) := \sup_{(\epsilon_i,\lambda_i)_{i=1}^{\infty}} \sum_{i=1}^{\infty} e^{-\beta i} \left(\int_{A\times K} cb_E \circ z_E d\epsilon_i + \int_{K\times K} b_L \circ z_L d\lambda_i \right) \tag{3.9}$$

$$\text{s.t.} \begin{cases} \pi_{1\#}\epsilon_i & = \alpha \\ \pi_{1\#}\lambda_i + \frac{1}{N'}\pi_{2\#}\lambda_i + \frac{1}{N}\pi_{2\#}\epsilon_i & = z_{E\#}\epsilon_{i-1} \end{cases} \qquad i = 1, 2, \dots$$

where $\epsilon_0 = (\mathbb{1} \times \mathbb{1})_{\#}\kappa_1$ by convention.

Note that the future wage function for students does not appear in the objective function. This is because the future skill of students will actually be included in the model for their role in the next generation. In (3.1), students' estimated wages are used in the objective function to represent their future contribution to the society, but this is not relevant in this new model.

By using a finite time horizon $T < \infty$ to approximate this infinite time horizon model, we will show in Section B.1 that the dual of this problem is given by the $T \to \infty$ limit:

$$C(\kappa_1) = \inf_{(u_i,v_i)_{i=1}^{\infty}} \int_K e^{-\beta} v_1(k) d\kappa_1(k) + \sum_{i=1}^{\infty} e^{-\beta i} \int_A u_i(s) d\alpha(s) \qquad \text{s.t} \tag{3.10}$$

$$u_i(s) + \frac{1}{N} v_i(t) \geq cb_E(z_E(s,t)) + e^{-\beta} v_{i+1}(z_E(s,t)) \qquad i = 1, 2, \dots \quad \text{educational stability;}$$

$$v_i(w) + \frac{1}{N'} v_i(m) \geq b_L(z_L(w,m)) \qquad i = 1, 2, \dots \quad \text{labour market stability.}$$

Let

$$H^C : \mathcal{P}(K) \to \mathcal{P}(K)$$
$$\kappa_1 \mapsto \kappa_2 = z_{E\#}\epsilon_1 \tag{3.11}$$

be a function that sends the initial distribution of adult skills to the distribution of adult skills at the next generation. Note that H^C depends on α, β, b_E, b_L, N, N', θ and θ'.

A priori it is not evident that $\kappa_{i+1} = H^C(\kappa_i)$, except in the defining case $i = 1$. However, in the next subsection, we will re-express (3.9) in order to show, in Corollary 3.2.4, that $\kappa_{i+1} = H^C(\kappa_i)$ for all $i \geq 1$. The proof is inspired by dynamic programming. Indeed, we'll show that optimizers of (3.9) satisfy a functional equation. Our goal thereafter will be to study dynamics of G^I and H^C.

Also, it is not evident that H^C is well defined. In fact, H^C will only be well defined when ϵ_1 in the solution of (3.9) is unique. This will be the case, whenever v_1 is strictly convex, increasing and bounded because in that case all following v_i are also strictly convex and increasing from Proposition 3.3.1 and then optimizer $(\epsilon_i, \lambda_i)_{i=1}^{\infty}$ will be unique (see [13, Theorem 15]).

3.2.1 Dynamic Programming Description

We begin by re-expressing the problem (3.9) using a functional equation. To do so, we need some new notation. Let

$$\mathcal{N} : \mathcal{P}(K) \to 2^{\mathcal{P}(K)} \tag{3.12}$$

($\mathcal{N}$ is for next) be a set valued function that sends each distribution of adults' skills $\kappa_1 \in \mathcal{P}(K)$ to the set of feasible adults' distributions of skills for the next generation. That is,

$$\mathcal{N}(\kappa_1) = \left\{ \kappa_2 \in \mathcal{P}(K) | \exists \epsilon \in \mathcal{P}(A \times K) \text{ such that } \pi_{1\#}\epsilon = \alpha,\ \frac{1}{N}\pi_{2\#}\epsilon \leq \kappa_1 \text{ and } z_{E\#}\epsilon = \kappa_2 \right\}.$$

We define the set of countable sequences, S, of feasible adults' distributions of skills for generations following an adults' distribution of skills κ_1 by:

$$S(\kappa_1) = \left\{ \{\, \kappa_i \,\}_{i=1}^{\infty} \in \prod_{i=0}^{\infty} \mathcal{P}(K) \,\middle|\, \kappa_{i+1} \in \mathcal{N}(\kappa_i) \text{ for all } i = 1, 2, \ldots \right\}. \tag{3.13}$$

For consecutive adults' skill distributions (κ_i, κ_{i+1}), the space of possible matching, M, is defined as

$$M(\alpha, \kappa_i, \kappa_{i+1}) = \left\{ (\epsilon, \lambda) \in \mathcal{P}(A \times K) \times \frac{N'}{N}\frac{N-1}{N'+1}\mathcal{P}(K \times K) \,\middle|\, {}_{\substack{\pi_{1\#}\epsilon = \alpha \\ \pi_{1\#}\lambda + \frac{1}{N'}\pi_{2\#}\lambda + \frac{1}{N}\pi_{2\#}\epsilon = \kappa_i \\ z_{E\#}\epsilon = \kappa_{i+1}}} \right\}. \tag{3.14}$$

We define the optimal production, P, for a sequence of feasible adults' distributions of skills to be:

$$\begin{aligned} &P : S(\kappa_1) \to \mathbf{R} \\ &\{\, \kappa_i \,\}_{i=1}^{\infty} \mapsto \sup_{\substack{\{\, \epsilon_i, \lambda_i \,\}_{i=1}^{\infty} \text{ s.t.} \\ (\epsilon_i, \lambda_i) \in M(\alpha, \kappa_i, \kappa_{i+1})}} \sum_{i=1}^{\infty} e^{-\beta i} \left(\int_{A \times K} c b_E \circ z_E d\epsilon_i + \int_{K \times K} b_L \circ z_L d\lambda_i \right). \end{aligned} \tag{3.15}$$

We restate (3.9) as the supremum of the optimal production for any finite sequence of feasible adults' distributions of skills.

Lemma 3.2.1. *For any $\kappa_1 \in \mathcal{P}(K)$ the complete information model* (3.9) *can be restated as follows:*

$$C(\kappa_1) = \sup_{\{\, \kappa_i \,\}_{i=1}^{\infty} \in S(\kappa_1)} P\left(\{\, \kappa_i \,\}_{i=1}^{\infty}\right),$$

where C is defined in (3.9) *and P is defined in* (3.15).

Proof. First, we prove the inequality:

$$C(\kappa_1) \leq \sup_{\{\, \kappa_i \,\}_{i=1}^{\infty} \in S(\kappa_1)} P\left(\{\, \kappa_i \,\}_{i=1}^{\infty}\right).$$

Let $\{\, (\epsilon_i^n, \lambda_i^n) \,\}_{i=1}^{\infty}$ be such that

$$C(\kappa_1) \leq \frac{1}{n} + \sum_{i=1}^{\infty} e^{-\beta i} \int_{A \times K} c b_E \circ z_E d\epsilon_i^n + \sum_{i=1}^{\infty} e^{-\beta i} \int_{K \times K} b_L \circ z_L d\lambda_i^n.$$

Let $\kappa_1^n = \kappa_1$ and $\kappa_i^n = z_{E\#}\epsilon_{i-1}^n$ for $i \in \{2, 3, \dots\}$. Then, $\{\kappa_i^n\}_{i=1}^\infty \in S(\kappa_1)$. Thus,

$$\sup_{\{\kappa_i\}_{i=1}^\infty \in S(\kappa_1)} P(\{\kappa_i\}_{i=1}^\infty) \geq \sum_{i=1}^\infty e^{-\beta i} \int_{A\times K} cb_E \circ z_E \epsilon_i^n + \sum_{i=1}^\infty e^{-\beta i} \int_{K\times K} b_L \circ z_L \lambda_i^n.$$

Taking the limit when $n \to \infty$, we get

$$\sup_{\{\kappa_i\}_{i=1}^\infty \in S(\kappa_1)} P(\{\kappa_i\}_{i=1}^\infty) \geq C(\kappa_1).$$

Now, we prove the other inequality. We fix κ_1 and let $\{\kappa_i^n\}_{i=1}^\infty \in S(\kappa_1)$ be an approximate optimizer for

$$\sup_{\{\kappa_i\}_{i=1}^\infty \in S(\kappa_1)} P(\{\kappa_i\}_{i=1}^\infty).$$

Then for all n there exists $\{\epsilon_i^{n,m}, \lambda_i^{n,m}\}_{i=0}^\infty$ such that $(\epsilon_i^{n,m}, \lambda_i^{n,m}) \in M(\alpha, \kappa_i, \kappa_{i+1})$ and $\{\epsilon_i^{n,m}, \lambda_i^{n,m}\}_{i=0}^\infty$ is an approximate optimizer for

$$\sup_{\substack{\{\epsilon_i,\lambda_i\}_{i=1}^\infty \text{ s.t.} \\ (\epsilon_i,\lambda_i)\in M(\alpha,\kappa_i^n,\kappa_{i+1}^n)}} \sum_{i=1}^\infty e^{-\beta i} \int_{A\times K} cb_E \circ z_E \epsilon_i + \sum_{i=1}^\infty e^{-\beta i} \int_{A\times K} b_L \circ z_L \lambda_i.$$

For all n, m, the set of measures $\{\epsilon_i^{n,m}, \lambda_i^{n,m}\}_{i=0}^\infty$ is feasible for $C(\kappa_1)$. Thus, taking the limit over n and m, we have

$$C(\kappa_1) \geq \sup_{\{\kappa_i\}_{i=1}^\infty \in S(\kappa_1)} P(\{\kappa_i\}_{i=1}^\infty).$$

□

Let

$$\mathcal{P}(\kappa_1, \kappa_2) = \sup_{(\epsilon,\lambda)\in M(\alpha,\kappa_1,\kappa_2)} e^{-\beta} \int_{K\times K} b_L \circ z_L \lambda + e^{-\beta} \int_{A\times K} cb_E \circ z_E \epsilon \tag{3.16}$$

be the optimal productivity of a generation whose adults' distribution of skills is κ_1, given that the distribution of skills for the next generation of adults will be κ_2 and the students' distribution of abilities is α. With this definition, we have:

$$P\big(\{\kappa_i\}_{i=1}^\infty\big) = \sum_{i=1}^\infty e^{-\beta(i-1)} \mathcal{P}(\kappa_i, \kappa_{i+1}).$$

We want to show that $C(\kappa_1)$ from (3.9) satisfies the following functional equation:

$$C(\kappa_1) = \sup_{\kappa_2 \in \mathcal{N}(\kappa_1)} \mathcal{P}(\kappa_1, \kappa_2) + e^{-\beta} C(\kappa_2). \tag{3.17}$$

That is, the full productivity of all generations, given that the distribution of skills for the first generation of adults is κ_1, is the supremum of the optimal productivity of the first generation plus the full productivity of all following generations, over all possible distributions of adults' skills for the second generation κ_2. This would imply that:

$$H^C(\kappa_1) \in \operatorname*{arg\,sup}_{\kappa_2 \in \mathcal{N}(\kappa_1)} \mathcal{P}(\kappa_1, \kappa_2) + e^{-\beta} C(\kappa_1)$$

and the analogous formula (3.23) for $i > 1$.

In order to show that C satisfies (3.17), we need the following lemma.

Lemma 3.2.2. *For all* $\{\kappa_i\}_{i=1}^\infty \in S(\kappa_1)$ *we have:*

$$P\big(\{\kappa_i\}_{i=1}^\infty\big) = \mathcal{P}(\kappa_1, \kappa_2) + e^{-\beta} P\big(\{\kappa_i\}_{i=2}^\infty\big),$$

where S, P *and* $\mathcal{P}$ *are defined by* (3.13), (3.15) *and* (3.16).

Proof. First, we simply use the definition of P:

$$P\big(\{\kappa_i\}_{i=1}^\infty\big) = \sup_{\substack{\{\epsilon_i, \lambda_i\}_{i=1}^\infty \text{ s.t.} \\ (\epsilon_i, \lambda_i) \in M(\alpha, \kappa_i, \kappa_{i+1})}} \left(\sum_{i=1}^\infty e^{-\beta i} \int_{A \times K} c b_E \circ z_E \epsilon_i + \sum_{i=1}^\infty e^{-\beta i} \int_{A \times K} b_L \circ z_L \lambda_i \right).$$

Now, we separate the first term of the series and then we use the definition of P and $\mathcal{P}$.

$$\begin{aligned} P\left(\{\kappa_i\}_{i=1}^\infty\right) = &\sup_{\substack{\{\epsilon_i, \lambda_i\}_{i=2}^\infty \text{ s.t.} \\ (\epsilon_i, \lambda_i) \in M(\alpha, \kappa_i, \kappa_{i+1})}} \left(\sum_{i=2}^\infty e^{-\beta i} \int_{A \times K} b_L \circ z_L \lambda_i + \sum_{i=2}^\infty e^{-\beta i} \int_{A \times K} c b_E \circ z_E \epsilon_i \right) \\ &+ \sup_{(\epsilon_1, \lambda_1) \in M(\alpha, \kappa_1, \kappa_2)} \left(e^{-\beta} \int_{A \times K} b_L \circ z_L \lambda_1 + e^{-\beta} \int_{A \times K} c b_E \circ z_E \epsilon_1 \right) \\ = &e^{-\beta} P\big(\{\kappa_i\}_{i=2}^\infty\big) + \mathcal{P}(\kappa_1, \kappa_2) \end{aligned}$$

□

Proposition 3.2.3. *For all* $\kappa_1 \in \mathcal{P}(K)$*,* $C(\kappa_1)$ *defined by* (3.9) *satisfies* (3.17).

Proof. The proof follows the proof of [36, Theorem 4.2]. By Lemma 3.2.1,

$$C(\kappa_1) = \sup_{\{\kappa_i\}_{i=1}^\infty \in S(\kappa_1)} P\Big(\{\kappa_i\}_{i=1}^\infty\Big),$$

which can be restated with the following two properties:

$$\forall \{\kappa_i\}_{i=1}^\infty \in S(\kappa_1), \qquad C(\kappa_1) \geq P\left(\{\kappa_i\}_{i=1}^\infty\right); \tag{3.18}$$

$$\forall \epsilon > 0,\ \exists \{\bar{\kappa}_i\}_{i=1}^\infty \in S(\kappa_1) \text{ s.t.} \qquad C(\kappa_1) \leq P\left(\{\bar{\kappa}_i\}_{i=1}^\infty\right) + \epsilon. \tag{3.19}$$

To prove that $C(\kappa_2)$ satisfies (3.17), we'll show the following two properties:

$$\forall \kappa_2 \in \mathcal{N}(\kappa_1) \qquad C(\kappa_1) \geq e^{-\beta} C(\kappa_2) + \mathcal{P}(\kappa_1, \kappa_2); \tag{3.20}$$

$$\forall \epsilon > 0,\ \exists \kappa_2 \in \mathcal{N}(\kappa_1) \text{ s.t.} \qquad C(\kappa_1) \leq e^{-\beta} C(\kappa_2) + \mathcal{P}(\kappa_1, \kappa_2) + \epsilon. \tag{3.21}$$

First, we show that C satisfies (3.20). Fix $\kappa_1 \in \mathcal{P}(K)$ and $\epsilon > 0$. Let $\kappa_2 \in \mathcal{N}(\kappa_1)$, by (3.19), there exists $\{\kappa_i\}_{i=2}^{\infty} \in S(\kappa_2)$ such that

$$P\big(\{\kappa_i\}_{i=2}^{\infty}\big) \geq C(\kappa_2) - \epsilon. \tag{3.22}$$

Thus, we have

$$\begin{aligned} C(\kappa_1) &\geq P\big(\kappa_1, \{\kappa_i\}_{i=2}^{\infty}\big) && \text{by (3.18)} \\ &= e^{-\beta} P\big(\{\kappa_i\}_{i=2}^{\infty}\big) + \mathcal{P}(\kappa_1, \kappa_2) && \text{by Lemma 3.2.2} \\ &\geq e^{-\beta} C(\kappa_2) - e^{-\beta}\epsilon + \mathcal{P}(\kappa_1, \kappa_2) && \text{by (3.22).} \end{aligned}$$

Since $\epsilon > 0$ and $\kappa_2 \in \mathcal{N}(\kappa_1)$ were arbitrary, this shows (3.20).

Now we prove (3.21). Fix $\kappa_1 \in \mathcal{P}(K)$ and $\epsilon > 0$. From (3.19), there exists $\{\kappa_i\}_{i=1}^{\infty}$ such that

$$\begin{aligned} C(\kappa_1) &\leq P\big(\{\kappa_i\}_{i=1}^{\infty}\big) + \epsilon \\ &= \mathcal{P}(\kappa_1, \kappa_2) + e^{-\beta} P\big(\{\kappa_i\}_{i=2}^{\infty}\big) + \epsilon && \text{by Lemma 3.2.2} \\ &\leq \mathcal{P}(\kappa_1, \kappa_2) + e^{-\beta} C\big(\{\kappa_i\}_{i=2}^{\infty}\big) + \epsilon && \text{by (3.18).} \end{aligned}$$

Letting $\epsilon \to 0$ yields the claim. □

Corollary 3.2.4. *The function H^C defined by* (3.11) *satisfies:*

$$H^C(\kappa_i) \in \underset{\kappa_{i+1} \in \mathcal{N}(\kappa_i)}{\arg\sup}\, \mathcal{P}(\kappa_i, \kappa_{i+1}) + e^{-\beta} C(\kappa_{i+1}) \tag{3.23}$$

for all i, where $\mathcal{N}$ is defined in (3.12)*, $\mathcal{P}$ is defined in* (3.16) *and C is defined in* (3.9)*.*

In Appendix B, we introduce a complete information model with finite time horizon. We prove in Section B.1 that in this case, strong duality holds and in Section B.2 that our infinite time horizon model can be expressed as a limit of the finite one.

3.3 Analysis of the Models and Preliminary Results

Before we study dynamics of G^I and of H^C, we need various results about the optimization problems (3.1), (3.4), (3.9) and (3.10).

Given a wage function for adults v, a benefit from education function for students u and an expected future wage function for students $\tilde{v}$, we can define the wage functions for different roles of adults:

$$\begin{aligned} &\mathcal{V} : (u, v, \tilde{v}) \mapsto (v_w, v_m, v_t) \\ &\text{where } v_w(k) = \sup_{m \in \bar{K}} b_L(z_L(k,m)) - \frac{v(m)}{N'}, \\ &\qquad v_m(k) = N' \sup_{w \in \bar{K}} b_L(z_L(w,k)) - v(w) \qquad \text{and} \\ &\qquad v_t(k) = N \sup_{s \in \bar{A}} c b_E(z_E(s,k)) + \tilde{v}(z_E(s,k)) - u(s). \end{aligned} \tag{3.24}$$

The subscripts indicate the roles; w for worker, m for manager and t for teacher. Although we generally deal with bounded convex payoffs $u, v \geq 0$ which extend continuously to $\bar{K}$, should we wish to allow unbounded payoffs we augment the definition of v_t with the convention $\infty - \infty = \infty$ as in [13].

Note that in the complete information case, we can find wage functions for different roles at every generation with $\mathcal{V}(u_i, v_i, e^{-\beta} v_{i+1})$. The wage functions for the steady state model (2.5) can be recovered from $\mathcal{V}(u, v, v)$.

Because we assumed that b_E and b_L are increasing and strictly convex, (and we will only consider $\tilde{v}$'s that are non-decreasing and convex) we can see, from the structure of $\mathcal{V}$, that v will be convex non-decreasing and strictly convex increasing as long as $\tilde{v}$ is or $c > 0$. To make this idea more precise we need Lemma 2.1.6 which allows us to quantify the derivative of the wage function for a fixed type of adult.

The next proposition gives conditions that ensure that the wage functions v of (3.4) and v_i of (3.10) are convex and non-decreasing and are equal to the maximum of the wage over the various professions.

Proposition 3.3.1. *Fix $c \geq 0$, $\theta, \theta' \in (0,1)$ and $N, N' > 1$. Let $A = [0, \bar{k})$ and let α be a Borel probability measure on $\bar{A}$ satisfying the doubling condition (3.7) at $\bar{a}$. Define $z_{E/L}$ and $b_{E/L}$ as in subsection 2.5.*

If $\tilde{v} \leq b$ is increasing and strictly convex, then the infimum (3.4) is attained by convex, non-decreasing functions (u, v) satisfying

$$u(s) = \sup_{k \in \bar{K}} c b_E(z_E(s,k)) + \tilde{v}(z_E(s,k)) - v(k) \tag{3.25}$$

and

$$v = \max\{ v_w, v_m, v_t \},$$

where $(v_w, v_m, v_t) = \mathcal{V}(u, v, \tilde{v})$ is defined by (3.24).

If $c > 0$, the infimum (3.10) is attained by convex, non-decreasing functions $(u_i, v_i)_{i=1}^{\infty}$ satisfying

$$u_i(s) = \sup_{k \in \bar{K}} c b_E(z_E(s,k)) + v_{i+1}(z_E(s,k)) - v_i(k)$$

and

$$v_i = \max\{ v_w^i, v_m^i, v_t^i \},$$

where $(v_w^i, v_m^i, v_t^i) = \mathcal{V}\left(u_i, v_i, e^{-\beta} v_{i+1}\right)$.

Note 3.3.2. *This proposition relies heavily on the hypotheses we made in subsection 2.5 that b_E and b_L are both strictly positive, increasing and strictly convex.*

Note 3.3.3. *It might be surprising to the reader who knows the work of [13] to see that we don't need to introduce a perturbed problem in order to show that dual functions (u, v) are non-decreasing and convex. This is because the wage function for teachers, v_t, is not recursive in our cases. For the same reason, we don't need to contemplate the possibility of unbounded wages or wage gradients in our models, except perhaps in the long time limit. Therefore, we don't need to assume a priori that v is non-decreasing and convex in order to conclude that v_t is also non-decreasing and convex. For the infimum of (3.4), it suffices to assume that $\tilde{v}$ is non-decreasing and convex. For the infimum of (3.10), we approximate the problem by a finite version and assume the wage function of the last generation is non-decreasing and convex.*

We still need a perturbation of our problems in order to show that $v = \max\{ v_w, v_m, v_t \}$. We won't give a full formulation of such a perturbed problem here and refer to [13] for details.

Proof of Proposition 3.3.1. Let (u, v) be feasible for (3.4). Let $\bar{v} = \max\{ v_w, v_m, v_t \}$, where

$$(v_w, v_m, v_t) = \mathcal{V}(u, v, \tilde{v}).$$

First, we prove that $\bar{v}$ is continuous on K, and convex and increasing on $\bar{K}$ following the proof of [13, Lemma 5]. To do so, we'll prove that $v_{w/m/t}$ are all convex increasing.

We start with v_t. First, we'll assume $\tilde{v}$ and b_E are twice differentiable. Define

$$f(a, k) = cb_E(z_E(a, k)) + \tilde{v}(z_E(a, k)).$$

As by assumption, b_E and $\tilde{v}$ are convex increasing and z_E is linear, we have that f is convex increasing in both variables. By Lemma 2.1.6, $v_t(k) = N \sup_{a\in A} f(a, k) - u(a)$ is also convex increasing and satisfies

$$\begin{aligned} v_t' &\geq\ N\theta \left(c\underline{b}_E' + \inf_b \tilde{v}'(b)\right)\ \geq 0 \\ v_t'' &\geq N\theta^2 \left(c\underline{b}_E'' + \inf_b \tilde{v}''(b)\right) \geq 0. \end{aligned}$$

If $\tilde{v}, b_E \notin C^2(\bar{K})$, we can approximate it uniformly on any compact subset of K by C^2 functions $\tilde{v}^j$. Using the same reasoning, we'll get that v_t^j is convex increasing. Thus, taking the limit, v_t is convex non-decreasing.

Now, we consider $v_{w/m}$. We apply Lemma 2.1.6 with $f(w, m) = b_L(z_L(w, m))$ which is jointly convex and increasing in both variables. If b_L isn't C^2, we can approximate it by functions that are. We get the following bounds:

$$(1 - \theta')\underline{b}_L' \leq f_w(w, m) \leq (1 - \theta')\overline{b}_L'$$

and

$$f_{ww}(w, m) \geq (1 - \theta')^2\underline{b}_L''.$$

Those bounds and Lemma 2.1.6 imply that v_w is also convex non-decreasing. Similarly, we can show that v_m is convex non-decreasing.

Thus, we conclude that $\bar{v}$ is convex non-decreasing and satisfies:

$$\begin{aligned} \bar{v}' &\geq \min\left\{(1 - \theta')\underline{b}_L', N'\theta'\underline{b}_L', N\theta\left(c\underline{b}_E' + \inf_b \tilde{v}'(b)\right)\right\} \\ \bar{v}'' &\geq \min\left\{(1 - \theta')^2\underline{b}_L'', N'\theta'^2\underline{b}_L'', N\theta^2\left(c\underline{b}_E'' + \inf_b \tilde{v}''(b)\right)\right\}. \end{aligned}$$

Because $\bar{v}$ is convex non-decreasing on $\bar{K}$, it is continuous on K.

Now, we show, also following the proof of [13, Lemma 5], that u satisfy (3.25) and is continuous on A, and convex and increasing on $\bar{A}$. As we assumed that u is an optimizer, it has to satisfy (3.25), otherwise we can lower the objective of (3.4) by replacing u by the right side of (3.25). Using Lemma 2.1.6, we can get a lower bound on u' and u'' from the fact that b_E and $\tilde{v}$ are convex and increasing and we conclude that u is convex and increasing.

Now, we'll prove that there exists optimizers for (3.4). To do so, we'll first add the extra hypothesis that feasible v are convex and non-decreasing and in that case use [13, Lemma 11] to have an optimizer.

We will then show that these extra hypotheses on v are non-binding. First, note that feasibility of $(u, v) = (b + c\bar{b}_E, \bar{b}_L)$, where b is an upper bound for $\tilde{v}$, yields an upper bound:

$$b + c\bar{b}_E + \bar{b}_L$$

for (3.4). Then [13, Lemma 11] allows us to extract from any approximating sequence of optimizers an optimizing pair (u, v) for (3.4) satisfying (3.5), (3.6) and the extra condition that garantees that v is non-decreasing and convex. Fatou's lemma ensures that (u, v) minimizes the objective.

Now we prove that $v = \max\{v_w, v_m, v_t\}$ by following [13, Theorem 13]. First, we suppose that κ is supported on K. Let $\bar{v} = \max\{v_w, v_m, v_t\}$. Feasibility of (u, v) implies $v \geq \bar{v}$. If $\eta = v - \bar{v}$ is positive somewhere, it is positive on an interval where no constraints can be binding. For small $\lambda > 0$, we define the perturbation $v^\lambda := (1 - \lambda)v + \lambda\bar{v}$. We will show that the pair (u, v^λ) is feasible. Thus, unless $\eta = 0$ on all K, (u, v^λ) lowers the objective functional. This is a contradiction, which implies that $v = \bar{v}$.

If spt $\kappa \subsetneqq K$, we can introduce a perturbed problem with spt $\kappa_\delta = K$ and $\lim_{\delta\to 0}\kappa_\delta = \kappa$ in order to show that $v = \bar{v}$.

Now, we show (u, v^λ) is feasible. First, we show that v^λ satisfies the stability of the labour market's constraint (3.6). Since $v^\lambda = v - \lambda\eta = \bar{v} + (1 - \lambda)\eta$, for any $w, m \in \bar{K}$, we have:

$$\begin{aligned} v^\lambda(w) + \frac{v^\lambda(m)}{N'} - b_L(z_L(w, m)) &= \bar{v}(w) + \frac{v(m)}{N'} - b_L(z_L(w, m)) + (1 - \lambda)\eta(w) - \frac{\lambda}{N'}\eta(m) \\ &\geq \eta(w)\left(1 - \lambda\left(1 + \frac{\eta(m)}{N'\eta(w)}\right)\right), \end{aligned}$$

and

$$\begin{aligned} v^\lambda(w) + \frac{v^\lambda(m)}{N'} - b_L(z_L(w, m)) &= v(w) + \frac{\bar{v}(m)}{N'} - b_L(z_L(w, m)) - \lambda\eta(w) + \frac{1 - \lambda}{N'}\eta(m) \\ &\geq \frac{\eta(m)}{N'}\left(1 - \lambda\left(1 + \frac{N'\eta(w)}{\eta(m)}\right)\right). \end{aligned}$$

If $\eta(m) = \eta(w) = 0$, then (3.6) is satisfied. If both $\eta(m), \eta(w) \geq 0$, taking $\lambda < 1/2$ ensures that (3.6) is satisfied. The same is true if one of $\eta(m)$ or $\eta(w)$ vanish.

Now, we prove that (u, v^λ) satisfies the stability of education market constraint (3.5). Because $\bar{v} \geq v$, by adding $u(s) - cb_E(z_E(s, t))$ to

$$\frac{v^\lambda(t)}{N} - v^\lambda(z_E(s, t)) = \frac{\bar{v}(t)}{N} - v(z_E(s, t)) + \frac{1 - \lambda}{N}\eta(t) + \lambda\eta(z_E(s, t)),$$

and considering constraint (3.5), we get

$$u(s) + \frac{v^\lambda(t)}{N} - cb_E(z_E(s, t)) - v^\lambda(z_E(s, t)) \geq \frac{1 - \lambda}{N}\eta(t) + \lambda\eta(z_E(s, t)) \geq 0.$$

This shows that (u, v^λ) is feasible and therefore, that $\bar{v} = v$ when (u, v) is an optimizing pair of (3.4).

Finally, we proved that $\bar{v}$ is strictly increasing and convex so the extra assumption that v is non-decreasing and convex are non-binding.

For the complete information model (3.10), we can optain an optimizing sequence $\{(u_i, v_i)\}_{i=1}^{\infty}$ similarly. First, we add the extra assumption that for all i, v_i is non-decreasing and convex. Note that

$\{ (u_i, v_i) \}_{i=1}^{\infty} = \{ (\bar{b}_L + c\bar{b}_E, \bar{b}_L) \}$ is feasible because $1 + \frac{1}{N} \geq e^{-\beta}$. So

$$e^{-\beta} \frac{2 - e^{-\beta}}{1 - e^{-\beta}} \bar{b}_L + \frac{e^{-\beta}}{1 - e^{-\beta}} c\bar{b}_E$$

is an upper bound for (3.10). Thus, there exists a sequence of approximate optimizers $\{ \{ (u_i^j, v_i^j) \}_{i=1}^{\infty} \}_{j=1}^{\infty}$ for (3.10) with extra assumption that v_i are convex non-decreasing.

By [13, Lemma 11], we can extract a converging subsequence for $i = 1$. Let j_l be the indices of this subsequence. We then extract a converging subsequence of $(u_2^{j_l}, v_2^{j_l})$ for $i = 2$ and continue in this fashion to get subsequences j_k such that $(u_i^{j_k}, v_i^{j_k})$ converges for all $i \leq k$. We can then use a diagonal argument[1] to get a subsequence such that $(u_i^j, v_i^j) \to (u_i, v_i)$ converges for all i.

By Fatou's lemma, $\{ (u_i, v_i) \}_{i=1}^{\infty}$ minimizes (3.10).

The proof that the infimum of (3.10) is attained by strictly convex, increasing functions $(u_i, v_i)_{i=1}^{\infty}$ uses the finite horizon problem introduced in Appendix B. For the finite horizon problem (B.2), the proof is done by induction over a perturbed problem. Indeed, we need to include v_i in the objective functional of (3.10) for all i. To do so, we can use a perturbed problem as the one from [13]. Then if we assume for fixed i the wage function, v_i, is continuous and convex non-decreasing, we can use a proof that is identical to the proof in the incomplete information case to show that $v_{i-1} = \max \{ v_w^{i-1}, v_m^{i-1}, v_t^{i-1} \}$ is continuous on $\bar{K}$ and strictly convex increasing on K (strictly convex and increasing because $c > 0$). As we assume that v_{T+1} is continuous and convex non-decreasing, we conclude that all v_i are strictly convex and increasing.

As we prove in Corollary B.2.3 that optimizers of (B.2) converge to optimizers of (3.10), we conclude that optimizers of (3.10), v_i, are non-decreasing and convex. Coupling this with the fact that strict convexity and growth of v_i is bounded, we get that optimizers of (3.10), v_i, are in fact increasing and strictly convex. □

Corollary 3.3.4. *The wage functions v_i solving the incomplete* (3.4) *and complete information models* (3.10) *are continuous on $\bar{K}$ and K respectively, and they are differentiable almost everywhere on their domain.*

The next proposition presents conditions to ensure the existence, uniqueness and positive assortativity of primal optimizer.

Proposition 3.3.5. *Fix $c \geq 0$ and $\theta, \theta' \in (0, 1)$. Let $A = K = [0, \bar{k})$ and fix a Borel probability measure $\alpha \geq 0$ and a Borel measure $\kappa_1 \geq 0$ both without atoms on $\bar{A}$ such that α satisfies the doubling condition* (3.7)*. Let $\tilde{v}$ be a positive, increasing and strictly convex function. Define*

$$z_E(a, k) = (1 - \theta)a + \theta k.$$

There exists maximizers (ϵ, λ) (resp. $(\epsilon_i, \lambda_i)_{i=1}^{\infty}$) of (3.1) *(resp. of* (3.9)*).*

Then, for any maximizing ϵ, λ of (3.4) *and $(\epsilon_i, \lambda_i)_{i=1}^{\infty}$ of* (3.9)*: λ and λ_i are positive assortative (see definition 2.1.4). Moreover, there exists a pair of optimizers for both problems such that ϵ and ϵ_i are also positive assortative. For those ϵ and ϵ_i, there exist non-decreasing functions $\phi, \phi_i : A \to K$ uniquely*

[1] A similar diagonal argument is used in the proof of Lemma 2.2.1 (c).

determined α-a.e. such that

$$\epsilon = (\mathbb{1} \times \phi)_{\#}\, \alpha,$$
$$\epsilon_i = (\mathbb{1} \times \phi_i)_{\#}\, \alpha \text{ for all } i = 1, \dots.$$

The measures λ and λ_i are unique. If $c > 0$ then ϵ and ϵ_i are also unique; if $c = 0$ this uniqueness extends to the incomplete information model provided $\tilde{v}$ is strictly convex.

Proof. To prove existence of primal optimizers for (3.1) (resp. (3.9)) we follow the proof of [13, Lemma 17]. Given the shape of the constraints, feasible pairs for (3.1) (resp. (3.9)) form a weak-$*$ compact subset of $C(\bar{A}^2)^*$, (resp. of $\left(C(\bar{A}^2)_{i=1}^{\infty}\right)^*$). As the linear functionals we are trying to optimize are weak-$*$ continuous, the maximum must be attained provided the set of feasible measures is non-empty.

To see that the feasible set of (3.1) is non-empty, let k_t be the largest element of K such that $\int_{k_t}^{\bar{k}} \kappa_1 = \frac{1}{N}$, and let $\epsilon = \alpha \times N\kappa_1|_{(k_t,\bar{k})}$. Then we let k_m be the largest element of K such that $\int_{k_m}^{k_t} \kappa_1 = \frac{N-1}{N(N'+1)}$ and let $\lambda = \kappa_1|_{(0,k_m)} \times N'\kappa_1|_{(k_m,k_t)}$.

To see that the feasible set of (3.9) is non-empty, we use the same construction to get ϵ_1 and λ_1. We then get ϵ_i and λ_i for $i > 1$ by setting $\kappa_i = z_{E\#}\epsilon_{i-1}$.

Strict positive assortativity of λ, λ_i, and positive assortativity of ϵ, ϵ_i is proven as in [13, Theorem 15]. We will present the proof in the incomplete information setting. Let

$$f(s,t) = u(s) + \frac{v(t)}{N} - cb_E(z_E(s,t)) - \tilde{v}(z_E(s,t)) \geq 0$$

and

$$g(w,m) = v(w) + \frac{v(m)}{N'} - b_L(z_L(w,m)) \geq 0.$$

As the problem satisfies competitive equilibrium[2], we have that $\epsilon(f) = \lambda(g) = 0$. Therefore, the support of ϵ is contained in the zero set $F \subset A \times K$ of f and the support of λ is contained in the zero set $G \subset K^2$ of g.

Now, as b_L and $\tilde{v}$ or cb_E are strictly convex, f and g are strictly submodular. Thus, their zero sets F and G are non-decreasing.

We will show that the measures ϵ and ϵ_i's are unique because f is strictly submodular, and that the measures λ and λ_i's are unique because g is strictly submodular. First, we prove that ϵ is unique. As f is strictly submodular, except potentially for a countable number of vertical segments, the non-decreasing set G is contained in the graph of a non-decreasing map $\varphi : A \to K$. Any measure $\epsilon \in \mathcal{P}(A \times K)$ such that $\pi_{1\#}\epsilon = \alpha$ has no mass on these segments because α doesn't have any atoms. Since the maximizer ϵ vanished outside the graph of φ, it is uniquely defined by $\epsilon = (\mathbb{1} \times \varphi)_{\#}\alpha$.

To show that λ is unique, we assume there exist two optimal matchings for the labour sector: λ_1 and λ_2. One can show that as α doesn't have any atoms, neither does κ_i or the measures of workers and managers: $\pi_{1\#}\lambda_i + \frac{\pi_{2\#}\lambda_i}{N'}$ for $i = 1, 2$[3]. Let $\Delta\lambda = \lambda_2 - \lambda_1$. We know that λ_1, λ_2 and therefore $\Delta\lambda$ must vanish outside G. Once again, G has at most countably many vertical and horizontal segments, but λ_i doesn't charge them, because κ_i doesn't have atoms. The positive part of the marginals of $\Delta\lambda$,

[2] I.e. there is no duality gap. For the proof that there is no duality gap between (3.1) and (3.4) see [13] and for the proof that there is no duality gap between (3.9) and (3.10) see Section B.1.

[3] See [13, Lemma 14].

$\pi_{1\#}\Delta\lambda_+$ and $\pi_{2\#}\Delta\lambda_+$ must have the same mass. Also, by feasibility,

$$N'\pi_{1\#}\Delta\lambda_+ - N'\pi_{1\#}\Delta\lambda_- + \pi_{2\#}\Delta\lambda_+ - \pi_{2\#}\Delta\lambda_- = 0.$$

As the positive and the negative parts of a measure have disjoint support, it means

$$N'\pi_{1\#}\Delta\lambda_+ = \pi_{2\#}\Delta\lambda_- \left(\text{ and } N'\pi_{1\#}\Delta\lambda_- = \pi_{2\#}\Delta\lambda_+\right).$$

Since $\pi_{1\#}\Delta\lambda_+$ and $\pi_{2\#}\Delta\lambda_-$ have the same mass, we have a contradiction, as $N' > 1$. □

The following lemma gives conditions that ensure that the distribution for skills of adults is without atoms.

Lemma 3.3.6. *Let $A = K = [0, \bar{k})$ and let α be a Borel probability measure without atoms on $\bar{A}$. Define $z_E(a, k) = (1-\theta)a + \theta k$. Let ϵ be a positive assortative measure on $A \times K$ such that $\pi_{1\#}\epsilon = \alpha$. Then $\kappa = z_{E\#}\epsilon$ has no atoms.*

If $\alpha \in L^\infty$ then $\kappa \in L^\infty$ also. More explicitely, when $X'_\alpha > \epsilon_c$ we have $X'_\kappa > \epsilon_c(1-\theta)$, where X_α (resp. X_κ) is the inverse of the cumulative distribution of α (resp. κ).[4]

Proof. The proof of the first claim follows the proof of [13, Lemma 14]. As ϵ is positive assortative, its support is non-decreasing, which means that almost everywhere, spt $(\epsilon) \subset \text{graph}(k_t)$ for some non-decreasing map $k_t : A \to K$. From [1, Lemma 3.1], $\epsilon = (\mathbb{1} \times k_t)_\# \alpha$.

The function $f(a) = z_E(a, k_t(a))$ is non-decreasing and $\kappa = f_\# \alpha$. Moreover, the derivative of f exists almost everywhere and is such that $f' \geq \frac{1}{1-\theta}$. Therefore f is one-to-one and there exists a function $g : K \to A$ such that $g' \leq 1-\theta$ and $g(\bar{f}(k)) = k$ for every non-decreasing extension $\bar{f}$ of f.

For all $K' \subset K$, we have

$$\kappa(K') = \alpha\big(f^{-1}(K')\big) = \alpha\big(g(K')\big).$$

As this function is valid when K' is a single point and we assumed that α has no atoms, we conclude that κ has no atoms.

We express the probability measures α and κ in terms of non-decreasing random variables:

$$\begin{aligned} X_\alpha : [0,1] &\to \bar{A} \\ r &\mapsto \text{ability of a student in the } r\text{-th percentile} \\ X_\kappa : [0,1] &\to \bar{K} \\ r &\mapsto \text{ability of an adult in the } r\text{-th percentile.} \end{aligned}$$

When $\alpha \in L^\infty$, then there exists an $\epsilon_c > 0$ such that ${X_\alpha}' > \epsilon_c$. The inverse function of the random variable X_α (resp. X_κ) is the cumulative distribution of α (resp. κ): F_α (resp. F_κ). By the inverse

[4] Recall that $\alpha \in L^\infty$ is equivalent to existence of an $\epsilon_c = 1/\|\alpha\|_\infty > 0$ such that $X'_\alpha > \epsilon_c$.

function theorem, we have that ${F_\alpha}' < \frac{1}{\epsilon_c}$. By definition of a random variable, we have

$$\begin{aligned} F_\kappa(k) &= \kappa([0,k]) \\ &= \alpha([g(0), g(k)]) \\ &= F_\alpha(g(k)). \end{aligned}$$

By taking the derivative with respect to k, we get:

$$\begin{aligned} F_\kappa'(k) &= g'(k) F_\alpha'(g(k)) \\ &< \frac{1}{\epsilon_c(1-\theta)}. \end{aligned}$$

So by applying the inverse function theorem again, we get that ${X_\kappa}'(k) > \epsilon_c(1-\theta)$. □

The next theorem gives conditions for the skills of teachers to be weakly above the skills of managers which will themselves be weakly above the skills of workers. This is an example where the rank function that sends the rank of students to the rank of their teachers R is the same for any generation.

Theorem 3.3.7. *Fix $c \geq 0$, $\theta, \theta' \in (0,1)$ and $N, N' > 1$. Let $A = [0, \bar{k})$ and let α be a Borel probability measure on $\bar{A}$ satisfying the doubling condition* (3.7)*. Define $z_E(a,k) = (1-\theta)a + \theta k$ and $b_{E/L}$ satisfying the bounds* (2.4)*. Assume*

$$N\theta c \underline{b}_E' \geq \overline{b}_L' \max\{ N'\theta', 1-\theta' \} \tag{a}$$

and

$$N'\theta' > (1-\theta') \sup_{k \in K} \frac{b_L'\left((1-\theta')k + \theta' \overline{k}^-\right)}{b_L'(\theta' k^+)}.^5 \tag{b}$$

If $\tilde{v}$ is non-decreasing and striclty convex, then the optimal measures (ϵ, λ) of (3.1) *are completely determined by the fact that the skills of workers are below the skills of managers which themselves are below teachers' skills and that ϵ and λ are positive assortative.*

The optimal measures $\{\epsilon_i, \lambda_i\}_{i=1}^\infty$ of (3.9) *are completely determined by the fact that the skills of workers are below the skills of managers which themselves are below teachers' skills and that ϵ_i and λ_i are positive assortative.*

Note 3.3.8. *Hypothesis* (a) *represents the fact that the benefit on productivity of having highly skilled teachers is large relative to the benefit of having highly skilled individuals in the labour market. Indeed, it requires that the number of students a teacher can teach (N), the impact of a teacher on its student's future skill (θ), and the benefit from education (c) are large. Hypothesis* (b) *represents the fact that the number N' of workers a manager can supervise is large and managers' skills have a large impact on the production (high θ').*

Proof of Theorem 3.3.7. We suppose (a) holds, and prove that for optimizers of (3.1) and of (3.9) the skills of teachers are weakly above the skills of other adults. This proof follows the proof of [13,

[5] We tag these equation (a) and (b) to be consistant with [13].

Proposition 7 (a)]. First, by Lemma 2.1.6, for all k where $v_{w/m}$ is differentiable there exists a $k_{m/w}$ such that

$$v'_w(k) = (1-\theta')b'_L(z_L(k,k_m)) \leq (1-\theta')\bar{b}'_L \tag{3.26}$$

$$v'_m(k) = N'\theta' b'_L(z_L(k_w,k)) \leq N'\theta'\bar{b}'_L \tag{3.27}$$

If a teacher of cognitive skill k_0 teaches a student of ability a, that student will end up having cognitive skill $k_1 = z_E(a, k_0)$. Using the fact that

$$v_t(k) = N \sup_{s\in A} cb_E(z_E(s,k)) + \tilde{v}(Z_E(s,k) - u(s)$$

and Lemma 2.1.6, we have that

$$v'_t(k_0) = N\theta\left(cb'_E(k_1) + \tilde{v}'(k_1)\right)$$

whenever v_t is differentiable. The fact that v_t is differentiable at k_0 follows from convexity. Indeed, we have that

$$u(a) + \frac{1}{N}v_t(k) - \tilde{v}(z_E(a)) - cb_E(z_E(a,k)) \geq 0$$

is convex and is equal to zero if $k = k_0$. Thus, k_0 is a minimizer. The first order necessary condition states that:

$$(\tilde{v}' + cb'_E)(z_E(a,k_0)^-) \geq \frac{1}{N}\frac{v'_t(k_0)}{\frac{\partial z_E(a,k_0)}{\partial k}} \geq (\tilde{v}' + cb'_E)(z(a,k_0)^+).$$

As $\tilde{v} + cb_E$ is differentiable by hypothesis,

$$v'_t(k_0) = N\frac{\partial z_E(a,k_0)}{\partial k}(\tilde{v}' + cb'_E)(z_E(a,k_0))$$

is well defined.

Thus, $v'_t(k_0) \geq N\theta c\underline{b}'_E$. Under the hypothesis of (a), it means that

$$v'_t \geq v'_w \qquad\qquad \text{and}$$
$$v'_t \geq v'_m.$$

Thus the functions

$$v_t - v_{w/m}$$

must be increasing. As the functions $v_t - v_{w/m}$ are non-positive on

$$\{\, k \mid \bar{v}(k) = v_{w/m}(k) \,\}$$

and non-negative on

$$\{\, k \mid \bar{v}(k) = v_t(k) \,\}$$

the first two sets must lie entirely to the left of the third, and so teachers' abilities are weakly above the abilities of other adults.

Now, we show that if (b) holds, then the skills of workers are weakly below the skills of managers. The proof follows the proof of [13, Proposition 7 (b)]. First, take the maximal case in (3.26), i.e. $m = \bar{k}$, and the minimal case if (3.27), i.e. $w = 0$. We get that

$$v'_w(k) \leq (1-\theta')b'_L(z_L(k,\bar{k}))$$
$$v'_m(k) \geq N'\theta' b'_L(\theta' k).$$

Thus if (b) holds, $v'_w < v'_m$. This implies that the function $v_m - v_w$ is strictly increasing. As the function $v_m - v_w$ 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 on the left of the second. □

We will also need some results from [13, Proposition 7] that we state now.

Proposition 3.3.9 (Specialization by type). *Fix $K = A = [0, \bar{k})$ and $c \geq 0$. Suppose $u, v, \tilde{v} : K \to \mathbf{R}$ are convex, non-decreasing and satisfy $v = \mathcal{V}(u, v, \tilde{v})$.*

If

$$N'\theta' > (1-\theta') \sup_{k \in K} b'_L(z_L(k, \bar{k}^-))/b'_L(\theta' k^+) \tag{b}$$

worker types lie weakly below all of the manager types.

If

$$N\theta \geq 1 \tag{d}$$

any student will be weakly less skilled than his or her teacher.

If

$$\text{either } c > 0 \text{ or } N\theta > 1 \tag{e}$$

and (d) *holds then any student will be strictly less skilled than his or her teacher.*

If

$$\text{either } c > 0 \text{ or } v'(0^+) > 0, \tag{f}$$

(d) *and* (e) *hold then 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.*

Finally, we need some duality theorem in order to prove that the assymptotic of our dynamic problem is indeed equivalent to (2.5) the steady state problem from [13]. The duality theorems for the incomplete information model (3.1)-(3.3) and (3.4)-(3.6) are proved exactly as the ones from [13] therefore we state them here without proof. The duality theorems for the complete information model (3.9) and (3.10) requires the introduction of a finite horizon model and is done in Section B.1.

Theorem 3.3.10 (c.f. [13, Corollary 9]). *Fix c, θ, θ', N, N', $\tilde{v}$, κ a distribution for student skills $\alpha \in \mathcal{P}(\bar{A})$. A pair of feasible measures $\epsilon, \lambda \geq 0$ maximizes* (3.1) *if there exists a pair of functions* (u, v)

that satisfy (3.5) *and* (3.6) *such that*

$$\alpha(u) + \kappa(v) = \epsilon\left(cb_E \circ z_E + \tilde{v} \circ z_E\right) + \lambda\left(b_L \circ z_L\right).$$

3.4 Convergence of the Adult Skill Distributions

In this section, we study the sequence of measures for skills of adults $\{\kappa_i\}$. We will prove that the sequence of measures converges to a fixed point from any initial measure in $\mathcal{P}(K)$ under some conditions. We'll assume that the optimal measures (ϵ_i, λ_i) for the complete information problem (3.9) and (ϵ, λ) for the incomplete information model (3.1) satisfy the two following properties:

(A) positive assortativity they are positive assortative;

(B) specialization for each generation, the profession of almost every adult is uniquely determined by his or her skill's rank among his or her peers.

To motivate these assumptions and give the reader examples of the simpler cases of (B) specialization, we'll present two examples of potential societies, where the ordering is the same for any generation. We proved in Theorem 3.3.7 that if (a) and (b) hold, the optimal measures (ϵ_i, λ_i) for (3.9) and (ϵ, λ) for (3.1) satisfy (A) positive assortativity and the skills of teachers are weakly above the skills of managers, which are weakly above the skills of workers, which is a specific case of (B) specialization.

Another important special case of (B) specialization is when the skills of teachers of teachers and of teachers of managers are weakly above the skills of managers, which are weakly above the skills of teachers of workers, which are weakly above the skills of workers. This is the case in the optimal solution to the simulation of [27, Section 4].

Note that for these two special cases, the adults with highest skill are teachers, but this is not a consequence of hypotheses (A) positive assortativity and (B) specialization. Consequently, we don't assume that the adults with highest skill are teachers.

If the solution satisfies (A) positive assortativity and (B) specialization, the pair of measures (ϵ_i, λ_i) only depends on the distribution for students' skills α and on the initial distribution for adults' skills κ_1. In this case, $\epsilon_i = (\mathbb{1} \times \varphi_i)_{\#}\alpha$ for $\varphi_i : A \to K$, non-decreasing sending the ability of a student to the skill of his or her teacher. The skills for the next generation of adults for both (3.9) and (3.1) will always be given by $\kappa_{i+1} = G(\kappa_i) = z_{E\#}\epsilon_i$. That is, $H^C = H^I = G$, where $H^I(\cdot) = H^I(\cdot, \tilde{v})$.

For all i, as α has no atoms and ϵ_i is strictly positive assortative, there exists a strictly increasing function

$$R_i : [0,1] \to [0,1]$$

that sends the rank of a student in the a-th percentile at generation i to the rank $R_i(a)$ of his or her teacher among adults in the teacher's generation. In order to understand the dynamics of the distribution of adults skills, we need to study properties of rank functions R_i's.

Lemma 3.4.1. *The rank function R_i, which is associated to a matching of students with teachers that satisfies (A) positive assortativity and (B) specialization, satisfies $R_i' = \frac{1}{N}$ a.e. on $[0,1]$ and $x \in [0,1] \mapsto R_i(x) - x/N$ is non-decreasing.*

Proof. Fix i. Consider the cumulative distribution F_α of α (resp. F_{κ_i} of κ_i). By (A) positive assortativity, the function $\varphi_i : A \to K$ sending the ability of a student to the skill of his or her teacher

$\left(\epsilon_i = (\mathbb{1} \times \varphi_i)_\# \alpha\right)$ is increasing. We can write R_i as $R_i = F_{\kappa_i} \circ \varphi_i \circ F_\alpha^{-1}$. As R_i is the composition of non-decreasing functions, it is non-decreasing. Therefore, R_i is differentiable a.e., and its distributional derivative is non-negative. We now prove that $R_i' = \frac{1}{N}$ a.e. on $[0,1]$ by showing both inequalities. The rank function R_i satisfies $R_i' \geq \frac{1}{N}$ a.e. on $[0,1]$. Indeed, if $R_i' < \frac{1}{N}$ on an interval $(a,b) \in [0,1]$, we have that

$$|R_i(b) - R_i(a)| < \frac{1}{N}\,|b-a|$$

which means that there aren't enough teachers for the quantity of students with ability rank between a and b. Similarly, by (B) specialization $R_i' \leq \frac{1}{N}$ a.e. on $[0,1]$. So we conclude that $R_i' = \frac{1}{N}$ a.e. on $[0,1]$. Thus the distributional derivative of R_i dominates $1/N$, to establish the monotonicity of $x \in [0,1] \mapsto R_i(x) - x/N$. □

In order to prove convergence of the inverse of the cumulative distribution functions of the adults' skill distributions we need to study some compositions of the R_i's. Let $R^{(i,j)}$ be the composition of R_{i-j+1} with R_{i-j+2} composed with the next one, continuing in this way until we reach R_i:

$$R^{(i,j)} = R_{i-j+1} \circ R_{i-j+2} \circ R_{i-j+3} \circ \cdots \circ R_{i-1} \circ R_i.$$

Using base case $R^{i,0} = \mathbb{1}$, we can define the $R^{(i,j)}$ recursively as

$$R^{(i,j)} = R_{i-j+1} \circ R^{(i,j-1)}. \tag{3.28}$$

Let

$$R^{(\infty,j)} = \underbrace{R_\infty \circ R_\infty \circ \cdots \circ R_\infty}_{j \text{ times}}.$$

We need the following assumptions on the rank functions and their compositions:

(C) the composition estimate there exists a norm $\|\cdot\| \leq \|\cdot\|_\infty$ on $L^\infty([0,1])$ (that might depend on α, θ and $\{R_i\}$) for which there exists a non-decreasing function $R_\infty : [0,1] \to [0,1]$ such that

$$\lim_{i\to 0} \left\| \sum_{j=1}^{i-1} \theta^j \left(X_\alpha \circ R^{(\infty,j)} - X_\alpha \circ R^{(i,j)} \right) \right\| = 0.$$

We also assume $\|\cdot\|$-convergence implies subsequential convergence Lebesgue a.e.

We present two examples where the condition (C) the composition estimate is satisfied. First, if the rank function is independent of i, that is if the rank function is the same for every generation: $R_i := R_\infty$.

For the second example, we prove in the next lemma and proposition that if R_∞ is continuous ($R_\infty(x) = \frac{1}{N}x + constant$ by Lemma 3.4.1) then (C) the composition estimate is satisfied if there exists an R_∞ such that $R_i \to R_\infty$ in $L^\infty([0,1])$. We begin with the following:

Lemma 3.4.2 (Implications of contracting limit)**.** *Let* $\mathrm{Lip}(R_\infty) \leq 1$ *and* $x \in [0,1] \mapsto R_i(x) - x/N$ *be non-decreasing for all* i. *If* $\|R_i - R_\infty\|_{L^\infty([0,1])} \leq \epsilon$ *for all* $i \geq I_\epsilon$, *then* $i - j \geq I_\epsilon$ *implies*

$$\|R^{(\infty,j)} - R^{(i,j)}\|_{L^\infty([0,1])} \leq j\epsilon.$$

Proof. Fix $i > I_\epsilon$ and $i - j \geq I_\epsilon$. We'll prove the lemma by adding and subtracting $R^{(\infty,k)} \circ R^{(i,j-k)}$ for

$k = 1, \dots j-1$, thus re-expressing $R^{(\infty,j)} - R^{(i,j)}$ as a sum of j terms of the form

$$R^{(\infty,j-1-k)} \circ (R_\infty - R_{i-k}) \circ R^{(i,k)}.$$

By coupling the terms as above, we get

$$\begin{aligned}\left\| R^{(\infty,j)} - R^{(i,j)} \right\|_{L^\infty(\mathcal{L})} &\le \sum_{k=0}^{j-1} \mathrm{Lip}(R_\infty)^{j-1-k} \left\| R_\infty - R_{i-k} \right\|_{L^\infty(R_{\#}^{(i,k)}\mathcal{L})} \\ &\le \sum_{k=0}^{j-1} \left\| R_\infty - R_{i-k} \right\|_{L^\infty(R_{\#}^{(i,k)}\mathcal{L})} \\ &\le j\epsilon.\end{aligned}$$

Here we have used the fact that $x \in [0,1] \mapsto R_i(x) - x/N$ is non-decreasing to conclude $R_{\#}^{(i,k)}\mathcal{L}$ has a density bounded by N^k, hence is absolutely continuous with respect to Lebesgue. □

Proposition 3.4.3 (L^∞ composition estimate)**.** *Fix a distribution α for students' abilities with $\log \alpha \in L^\infty(\bar{A})$, so that the inverse X_α of its cumulative distribution function is Lipschitz. Under the hypotheses of Lemma 3.4.2, (C) the composition estimate holds in L^∞:*

$$\lim_{i \to 0} \left\| \sum_{j=1}^{i-1} \theta^j \left(X_\alpha \circ R^{(\infty,j)} - X_\alpha \circ R^{(i,j)} \right) \right\|_{L^\infty([0,1])} = 0.$$

Proof. As $\{ R_i \}$ converges uniformly to R_∞, for all $\epsilon > 0$, there exists an I_ϵ such that for a.e. $x \in [0,1]$:

$$|R_\infty(x) - R_i(x)| < \epsilon$$

for all $i \ge I_\epsilon$.

Fix $\epsilon > 0$ and

$$i > \max \left\{ I_\epsilon, \frac{\ln\left(\frac{\epsilon(1-\theta)}{2k(1-\theta^{I_\epsilon})}\right)}{\ln \theta} + I_\epsilon \right\}.$$

Now to show that

$$\left\| \sum_{j=1}^{i-1} \theta^j \left(X_\alpha \circ R^{(\infty,j)} - X_\alpha \circ R^{(i,j)} \right) \right\|_{L^\infty} < C\epsilon$$

we use the fact that, for large i, when j is small $R^{(i,j)}$ is close to $R^{(\infty,j)}$ and when j is large, θ^j is small. Indeed, we split the sum into two:

$$\begin{aligned}\left\| \sum_{j=1}^{i-1} \theta^j \left(X_\alpha \circ R^{(\infty,j)} - X_\alpha \circ R^{(i,j)} \right) \right\|_{L^\infty} &\le \sum_{j=0}^{i-I_\epsilon} \theta^j \left\| \left(X_\alpha \circ R^{(\infty,j)} - X_\alpha \circ R^{(i,j)} \right) \right\|_{L^\infty} \\ &\quad + \sum_{i-I_\epsilon+1}^{i-1} \theta^j \left\| \left(X_\alpha \circ R^{(\infty,j)} - X_\alpha \circ R^{(i,j)} \right) \right\|_{L^\infty}.\end{aligned}$$

For the first term, we have:

$$\begin{aligned}\sum_{j=0}^{i-I_\epsilon}\theta^j\left\|\left(X_\alpha\circ R^{(\infty,j)}-X_\alpha\circ R^{(i,j)}\right)\right\|_{L^\infty}&\le\sum_{j=0}^{i-I_\epsilon}\theta^j\mathrm{Lip}X_\alpha\left\|R^{(\infty,j)}-R^{(i,j)}\right\|_{L^\infty}\\&\le\mathrm{Lip}X_\alpha\epsilon\sum_{j=0}^{i-I_\epsilon}j\theta^j\qquad\text{by Lemma 3.4.2}\\&\le\frac{\mathrm{Lip}X_\alpha}{(1-\theta)^2}\epsilon\end{aligned}$$

where the last inequality is obtained using the fomula for the derivative of a geometric series.

For the second term,

$$\begin{aligned}\sum_{i-I_\epsilon+1}^{i-1}\theta^j\left\|\left(X_\alpha\circ R^{(\infty,j)}-X_\alpha\circ R^{(i,j)}\right)\right\|_{L^\infty}&\le\sum_{i-I_\epsilon+1}^{i-1}\theta^j2\bar{k}\\&=\theta^{i-I_\epsilon}\frac{2\bar{k}(1-\theta^{I_\epsilon})}{1-\theta}<\epsilon.\end{aligned}$$

The last inequality holds because of the choice defining of i. □

We prove convergence of the cumulative distribution functions of the adults' skill distributions and their inverses. The proof is based on a closed form representation of the dynamics in terms of the inverses of the cumulative distributions of the adults' skill distributions.

Theorem 3.4.4 (Long time limit of adult skills)**.** *Fix $c\ge0$, $\theta,\theta'\in[0,1]$, $N>1$, $N'>1$, $\kappa_1\in\mathcal{P}(\bar{K})$ and $\tilde{v}_1:\bar{K}\to\mathbf{R}$ assumed to be positive, convex and non-decreasing. Fix a distribution for students' abilities α with $\log\alpha\in L^\infty(\bar{A})$, so that the inverse X_α of the cumulative distribution function of α is Lipschitz continuous.*

Suppose solutions $\{(\epsilon_i,\lambda_i)\}$ of the primal overlapping generation problems (3.1) *or* (3.9) *satisfy (A) positive assortativity, (B) specialization and the rank functions R_i's satisfy (C) the composition estimate with a suitable norm $\|\cdot\|$.*

Then as $i\to\infty$, the cumulative distribution functions $F_{\kappa_i}(k):=\kappa_i([0,k])$ of the adult skill distributions converge uniformly on $[0,\bar{k}]$, and their inverses $X_{\kappa_i}:=F_{\kappa_i}^{-1}$ converge in $\|\cdot\|$ to

$$X_\infty:=\sum_{j=0}^{\infty}\theta^j(1-\theta)X_\alpha\circ R^{(\infty,j)}.\tag{3.29}$$

Proof. We express the probability measures α and κ in terms of non-decreasing random variables:

$$\begin{aligned}&X_\alpha:[0,1]\to\bar{A}\\&\qquad r\mapsto\text{ability of a student in the }r\text{-th percentile;}\\&X_\kappa:[0,1]\to\bar{K}\\&\qquad r\mapsto\text{ability of an adult in the }r\text{-th percentile.}\end{aligned}$$

The inverse function F_α of the random variable X_α (resp. F_κ of X_κ) is the cumulative distribution function of α (resp. of κ).

Using the equation z_E that sends the ability of a student and his or her teacher's skill to his or her future ability as an adult (2.3) it is possible to express the random variables representing the next generation of adults $X_{\kappa_{i+1}}$ using X_α, R_i and X_{κ_i}:

$$X_{\kappa_{i+1}} = (1-\theta)X_\alpha + \theta X_{\kappa_i} \circ R_i.$$

Thus, by induction we get:

$$X_{\kappa_{i+1}} = \theta^i X_{\kappa_1} \circ R^{(i,i)} + \sum_{j=0}^{i} \theta^j (1-\theta) X_\alpha \circ R^{(i,j)}, \tag{3.30}$$

where $R^{(i,j)}$ is the composition of R_{i-j+1} with R_{i-j+2} composed with the next one, continuing in this way until we reach R_i:

$$R^{(i,j)} = R_{i-j+1} \circ R_{i-j+2} \circ R_{i-j+3} \circ \cdots \circ R_{i-1} \circ R_i.$$

Let X_∞ defined by (3.29) be a candidate for the limit of the sequence $\{X_{\kappa_i}\}$, where $R^{(\infty,j)}$ is from (3.28).

We have that $R_\infty(x) \in [0,1]$. Therefore, by the monotonicity of X_α, there exists a global bound for X_∞:

$$X_\infty(x) \leq (1-\theta) \sum_{j=0}^{\infty} \theta^j X_\alpha(1) \leq X_\alpha(1).$$

The series X_∞ is convergent because it is a bounded series of positive terms.

We now prove that $\lim_{i\to\infty} X_{\kappa_i} = X_\infty$ in the norm $\|\cdot\|$ from (C) the composition estimate. Because $\|\cdot\| \leq \|\cdot\|_\infty$,

$$\begin{aligned} \|X_\infty - X_{\kappa_i}\| &= \left\| \sum_{j=0}^{\infty} (1-\theta)\theta^j X_\alpha \circ R^{(\infty,j)} - \theta^i X_{\kappa_1} \circ R^{(i,i)} - \sum_{j=0}^{i-1} (1-\theta)\theta^j X_\alpha \circ R^{(i,j)} \right\| \\ &\leq \sum_{j=i}^{\infty} (1-\theta)\theta^j \left\| X_\alpha \circ R^{(\infty,j)} \right\|_\infty + \theta^i \left\| X_{\kappa_1} \circ R^{(i,i)} \right\|_\infty \\ &\quad + (1-\theta) \left\| \sum_{j=0}^{i-1} \theta^j \left(X_\alpha \circ R^{(\infty,j)} - X_\alpha \circ R^{(i,j)} \right) \right\|. \end{aligned}$$

Since X_α, $X_{\kappa_1} \in [0, \bar{k}]$ and $0 < \theta < 1$, the first two terms go to 0 when $i \to \infty$. The last term also goes to zero by (C) the composition estimate. This concludes the proof that X_{κ_i} converges to X_∞.

As α doesn't have any atoms (by assumption), X_α is strictly increasing. Also R_∞ is strictly increasing by Lemma 3.4.1. Because X_α and R_∞ are strictly increasing,

$$X_\infty(x) = (1-\theta) \sum_{j=0}^{\infty} \theta^j X_\alpha \circ R^{(\infty,j)}(x)$$

is strictly increasing and therefore invertible on its image. Let F_∞ be the inverse of X_∞, extended monotonically to K.

We now prove that $\{ F_{\kappa_i} \}$ converges uniformly to F_∞. Suppose for contradiction that there exists a subsequence $\{ F_{\kappa_{i_j}} \}_{j=1}^{\infty}$ and a sequence of points $\{ x_{i_j} \}_{j=1}^{\infty}$ such that

$$\left| F_{\kappa_{i_j}}\left(x_{i_j}\right) - F_\infty\left(x_{i_j}\right) \right| > \epsilon. \tag{3.31}$$

By Lemma 3.3.6, $\{ F_{\kappa_i} \}$ is equi-Lipschitz (and therefore equicontinuous), so by the Arzelà-Ascoli theorem, there exists a subsequence that converges uniformly to a limit $\bar{F}_\kappa \neq F_\infty$. A further subsequence $\{ F_{\kappa_l} \}$ of the converging subsequence has the additional property that its inverse cumulative distributions converge pointwise to X_∞.

We now prove that $\{ F_{\kappa_l} \}$ converges pointwise to F_∞, which will contradict (3.31). First, we consider the subdomain $\bar{K} \cap \Im(X_\infty)$ and then we'll extend F_∞ on its complement.

By Lemma 3.3.6, $\{ F_{\kappa_i} \}$ is equi-Lipschitz (and therefore equicontinuous), so for $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|F_{\kappa_i}(x) - F_{\kappa_i}(y)| < \epsilon$$

whenever $|x - y| < \delta$.

By hypothesis, $\|X_{\kappa_l} - X_\infty\| \to 0$ implies there exists a subsequence $X_{\kappa_{l_j}}$ that converges to X_∞ a.e. We fix $x \in [0,1]$ such that $X_{\kappa_{l_j}}(x) \to X_\infty(x)$ and let

$$y = X_\infty(x) \qquad \text{and} \qquad y_j = X_{\kappa_{l_j}}(x).$$

As $\{ X_{\kappa_{l_j}}(x) \}$ converges to $X_\infty(x)$, there exists an $I \in \mathbf{N}$ such that $|y - y_j| < \delta$ for all $j > I$. Then for a.e. $y \in \mathrm{Im}\,(X_\infty)$,

$$\begin{aligned} \left| F_\infty(y) - F_{\kappa_{l_j}}(y) \right| &= \left| x - F_{\kappa_{l_j}}(y) \right| \\ &= \left| F_{\kappa_{l_j}}(y_j) - F_{\kappa_{l_j}}(y) \right| \\ &< \epsilon && \text{by equicontinuity of } \{ F_{\kappa_i} \}. \end{aligned}$$

There exists a unique continuous extension for F_∞ to $y \in \bar{K} \setminus \mathrm{Im}\,(X_\infty)$.

So we conclude that

$$F_{\kappa_{l_j}} \to F_\infty$$

pointwise a.e. which contradicts (3.31).

□

Given the difference between the two models, we now separate the analysis to deal with them individually. For the incomplete information model (introduced in Section 3.1), we prove, using duality, that the sequence of solutions converges to the solution of a limiting problem which is also a solution to the steady state problem from [13] (see Section 2.6).

For the complete information model (introduced in Section 3.2), we introduce a steady state model with discounting, and show that solutions to (3.9) and (3.10) converge to a solution of this new steady state problem.

3.5 Limit of Adult Skill Distributions for Incomplete Information Model

We prove that a converging sequence of solutions of the incomplete information model from Section 3.1 converges to a solution of the steady state model introduced in Section 2.6.

As the dual optimizers (u_i, v_i) of (3.4) are convex in $A \times K$, they are locally Lipschitz. Since the asymptotic analysis of [13] indicates boundedness of the steady state solution, it is reasonable to expect that the v_i are uniformly bounded, as we henceforth assume. Lemma 2.2.1 (c) then yields a subsequence (u_{i_j}, v_{i_j}) that converges uniformly on compact subsets of $A \times K$ to a limit (u_∞, v_∞).

We are not able to rule out the case where there exist different limits for the wage functions. For example, it might be possible that an asymptotic society would alternate between two states, with two different wage functions, though the author doesn't think it possible if b_E and b_L are convex. Therefore, we need to assume that the whole sequence of solutions from the dual problem (3.4) converges to a limit (u, v). In this case, a sequence of solutions from (3.1) (resp. from (3.4)) converges to a solution of the steady state model (2.5) (resp. of the dual of the steady state model (2.8)) from [13]

Proposition 3.5.1 (Steady characterization of eventual wages and skills)**.** *Fix $c \geq 0$, $\theta, \theta' \in [0,1]$, $N, N' > 1$, and $\kappa_1 \in \mathcal{P}(\bar{K})$. Fix a distribution for students' abilities α with $\log \alpha \in L^\infty(\bar{A})$.*

Suppose a sequence of solution $\{(\epsilon_i, \lambda_i)\}$ for the primal non-recursive incomplete information problem (3.1) *following the process explained in Section 3.1 satisfy (A) positive assortativity, (B) specialization, and (C) the composition estimate.*

Suppose (u_i, v_i) are optimizers of (3.4) *and the wage functions v_i are bounded independently of i. Let $\{(\epsilon_{i_l}, \lambda_{i_l}; u_{i_l}, v_{i_l})\}$ be a subsequence such that $(\epsilon_{i_l}, \lambda_{i_l})$ weak-$*$ converges to measures $(\epsilon_\infty, \lambda_\infty)$. and $\{(u_{i_l}, v_{i_l})\}$ is a subsequence of optimal wages that converges uniformly on compact subset of K to wage functions (u_∞, v_∞).*

By possibly taking another subsequence, suppose $\{v_{i_l - 1}\}$ converges to $\hat{v}$ which might be different then v_∞.

The functions (u_∞, v_∞) and measures $(\epsilon_\infty, \lambda_\infty)$ optimize the dual and primal non-recursive incomplete information problems (3.1)*–*(3.3) *and* (3.4)*–*(3.6) *with $\kappa = \kappa_\infty$ and $\tilde{v} = \hat{v}$, where the limit of the cumulative distribution functions $F_{\kappa_i}(x) = \kappa_i[0, x]$ of the adult skill distributions*

$$\lim_{i \to \infty} F_{\kappa_i}(k) := F_\infty(k)$$

is the cumulative distribution function of κ_∞.

Proof. Fix i. By Proposition 3.3.5, there exists a solution (ϵ_i, λ_i) of the primal overlapping generation problems (3.1). By Proposition 3.3.1, there exists an feasible optimizing pair (u_i, v_i) for (3.4). Therefore, we have

$$\epsilon_i (cb_E \circ z_E + v_{i-1} \circ z_E) + \lambda_i (b_L \circ z_L) = \alpha(u_i) + \kappa_i(v_i). \tag{3.32}$$

We take the limit for the subsequence i_l. As ϵ_{i_l} and λ_{i_l} weak-$*$ converge, we have

$$\epsilon_{i_l}(cb_E \circ z_E) \to \epsilon_\infty(cb_E \circ z_E)$$

and

$$\lambda_{i_l}(b_L \circ z_L) \to \lambda_\infty(b_L \circ z_L).$$

As u_{i_l} d_K-converges to u_∞ and $\alpha \in L^\infty(\bar{A})$, by Lemma 2.2.1 (b),

$$\alpha(u_{i_l}) \to \alpha(u_\infty).$$

By Lemma 3.3.6 the κ_i's are uniformly bounded in L^∞. So as $v_{i_l} \to v_\infty$ in d_K and $\kappa_i \overset{w^*}{\to} \kappa_\infty$ by Lemma 2.2.1 (b),

$$\kappa_{i_l}(v_{i_l}) \to \kappa_\infty(v_\infty).$$

By Lemma 3.3.6 the $z_{E\#}\epsilon_i$'s are also uniformly bounded in L^∞. So as $v_{i_l-1} \to \hat{v}$ in d_K and $z_{E\#}\epsilon_{i_l} \overset{w^*}{\to} z_{E\#}\epsilon_\infty$ by Lemma 2.2.1 (b),

$$z_{E\#}\epsilon_{i_l}(v_{i_l-1}) \to \kappa_\infty(\hat{v}).$$

Using these limit in (3.32), we have

$$\epsilon_\infty(cb_E \circ z_E) + \epsilon_\infty(\hat{v}(z_E)) + \lambda_\infty(b_L \circ z_L) = \alpha(u_\infty) + \kappa_\infty(v_\infty)$$

From Theorem 3.3.10 we conclude that $(\epsilon_\infty, \lambda_\infty)$ is an optimal solution of (3.1) and (v_∞, u_∞) is an optimal solution of (3.4) with $\kappa = \kappa_\infty$ and $\tilde{v} = \hat{v}$. □

When $\hat{v}$ is equal to v_∞ in Proposition 3.5.1 we'll prove in Theorem 3.5.3 that the limit of solutions of the incomplete information model is a solution of the steady state problem from [13] which is recalled in Section 2.6. This will be the case when optimal wage functions are unique for every generation and converge globally. We present a specific regime when this is the case.

Proposition 3.5.2. *Fix $c \geq 0$, $\theta, \theta' \in [0,1]$, $N, N' > 1$, and $\kappa_1 \in \mathcal{P}(\bar{K})$ is connected. Fix a distribution for students' abilities α supported on $\bar{A}$ with $\log \alpha \in L^\infty(\bar{A})$.*

Suppose hypotheses (a) *and* (b) *from Theorem 3.3.7 and hypotheses* (d), (e) *and* (f) *from Proposition 3.3.9 hold, as well as (A) positive assortativity, (B) specialization and (C) the composition estimate from Theorem 3.4.4.*

Then the sequence of wage functions $\{ v_i \}$ that solves the incomplete information model introduced in section 3.1 converges uniformly on compact subsets to a wage function v_∞.

Proof. Because of Theorem 3.3.7, the skill of workers is weakly below the skill of managers which is weakly below the skill of teachers for any generation. As the support of κ_1 is connected, that means that the support for the skill of teachers is also connected. Therefore as the support of α is $\bar{A}$, which is a connected set, we know that the support of κ_2 will also be connected and, by induction, we conclude that the support of κ_i is connected for all i.

As $\{ \kappa_i \}$ is convergent by Theorem 3.4.4 and the order of the professions is constant, we conclude that $\lambda_i \to \lambda_\infty$ and $\epsilon_i \to \epsilon_\infty$ are also convergent. We will show that (u_i, v_i) is convergent by constructing a unique pair (u_∞, v_∞) that solves the incomplete information model (3.1) with $\kappa_1 = \kappa_\infty$ and $\tilde{v} = v_\infty$. To construct v_∞, we begin with a wage function v_0 and we construct the sequence $\{ v_i \}$ such that there exists a u_i such that (u_i, v_i) is a solution of (3.4) with $\kappa_1 = \kappa_\infty$ and $\tilde{v} = v_{i-1}$. Note that with our assumptions, the solution of the primal incomplete information model (3.1) with $\kappa_1 = \kappa_\infty$ and $\tilde{v} = v_{i-1}$ is constant $(\lambda_\infty, \epsilon_\infty)$.

Because of our strict hypotheses, the optimal wage function v is uniquely dertermined when it doesn't depend on the initial wage function v_0. The seed velocity v_0 affects only the wages of teachers at the first generation, not of managers and workers. Hence it affects only the wages of teachers of teachers at the second generation, not of managers, workers, or their teachers. Similarly, at the n-th generation, it affects only the wages of teachers whose first $n-1$ generations of academic descendants are all teachers, not workers, managers, or teachers whose academic descendants in the next $n-1$ generations include workers or managers. By Proposition 3.3.9 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, so the teachers whose academic descendants are not workers nor managers for the next $n-1$ generations become an arbitrarily small proportion of the population as n tends to infinity.

We'll denote $K_0 \subset K$ to represent the support of workers and managers, $\mathrm{spt}(\kappa^w + \kappa^m)$, and $K_j \subset K$ to represent the support of workers, managers and teachers with at most j descendants (therefore K_0 is an extension of this definition where workers and managers are considered to be teachers with no descendants).

By the first order condition, we have that $v_1'(k) = (1-\theta')b_L'(z_l(k,m))$ on the support of workers, where $m \in K$ is such that $\lambda_1(k,m) > 0$ and $v_1'(k) = N'\theta' b_L'(z_l(w,k))$ on the support of managers, where $w \in K$ is such that $\lambda_1(w,k) > 0$. We will now prove that v_1 is uniquely determined on K_0. This is because K_0 is connected. Suppose there exist two $v_1, \tilde{v}_1$ that are part of an optimizing pair (u_1, v_1) and $(\tilde{u}_1, \tilde{v}_1)$. We have that $v_1' = \tilde{v}_1'$ and that both functions are non-decreasing and convex, so as K_0 is connected, we conclude that $v_1 = \tilde{v}_1 + l$ on K_0. Suppose WLOG that $l \geq 0$. We have

$$b_L(z_L(w,m) \leq v_1(w) + \frac{v_1(m)}{N'}$$

with equality on the support of λ_1. Replacing v_1 by $\tilde{v}_1 + l$, we get

$$b_L(z_L(w,m) \leq \tilde{v}_1(w) + \frac{\tilde{v}_1(m)}{N'} + \left(1 + \frac{1}{N'}\right) l$$

with equality on the support of λ_1. We conclude that $l = 0$ and $v_1 = \tilde{v}_1$ on K_0.

Now suppose v_i is uniquely determined on K_{i-1}. We can show similarly that v_{i+1} is uniquely determined on K_i. Now suppose that there exists (u_{i+1}, v_{i+1}) and $(\tilde{u}_{i+1}, \tilde{v}_{i+1})$ that are both optimal on K_i. We have $v_{i+1}' = \tilde{v}_{i+1}'$ and $u_{i+1}' = \tilde{u}_{i+1}'$ thus $v_{i+1} = \tilde{v}_{i+1} + l$ on $K_i \setminus K_0$ because the skill of teachers is connected and $u_{i+1} = \tilde{u}_{i+1} + n$ on A. As it is optimal,

$$\alpha(u_{i+1}) + \kappa_\infty(v_{i+1}) = \alpha(\tilde{u}_{i+1}) + \kappa_\infty(v_{i+1})$$

and therefore $l + n = 0$. We have

$$b_E(z_E(s,t)) + v_i(z_E(st)) \leq u_{i+1}(s) + \frac{v_{i+1}(t)}{N}$$

with equality on the support of λ_{i+1}. Replacing v_{i+1} by $\tilde{v}_{i+1} + l$, and u_{i+1} by $\tilde{u}_{i+1} + n$ we get

$$b_E(z_E(s,t)) + v_i(z_E(st)) \leq \tilde{u}_{i+1}(s) + \frac{\tilde{v}_{i+1}(t)}{N} + n + \frac{l}{N} = \tilde{u}_{i+1}(s) + \frac{\tilde{v}_{i+1}(t)}{N} + \left(1 - \frac{1}{N}\right) n$$

and we conclude that $n = l = 0$ and v_{i+1} is unique on K_i.

We constructed a unique v_∞ on $\cup_{i=1}^{\infty} K_i$ which is the support of κ_∞ by Proposition 3.3.9. Any other candidate $\tilde{v}_\infty$ can't solve (3.4) with $\kappa_1 = \kappa_\infty$ and $\tilde{v} = \tilde{v}_\infty$ because we would get a sequence of wage function converging to v_∞ by using the process we just described. Therefore v_∞ is unique and we conclude that $\{ v_i \}$ is convergent. □

Theorem 3.5.3. *Under the hypotheses of Proposition 3.5.1, if $\hat{v} = v_\infty$ the functions (u_∞, v_∞) and measures $(\epsilon_\infty, \lambda_\infty)$ optimize the dual and primal steady state problems from [13] which are recalled in Section 2.6. The limit of the cumulative distribution functions $F_{\kappa_i}(x) = \kappa_i[0,x]$ of the adult skill distributions*

$$\lim_{i\to\infty} F_{\kappa_i}(k) := F_\infty(k)$$

is the cumulative distribution function of the adult skill distribution associated to a solution of (2.5), the steady state problem from [13]. That is, there exists an optimal pair (ϵ, λ) of (2.5) such that $\kappa_\infty = z_{E\#}\epsilon$ has cumulative distribution F_∞.

Proof. From Theorem 3.3.10, we have

$$\epsilon_\infty(cb_E \circ z_E) + \epsilon_\infty(\hat{v}(z_E)) + \lambda_\infty(b_L \circ z_L) = \alpha(u_\infty) + \kappa_\infty(v_\infty)$$

Setting $\hat{v} = v_\infty$, and noting that $\kappa_\infty = z_{E\#}\epsilon_\infty$, we have $\epsilon_\infty(\hat{v} \circ z_E) = \kappa_\infty(v_\infty)$ and therefore

$$\epsilon_\infty(cb_E \circ z_E) + \lambda_\infty(b_L \circ z_L) = \alpha(u_\infty).$$

From the duality theory of [13] we conclude that $(\epsilon_\infty, \lambda_\infty)$ is an optimal solution of (2.5) and (u_∞, v_∞) is an optimal solution of (2.8).

□

Note that Theorem 3.5.3 and the proof of Theorem 3.4.4 can be used to get an explicit formulation for the distribution of adults' skills for the steady state case [13] (see Section 2.6), when the associate solution to the non-recursive incomplete information model is unique.

Corollary 3.5.4 (Explicit form of steady state)**.** *Fix $c \geq 0$, $\theta, \theta' \in [0,1]$, $N, N' > 1$. Fix a distribution for students' abilities α with $\log \alpha \in L^\infty(\bar{A})$. Consider the solution to the steady state problem introduced in Section 2.6. Suppose λ is positive assortative and ϵ is strictly positive assortative. Let $\bar{\kappa}$ be the distribution for skills of adults.*

Let $F_\alpha(x) := \alpha[0,x]$ (resp. $F_{\bar{\kappa}}$) be the cumulative distribution function of α (resp. $\bar{\kappa}$) and let X_α (resp. $X_{\bar{\kappa}}$) be its inverse. Let R_∞ be the function that sends the rank of a student to the rank of his or her teacher.

If hypothesis (f) from Proposition 3.3.9 is satisfied, and the skill of workers and managers are connected in K then the steady state $X_{\bar{\kappa}} = X_\infty$ is given by (3.29).

Proof. As (f) is satisfied, and the skill of workers and managers are connected in K the solutions (ϵ, λ), (u, v) are unique solution of (3.1), with $\kappa = z_{E\#}\epsilon = \bar{\kappa}$ and $\tilde{v} = v$. Therefore, we can use Theorem 3.4.4 to deduce $X_{\bar{\kappa}}$. □

3.6 Limit of Adult Skill Distributions for Complete Information Model

For the complete information model we adopt a similar approach. However, discounting prevents the solutions from converging to solutions of the steady state model of [13]. Thus we need to replace it with a suitably adapted limiting model.

In order to create a steady state problem arising from the complete information model, we adapt the Lagrangian from the steady state model by replacing part of the resulting distribution $z_{E\#}\epsilon$ of adult skills with an anticipated distribution κ_∞ which, along with the discount factor $\beta < 1$, becomes a parameter in the model. This division is calculated to preserve the mass of the skill distribution for adults, while discounting future wages relative to tuition expenses in the education market's stability constraint. The resulting Lagrangian becomes

$$\begin{aligned} L(\epsilon, \lambda; u, v) = \alpha(u) &+ (1 - e^{-\beta})\kappa_\infty(v) \\ &+ c\epsilon(b_E \circ z_E) - \pi_{1\#}\epsilon(u) + e^{-\beta} z_{E\#}\epsilon(v) - \frac{\pi_{2\#}\epsilon(v)}{N} \\ &+ \lambda(b_L \circ z_L) + \alpha(u) - \pi_{1\#}\lambda(v) - \frac{\pi_{2\#}\lambda(v)}{N'} \end{aligned}$$

and depends on the parameters $\beta \in (0, 1)$ and $\kappa_\infty \in \mathcal{P}(K)$. Since this Lagrangian is bilinear in its variables we anticipate the minimax relation

$$\begin{aligned} S_\beta(\kappa_\infty) &:= \sup_{\epsilon,\lambda\geq 0} \inf_{u,v\in C(\bar{K})} L(\epsilon, \lambda; u, v) \\ &\leq \inf_{u,v\in C(\bar{K})} \sup_{\epsilon,\lambda\geq 0} L(\epsilon, \lambda; u, v) \\ &=: S^*_\beta(\kappa_\infty) \end{aligned} \tag{3.33}$$

becomes an equality. Under simplifying hypotheses we'll establish this presently. In the meantime it is straightforward to compute

$$\begin{aligned} S_\beta(\kappa_\infty) = \max_{0\leq\epsilon,\lambda} &\int_{A\times K} cb_E\big(z_E(s,t)\big)d\epsilon(s,t) + \int_{K\times K} b_L\big(z_L(w,m)\big)d\lambda(w,m) \\ \text{s.t.} \quad & \alpha = \pi_{1\#}\epsilon \qquad \text{steady state for students} \\ & e^{-\beta} z_{E\#}\epsilon + (1 - e^{-\beta})\kappa_\infty = \frac{\pi_{2\#}\epsilon}{N} + \pi_{1\#}\lambda + \frac{\pi_{2\#}\lambda}{N'} \qquad \text{steady state for adults} \end{aligned} \tag{3.34}$$

by carrying out the infimum over the wage function $v : K \to \mathbf{R}$ for adults and the indirect utility $u : K \to \mathbf{R}$ for students. Similarly, carrying out the supremum over the education and labour pairings $\epsilon \geq 0$ and $\lambda \geq 0$ yields the linear program dual to (3.34):

$$\begin{aligned} S^*_\beta(\kappa_\infty) := \inf_{u,v} &\int_A u(s)\alpha(s) + (1 - e^{-\beta}) \int_K v(k)\kappa_\infty(k) \\ \text{s.t.} \quad & cb_E(z_E(s,t)) + e^{-\beta} v(z_E(s,t)) \leq u(s) + \frac{v(t)}{N} \qquad \text{stability of the education market} \\ & b_L(z_L(w,m)) \leq v(w) + \frac{v(m)}{N'} \qquad \text{stability of the labour market.} \end{aligned} \tag{3.35}$$

Naturally, the optimizers (ϵ, λ) of (3.34) and (u, v) of (3.35) depend on the parameters (β, κ_∞). Given $0 < \beta < 1$, our "steady state model with discounting" is to find a distribution κ_∞ of adult skills for which the optimal (ϵ, λ) in (3.34) satisfy the self-consistency condition

$$z_{E\#}\epsilon = \kappa_\infty, \tag{3.36}$$

meaning the resulting distribution of adult skills $(z_E)_{\#}\epsilon$ agrees with the proposed distribution κ_∞.[6]

By extracting a long time limit from our complete information model under conditions analogous to those imposed in the incomplete information model, we shall show that the steady state with discounting model (3.34)–(3.36) admits a self-consistent solution.

First, we prove the easy inequality $S_\beta(\kappa_\infty) \leq S^*_\beta(\kappa_\infty)$ from (3.33) which relates (3.34) to (3.35) continues to hold when the continuity of $u, v \in C(K)$ in relaxed to permit the possibility of unboundedness at the upper endpoint $\bar{k}$. As shown in [13, Proposition 8] the doubling condition (3.7) on α then guarantees that $v \in L^1(\bar{K}, z_{\#}\epsilon)$, which is sufficient to prove the desired inequality.

Lemma 3.6.1 (Easy direction of duality)**.** *Fix $c \geq 0$, $\theta, \theta' \in [0,1]$, $N > 1$, $N' > 1$. Let $A = K = [0, \bar{k})$ and $\beta \geq 0$. Suppose $e^{-\beta} > \frac{1}{N}$. Let κ_∞ be a Borel probability measure on $\bar{K}$ and fix a distribution for students' abilities α with $\alpha \in L^\infty(\bar{A})$.*

Define $b_{E/L}$ and $z_{E/L}$ as in subsection 2.5. Let (ϵ, λ) be a feasible candidate for the primal problem (3.34). Let $(u, v) = (u_0 + u_1, v_0 + v_1)$ be feasible for the dual (3.35) which differ from bounded continuous functions $u_0, v_0 \in C(\bar{A})$ by non-decreasing functions $u_1, v_1 : \bar{A} \to [0, \infty)$. Suppose moreover that $u \in L^1\left(\bar{A}, \alpha\right)$ and $v \in L^1\left(\bar{A}, z_{E\#}\epsilon\right) \cap L^1\left(\bar{K}, \kappa_\infty\right)$ then

$$c\epsilon(b_E \circ z_E) + \lambda(b_L \circ z_L) \leq \alpha(u) + \kappa_\infty(v).$$

If α satisfies the doubling condition (3.7), then $v \in L^1(\bar{A}, z_{E\#}\epsilon)$.

Proof. This proof follows the proof of [13, Proposition 8]. Let (ϵ, λ) be feasible for (3.34) and (u, v) feasible for (3.35) such that $u \in L^1(\bar{A}, \alpha)$ and $v \in L^1\left(\bar{A}, z_{E\#}\epsilon\right) \cap L^1\left(\bar{K}, \kappa_\infty\right)$. By the stability constraint for the education sector, we have

$$u(s) - cb_E(z_E(s,t)) \geq e^{-\beta} v(z_E(s,t)) - \frac{v(t)}{N}$$

on $\bar{A} \times \bar{K}$. By hypothesis the left hand side is in $L^1\left(\bar{A} \times \bar{K}, \epsilon\right)$, so

$$\infty > \alpha(u) - c\epsilon(b_E \circ z_E) \geq \int_{A \times K} \left(e^{-\beta} v(z_E(s,t)) - \frac{1}{N} v(t)\right) \epsilon(s,t). \tag{3.37}$$

Now integrating v over the steady state constraint for adults gives

$$\int_K v\left(e^{-\beta} z_{E\#}\epsilon + (1 - e^{-\beta})\kappa_\infty - \frac{\pi_{2\#}\epsilon}{N}\right) = \int_K v\left(\pi_{1\#}\lambda + \frac{\pi_{2\#}\lambda}{N'}\right) \tag{3.38}$$

$$\geq \int_{K \times K} b_L \circ z_L \lambda > 0 \tag{3.39}$$

where the last inequality follows from stability of the labour sector.

[6] We could consider this problem recursively as we did for the incomplete information model from Section 3.1 in order to get an incomplete information model with discounting, but we won't go there here.

Adding $(1-e^{-\beta})\kappa_\infty(v) < \infty$ to (3.37) makes its right hand side equal the left hand side of (3.38). In that case, $(1-e^{-\beta})\kappa_\infty$ plus the left hand side of (3.37) is greater or equal to (3.39) which is the desired inequality.

Now, we show that the doubling condition (3.7) on α at $\bar{a}$ implies that $0 \le v \in L^1(\bar{A}, z_{E\#}\epsilon)$. Recall that $(u,v) = (u_0+u_1, v_0+v_1)$ with $u_0, v_0 \in C(\bar{A})$ and $u_1, v_1 : \bar{A} \to [0,\infty]$ non-decreasing. Since v_0 is bounded it is integrable. We will show that $v_1 \in L^1(\bar{A}, z_{E\#}\epsilon)$. Since v_1 is strictly increasing, $v_1^{-1}(y) \in \bar{A}$ is well defined. We have:

$$
\begin{aligned}
\int_{\bar{K}} v_1(k) z_{E\#}\epsilon(dk) &= \int_0^\infty z_{E\#}\epsilon\left(v_1^{-1}[y,\infty]\right) dy && \text{by the layer-cake representation}\\
&= \int_0^\infty z_{E\#}\epsilon\left(\bar{a} - \left(\bar{a} - v_1^{-1}[y,\infty]\right)\right) dy\\
&\le \int_0^\infty \alpha\left(\bar{k} - \frac{1}{1-\theta}(\bar{a} - v_1^{-1}(y)), \bar{a}\right) dy && \text{by [13, Lemma 14]}\\
&\le C^{\frac{1}{\theta}-1}\int_0^1 \alpha[v_1^{-1}(y), \bar{a}] dy && (3.40)
\end{aligned}
$$

for some $C < \infty$. From stability of the education market and $e^{-\beta} > \frac{1}{N}$ we have $v < u$. Thus, $v_1 \le u_0 + u_1 - v_0 \le u_1 + l$ where l is a constant. Therefore, $u_1^{-1}(y-l) \le v_1^{-1}(y)$, so

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

implies that (3.40) is finite and so $v_1 \in L^1(\bar{K}, z_{E\#}\epsilon)$. □

We could prove that there is no duality gap using Section B.1 and we expect that existence of optimizing wages can be proved as in [13]. However, it is not obvious how the self-consistency condition (3.36) might be achieved using such techniques. Instead of following this path, we shall extract a limit from the optimizer of the complete information model from Section 3.2 and show that it provides self-consistent solution to the steady state with discounting, thus establishing existence of an optimizer and absence of a duality gap under suitable (but somewhat restrictive) hypotheses.

Lemma 3.6.2 (C is continuous with respect to the weak-$*$ topology)**.** *Fix $c \ge 0$, $\theta, \theta' \in [0,1]$, $N, N' > 1$. Let $A = K = [0, \bar{k})$ and $\beta \ge 0$. Fix a distribution for students' abilities α with $\log\alpha \in L^\infty(\bar{A})$.*

Let $\{\kappa_j\}_{j=1}^\infty$ be a sequence that is uniformly bounded in L^∞ such that $\kappa_j \to \kappa_\infty$ in the weak-$*$ topology. Then $\lim_{j\to\infty} C(\kappa_j) = C(\kappa_\infty)$ in (3.9).

Proof. Fix j. By Proposition 3.3.5 there exists an optimizing sequence for $C(\kappa_j)$

$$
\{(\epsilon_i^j, \lambda_i^j)\}_{i=1}^\infty .
$$

By Proposition 3.3.1 there exists an optimizing sequence $\{(u_i^j, v_i^j)\}_{i=1}^\infty$. By Corollary B.2.3, we can take the limit $T \to \infty$ in Corollary B.1.3 and we have

$$
\sum_{i=1}^\infty e^{-\beta i}\left(c\epsilon_i^j(b_E \circ z_E) + \lambda_i^j(b_L \circ z_L)\right) = e^{-\beta}\kappa_j(v_1^j) + \sum_{i=1}^\infty e^{-\beta i}\alpha(u_i^j).
$$

For all i, $\{(\epsilon_i^j, \lambda_i^j)\}_{j=1}^\infty$ forms a sequence and the Prokhorov Theorem [30] couples with a diagonal argument to produce a weak-$*$ converging subsequence $\{ (\epsilon_i^{j_l}, \lambda_i^{j_l}) \}_{l=1}^{\infty}$ that converges for all i. Let $(\hat{\epsilon}_i, \hat{\lambda}_i)$ be the limit of that converging subsequence. By Proposition 3.3.1, for all i and j, u_i and v_i are non-decreasing and convex. By Lemma 2.2.1 (c) $(u_i^{j_l}, v_i^{j_l})$ has a d_K-converging subsequence. Taking the limit as $j \to \infty$ along the converging subsequence (using Lemma 2.2.1 (b) for the $\kappa_j(v_1^j)$ term), we get:

$$\sum_{i=1}^{\infty} e^{-\beta i} \left(c\epsilon_i^{\infty}(b_E \circ z_E) + \lambda_i^{\infty}(b_L \circ z_L) \right) = e^{-\beta} \kappa_j(v_1^{\infty}) + \sum_{i=1}^{\infty} e^{-\beta \infty} \alpha(u_i^{\infty})$$

and therefore, $\{ (\epsilon_i^{\infty}, \lambda_i^{\infty}) \}_{i=1}^{\infty}$ is an optimizer for $C(\kappa_\infty)$ and the result follows from linearity of the objective functional. □

Similarly, we can prove the following lemma.

Lemma 3.6.3. *When κ_i weak-$*$ converges to κ_∞ we have*

$$\mathcal{P}(\kappa_i, \kappa_{i+1}) \to \mathcal{P}(\kappa_\infty, \kappa_\infty),$$

where the doubly constrained maximum $\mathcal{P}$ over one generation is defined by (3.16)*.*

We can now prove the main theorem of this section.

Theorem 3.6.4 (Optimizers of the complete information model converge to optimizers of the steady state with discounting)**.** *Fix $c > 0$, $\theta, \theta' \in [0,1]$, $N > 1$, $N' > 1$. Let $\kappa_1 \in \mathcal{P}(\bar{K})$ and fix a distribution for students' abilities α with $\log \alpha \in L^{\infty}(\bar{A})$. Suppose hypothesis* (b) *from Proposition 3.3.9 is satisfied.*

Suppose (ϵ_i, λ_i) is an optimizer of (3.9) *and satisfies (A) positive assortativity, (B) specialization and (C) the composition estimate.*

Suppose (u_i, v_i) is an optimizer of (3.10) *and the wage functions v_i are bounded independently of i.*

Then (u_i, v_i) has a subsequence that converges uniformly on bounded subsets of $A \times K$ to wage functions (u_∞, v_∞). The solutions of the primal complete information problem (3.9) *$\{ \epsilon_i, \lambda_i) \}$ have a weak-$*$ convergent subsequence that converges to $(\epsilon_\infty, \lambda_\infty)$.*

Suppose hypothesis (d) *from Proposition 3.3.9 are satisfied and (u_∞, v_∞) are strictly convex and the skills of workers and managers form a connected subset of K. The constant sequence of measures $\{ (\epsilon_\infty, \lambda_\infty) \}_{i=1}^{\infty}$ solves* (3.9) *and the constant sequence of functions $\{ (u_\infty, v_\infty) \}_{i=1}^{\infty}$ solves* (3.10) *with $\kappa_1 = z_{E\#}\epsilon_\infty$. In that case, $(\epsilon_\infty, \lambda_\infty)$ solves the steady state model with discounting* (3.34) *with $\kappa_\infty = z_{E\#}\epsilon_\infty$ and (u_∞, v_∞) solve its dual* (3.35)*.*

Proof. Consider solutions of the complete information model $\{(\epsilon_i, \lambda_i)\}$ of (3.9) and $\{(u_i, v_i)\}$ of (3.10). By Prokhorov's theorem [30], $\{ (\epsilon_i, \lambda_i) \}$ has a convergent subsequence which converges to a limit $(\epsilon_\infty, \lambda_\infty)$. As u_i, v_i are convex and bounded, by Lemma 2.2.1 (c), we know that there exists a convergent subsequence:

$$(u_{i_j}, v_{i_j}) \to (u_\infty, v_\infty).$$

By Theorem 3.4.4, κ_i weak-$*$ converges to κ_∞.

We now consider a problem with the same hypotheses as above except κ_1 is replaced with $\tilde{\kappa}_1 = z_{E\#}\epsilon_\infty = \kappa_\infty$. This new problem has its own solutions for the primal $(\tilde{\epsilon}_i, \tilde{\lambda}_i)$ and for the dual $(\tilde{u}_i, \tilde{v}_i)$. First, we prove that for all i, $\tilde{\epsilon}_i = \epsilon_\infty$, $\tilde{\lambda}_i = \lambda_\infty$, and $\tilde{u}_i = u_\infty$.

By Proposition 3.2.3 we get

$$\begin{aligned}C(\kappa_i) &= \sup_{\kappa \in \mathcal{N}(\kappa_i)} \mathcal{P}(\kappa_i, \kappa) + e^{-\beta} C(\kappa) \\ &= \mathcal{P}(\kappa_i, \kappa_{i+1}) + e^{-\beta} C(\kappa_{i+1}).\end{aligned}$$

so by Lemmas 3.6.2 and 3.6.3 when we take the limit $i \to \infty$ we get:

$$C(\kappa_\infty) = \mathcal{P}(\kappa_\infty, \kappa_\infty) + e^{-\beta} C(\kappa_\infty).$$

Since $\tilde{\kappa}_1 = \kappa_\infty$ this shows we may take $\tilde{\kappa}_2 = \kappa_\infty$ and hence $\tilde{\kappa}_i = \kappa_\infty$ for all i. As $\tilde{\kappa}_i$ is in a steady state, we can conclude from Proposition 3.3.5 that $\tilde{\epsilon}_i = \tilde{\epsilon}_\infty$ and $\tilde{\lambda}_i = \tilde{\lambda}_\infty$ for all i. Also, λ_i and ϵ_i are all positive assortative.

For all i, the proof that $\tilde{v}_i'$ doesn't depend on i for workers, i.e. on $\pi_1(\text{spt }\lambda) \cap \text{Dom }\tilde{v}_i'$, and for managers, i.e. on $\pi_2(\text{spt }\lambda) \cap \text{Dom }\tilde{v}_i'$ is identical to the one from [13, Theorem 15].

Now for teachers, given that we assume hypothesis (d) from Proposition 3.3.9 and $c > 0$, by Proposition 3.3.9 almost every teacher has at most finitely many descendants and therefore we'll be able to establish the κ_∞-a.e. uniqueness of v_i' from the uniqueness of the wage gradient for workers or managers of future generations v_{i+n}'.

Using uniqueness of $\tilde{v}_i'$, we can then prove uniqueness of $\tilde{u}_i'$. Given that $\tilde{\epsilon}_i = \epsilon_\infty$ and $\tilde{\lambda}_i = \lambda_\infty$ are constant from generation to generation, we have that $\tilde{v}_i' = v_\infty'$ and $\tilde{u}_i' = u_\infty'$ are also constant from generation to generation.

Now, uniqueness of u_∞ can be shown as in [13, Theorem 15] because u_∞ is in the objective function of (3.10) and α is supported on all A. Finally, we show κ_∞-a.e. uniqueness of v_∞.

As in [13, Theorem 15], we also have that v_∞' is unique. Suppose there exists an i such that there exist two different potential values v and $\hat{v}$ for v_i.

We know that $v_i' = v' = \hat{v}'$. As we assumed that the skills of workers and managers are connected in K, we have $v = \hat{v} + l$ on the support of the skills of workers and managers. Suppose WLOG that $l \geq 0$. As $(u_\infty, v_\infty)_{i=1}^\infty$ is optimal for (3.10) with $\kappa_i = \kappa_\infty$, we have

$$b_L(z_L(w, m)) \leq v(w) + \frac{v(m)}{N'}$$

with equality on the support of λ_∞. We also have

$$b_L(z_L(w, m)) \leq \hat{v}(w) + \frac{\hat{v}(m)}{N'} + \left(1 + \frac{1}{N'}\right) l$$

with equality on the support of λ_∞, which is impossible unless $l = 0$. We conclude that $l = 0$ and $v = \hat{v}$ on the support of the skills of workers and managers.

Now by Proposition 3.3.9 the wages of teachers will eventually be expressed with the functions b_E, u, and v on the support of the skills of workers and managers. Therefore v is also unique for the skills of teachers. We conclude that $v_i = v_\infty$ for all i on the support of the skills of adults.

Using Corollary 2.3, we get the equivalent of Corollary B.1.3 in the infinite horizon case, which implies

$$\sum_{i=1}^{\infty} e^{-\beta i} \left(c\tilde{\epsilon}_i(b_E \circ z_E) + \tilde{\lambda}_i(b_L \circ z_L) \right) = e^{-\beta}\tilde{\kappa}_1(\tilde{v}_1) + \sum_{i=1}^{\infty} e^{-\beta i}\alpha(\tilde{u}_i).$$

So

$$\begin{aligned} &\sum_{i=1}^{\infty} e^{-\beta i} \left(c\epsilon_\infty(b_E \circ z_E) + \lambda_\infty(b_L \circ z_L) \right) = e^{-\beta}\kappa_\infty(v_\infty) + \sum_{i=1}^{\infty} e^{-\beta i}\alpha(u_\infty) \\ \iff &c\epsilon_\infty(b_E \circ z_E) + \lambda_\infty(b_L \circ z_L) = \alpha(u_\infty) + (1 - e^{-\beta})\kappa_\infty(v_\infty). \end{aligned}$$

This is the equilibrium formula of the steady state with discounting models (3.34) and (3.35). □

Note 3.6.5. *Under the hypotheses of Theorem 3.6.4, this construction yields a self-consistent solution for the steady state with discounting* (3.34)*.*

Note 3.6.6. *Under the hypotheses of Theorem 3.3.7 and Proposition 3.3.9, the solution of the primal complete information problem* (3.9) $\{ (\epsilon_i, \lambda_i) \}$ *converges weak-∗ to* $(\epsilon_\infty, \lambda_\infty)$*. Therefore, the solution to the dual problem is also convergent if the solution of the steady state with discounting is unique (see Proposition 3.5.2).*

3.7 Examples

We now study some specific examples. First, we consider the case where the skills of teachers are weakly above the skills of managers, which are weakly above the skills of workers. Conditions for this to be realized are presented in Theorem 3.3.7.

In this case, the function that sends the rank of a student to the rank of his or her teacher for any generation i is linear and is given by

$$R_i(r) = 1 + \frac{r-1}{N}.$$

As R_i doesn't depends on i, $R^{(i,j)} = R_i^{(j)}$ and we get:

$$R^{(i,j)}(r) = 1 + \frac{r-1}{N^j}.$$

Therefore, from Corollary 3.5.4

$$\begin{aligned} X_{\bar{\kappa}}(r) &= (1-\theta)\sum_{j=0}^{\infty} \theta^j X_\alpha \circ R^{(i,j)}(r) \\ &= (1-\theta)\sum_{j=0}^{\infty} \theta^j X_\alpha \left(1 + \frac{r-1}{N^j}\right). \end{aligned}$$

If we also assume that α is the uniform measure on $[0,1]$, then the random variable X_α sends r to r. In this case

$$\begin{aligned}
X_{\bar{\kappa}}(r) &= (1-\theta)\sum_{j=0}^{\infty}\theta^j\left(1+\frac{r-1}{N^j}\right)\\
&= (1-\theta)\sum_{j=0}^{\infty}\theta^j + (1-\theta)(r-1)\sum_{j=0}^{\infty}\frac{\theta^j}{N^j}\\
&= 1+(1-\theta)(r-1)\frac{1}{1-\frac{\theta}{N}}\\
&= \frac{N(1-\theta)}{N-\theta}r + \frac{\theta(N-1)}{N-\theta}.
\end{aligned}$$

We deduce that $\bar{\kappa}$ is the uniform probability measure on $\left[\frac{\theta(N-1)}{N-\theta},1\right]$:

$$\bar{\kappa} = \frac{N-\theta}{(1-\theta)N}\mathbb{1}|_{\left[\frac{\theta(N-1)}{N-\theta},1\right]}.$$

More generally, we can prove similarly that if α is a uniform probability measure and the R_i's are linear functions that are independant of i, $\bar{\kappa}$ will also be a uniform probability measure.

3.7.1 Explicit Calculation of the Dual

In this subsection, we use Corollary 3.5.4 and more precisely the primal solution we just obtained to find explicit solutions to the dual problem from the steady state labour and education matching model from Erlinger et al., when λ is positive assortative and ϵ is strictly positive assortative, the skills of teachers are weakly above the skills of managers, which are weakly above the skills of workers and α is the uniform measure on $[0,1]$.

Support for skills and abilities

We have that the adults' skills take values between $\frac{\theta(N-1)}{N-\theta}$ and 1. As we know the ratio of adults for each profession (see [13, p.15]), we get that workers' skills are between $\frac{\theta(N-1)}{N-\theta}$ and $\frac{(N'+\theta)(N-1)}{(1+N')(N-\theta)}$, and managers' skills are between $\frac{(N'+\theta)(N-1)}{(1+N')(N-\theta)}$ and $\frac{N-1}{N-\theta}$. Then, we can subdivide the rest of the skills, i.e. between $\frac{N-1}{N-\theta}$ and 1, between all teachers, i.e. teachers of workers, teachers of managers, teachers of teachers of workers, etc.. Let

$$a_0 = \frac{\theta(N-1)}{N-\theta}, a_1 = \frac{(N'+\theta)(N-1)}{(1+N')(N-\theta)}, a_2 = \frac{N-1}{N-\theta}.$$

The skills of workers are between a_0 and a_1, the skills of managers are between a_1 and a_2 and we'll keep denoting the "boundary skills" by a_n, where the skills of teachers of individuals that will eventually become workers are between a_{2n} and a_{2n+1} where n represents the number of student descendent our teacher has, and the skills of teachers of individuals that will eventually become managers are between a_{2n+1} and a_{2n+2} where n represents the number of student descendent our teacher has. The boundary skills can be obtained from a_0, a_1, a_2 as follow:

$$a_{2n-1} = \frac{1}{N^{n-2}}\left(\sum_{i=0}^{n-2} N^i\right) a_2 + \frac{a_1}{N^{n-1}} - \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-2} N^i\right) a_0 \tag{3.41}$$

$$a_{2n} = \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-1} N^i\right) a_2 - \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-2} N^i\right) a_0 \tag{3.42}$$

for $n \geq 2$. We prove these formulas by induction.

Proof of (3.41)–(3.42). For the base cases a_3 and a_4 ($n = 2$) we have:

$$\begin{aligned} a_3 &= \text{right side of the managers interval} + \text{quantity of teacher of workers} \\ &= a_2 + \frac{a_1 - a_0}{N} \\ &= \frac{1}{N^0}\left(\sum_{i=0}^{0} N^i\right) a_2 + \frac{a_1}{N^1} - \frac{1}{N^1}\left(\sum_{i=0}^{0} N^i\right) a_0 \qquad \text{and} \\ a_4 &= \text{right side of the teachers of workers interval} + \text{quantity of teachers of managers} \\ &= a_3 + \frac{a_2 - a_1}{N} \\ &= a_2 + \frac{a_1}{N} - \frac{a_0}{N} + \frac{a_2}{N} - \frac{a_1}{N} \\ &= \frac{1+N}{N} a_2 - \frac{a_0}{N} \\ &= \frac{1}{N^1}\left(\sum_{i=0}^{1} N^i\right) a_2 - \frac{1}{N^1}\left(\sum_{i=0}^{0} N^i\right) a_0 \end{aligned}$$

Fix n. Suppose the formulas are true for all $m < 2n$. Then

$$\begin{aligned} a_{2n+1} =& a_{2n} + \frac{1}{N}\left(a_{2n-1} - a_{2n-2}\right) \\ =& \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-1} N^i\right) a_2 - \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-2} N^i\right) a_0 + \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-2} N^i\right) a_2 + \frac{a_1}{N^n} \\ & - \frac{1}{N^n}\left(\sum_{i=0}^{n-2} N^i\right) a_0 - \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-2} N^i\right) a_2 + \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-3} N^i\right) a_0 \\ =& \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-1} N^i\right) a_2 + \frac{a_1}{N^n} - \frac{1}{N^n}\left(\sum_{i=0}^{n-1} N^i\right) a_0 \end{aligned}$$

and

$$\begin{aligned}
a_{2n+2} =& a_{2n+1} + \frac{1}{N}\left(a_{2n} - a_{2n-1}\right) \\
=& \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-1} N^i\right) a_2 + \frac{a_1}{N^n} - \frac{1}{N^n}\left(\sum_{i=0}^{n-1} N^i\right) a_0 + \frac{1}{N^n}\left(\sum_{i=0}^{n-1} N^i\right) a_2 \\
& - \frac{1}{N^n}\left(\sum_{i=0}^{n-2} N^i\right) a_0 - \frac{1}{N^{n-1}}\left(\sum_{i=0}^{n-2} N^i\right) a_2 - \frac{a_1}{N^n} + \frac{1}{N^n}\left(\sum_{i=0}^{n-2} N^i\right) a_0 \\
=& \frac{1}{N^n}\left(\sum_{i=0}^{n} N^i\right) a_2 - \frac{1}{N^n}\left(\sum_{i=0}^{n-1} N^i\right) a_0.
\end{aligned}$$

□

Similarly, the skills of students will be separated between skills of future workers, of future managers, future teachers of workers, teachers of managers, ..., with the "boundary skills":

$$s_i = \frac{N-\theta}{N(1-\theta)} a_i + \frac{\theta(1-N)}{N(1-\theta)}.$$

Matching worker/manager and student/teacher

The relationship between the ability of a student s and the skill of his or her teacher t is given by the affine transformation between the support of students' abilities and the support of teachers' skills:

$$s = \frac{N-\theta}{N(1-\theta)} t + \frac{\theta(1-N)}{N(1-\theta)} \qquad \text{and} \qquad t = \frac{N(1-\theta)}{N-\theta} s - \frac{\theta(1-N)}{N-\theta}.$$

Similarly, the relationship between the skill of a worker w and the skill of his or her manager m is given by the affine transformation between the support of workers' skills and the support of managers' skills:

$$w = \frac{a_1 - a_0}{a_2 - a_1} m - \frac{a_2 a_0 - a_1^2}{a_2 - a_1} \qquad \text{and} \qquad m = \frac{a_2 - a_1}{a_1 - a_0} w - \frac{a_2 a_0 - a_1^2}{a_1 - a_0}.$$

Wage Functions for Workers and Managers

We find the wage functions for workers and managers. To do so, we assume $b_E(x) = b_L(x) = e^x$. We have that

$$v_m(k) = N' \sup_{w \in K} \exp(z_L(w,k)) - v_w(w)$$

so by the first order necessary condition, for an interior maximum we have

$$\begin{aligned}
& \left.\frac{\partial}{\partial w}\right|_{w=\frac{a_1-a_0}{a_2-a_1}m - \frac{a_2 a_0 - a_1^2}{a_2-a_1}} \exp(z_L(w,k)) - v_w(w) = 0 \\
\iff & v_w'(w) = (1-\theta') \exp\left((1-\theta')w + \theta'\left(\frac{a_2-a_1}{a_1-a_0} w - \frac{a_2 a_0 - a_1^2}{a_1 - a_0}\right)\right) \\
\iff & v_w(w) = \frac{1-\theta'}{1-\theta'+\theta'\frac{a_2-a_1}{a_1-a_0}} \exp\left((1-\theta')w + \theta'\left(\frac{a_2-a_1}{a_1-a_0} w - \frac{a_2 a_0 - a_1^2}{a_1 - a_0}\right)\right) + c_w
\end{aligned}$$

and as

$$v_w(k) = \sup_{m \in K} \exp(z_L(k,m)) - \frac{v_m(m)}{N'},$$

by the first order necessary condition, we have:

$$\begin{aligned} &\frac{\partial}{\partial m}\bigg|_{m=\frac{a_2-a_1}{a_1-a_0}w-\frac{a_2a_0-a_1^2}{a_1-a_0}} \exp(z_L(k,m)) - \frac{v_m(m)}{N'} = 0 \\ \iff & v_m'(m) = \theta' N' \exp\left((1-\theta')\frac{a_1-a_0}{a_2-a_1}m - \frac{a_2a_0-a_1^2}{a_2-a_1} + \theta' m\right) \\ \iff & v_m(m) = \frac{\theta' N'}{(1-\theta')\frac{a_1-a_0}{a_2-a_1}+\theta'} \exp\left((1-\theta')\frac{a_1-a_0}{a_2-a_1}m - \frac{a_2a_0-a_1^2}{a_2-a_1} + \theta' m\right) + c_m. \end{aligned}$$

To find c_w and c_m, we use the fact that

$$v_w\left(a_1\right) = v_m\left(a_1\right)$$

and the fact that when $m(w) = \frac{a_2-a_1}{a_1-a_0}w - \frac{a_2a_0-a_1^2}{a_1-a_0}$.

$$v_w(w) + \frac{1}{N'}v_m\left(m(w)\right) = \exp\left((1-\theta')w + \theta' m(w)\right).$$

Wage Functions for Teachers

We get that

$$v_k'(t) = cN\theta \exp((1-\theta)s(t) + \theta t) + N\theta\tilde{v}'((1-\theta)s(t)+\theta t)$$

where $\tilde{v}' = v_{k-2}'$. For that, we get $v_k(t) = \int v_k'(t) + c_k$. We'll write v_2 for the wage function of teachers of workers, v_3 for the wage function of teachers of managers, v_4 for the wage function of teachers of teachers of workers, etc. and we'll denote $v_0 = v_m$ and $v_1 = v_w$. The function v_k is defined on $[a_k, a_{k+1}]$. Using the fact that v_k takes the same value as the previous wage function on its left boundary a_k ($v_k(a_k) = v_{k-1}(a_k)$), we obtain the values of c_k from the constant for the wage of manager.

The benefit from education function

We have that

$$v_{k+2}(k) = N \sup_{s \in A} cb_E(z_E(s,k)) + v_k(z_E(s,k)) - u_k(s)$$

and so we have that

$$u_k'(s) = cN(1-\theta)\exp\left((1-\theta)s + \theta t(s)\right) + (1-\theta)Nv_k\left((1-\theta)s + \theta t(s)\right)$$

Using the fact that

$$v_{k+2}(k) = Ncb_E(z_E(s,k)) + v_k(z_E(s,k)) - u_k(s)$$

when $s = t(s)$, we can get the constant of integration.

Derivative of the Wage Function Near $\bar{k}$

We have that

$$v'_{t_{2n}}(a_{2n}) = (N\theta)^{n+1}\, v'_w(a_{-2}) + cN\theta \sum_{i=0}^{n} (N\theta)^i \exp(a_{2i-2})$$
$$v'_{t_{2n+1}}(a_{2n+1}) = (N\theta)^{n+1}\, v'_m(a_{-1}) + cN\theta \sum_{i=0}^{n} (N\theta)^i \exp(a_{2i-1})$$

which reinforced the belief that, in this specific case (i.e. if α is uniform and skills of teachers are above skills of managers which are above skills of workers), when the derivative of the wage function approaches $\bar{k} = 1 = a_\infty$, it is bounded if $N\theta < 1$ and unbounded if $N\theta > 1$ as predicted in [13, Theorem 16]. In fact, the leading order rates of convergence predicted there are attained precisely, in spite of the fact that one of the hypotheses upon which these predictions were based is violated, because the wage functions found here have derivatives whose discontinuities accumulate at $\bar{k}$.

We now give an example for the case $N\theta > 0$. Let $N = N' = 5$, $\theta = \theta' = \frac{1}{2}$, $c = 4$ and $b_E(x) = b_L(x) = e^x$. We get

$$v'_t\,(1-\delta) \sim \frac{1}{\delta^{1-\frac{\log 2}{\log 5}}}.$$

For the case $N\theta < 1$, let $N = 2$, $N' = 5$, $\theta = \theta' = \frac{1}{3}$, $c = 12$ and $b_E(x) = b_L(x) = e^x$. We get

$$\begin{aligned} v'_t(1-\delta) &\sim 24e\left(1 - \left(\frac{2}{3}\right)^{2-\frac{\log 5}{\log 2}} \delta^{\frac{\log 3}{\log 2}-1}\right) \\ &\sim 24e. \end{aligned}$$

3.7.2 Explicit Examples

In [13, Theorem 16], the authors prove that, if the wage function is continuously differentiable at $\bar{k}$ and $N\theta \neq 1$, the derivative of the wage function v is bounded at $\bar{k}$ if and only if $N\theta < 1$. We'll show two explicit examples, one where $N\theta > 1$ and one where $N\theta < 1$. For both of our examples, the derivative of the wage function v is not continuous at $\bar{k}$, but it still behaves as predicted by [13, Theorem 16]'s result.

We show the different functions u, v, and v' in figures 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6. In all of these pictures workers, future workers, teachers of individuals that will eventually teach workers and future teachers of individuals that will eventually teach workers are in blue. Managers, future managers, teachers of students that will eventually become teachers and future teachers of individuals that will eventually teach managers are in red.

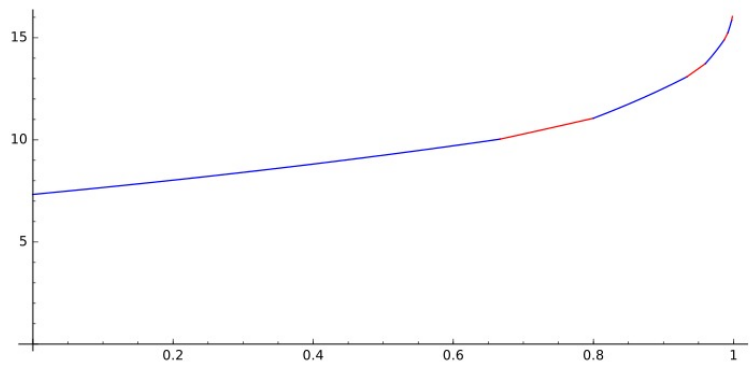

Figure 3.1: The benefit from education $u : A \to \mathbf{R}$, with $N = N' = 5$, $\theta = \theta' = \frac{1}{2}$, $c = 4$, and $b_E(x) = b_L(x) = e^x$.

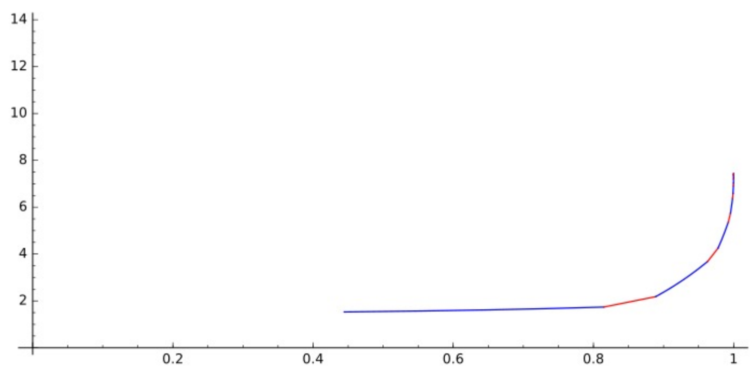

Figure 3.2: The wage function $v : K \to \mathbf{R}$, with $N = N' = 5$, $\theta = \theta' = \frac{1}{2}$, $c = 4$, and $b_E(x) = b_L(x) = e^x$.

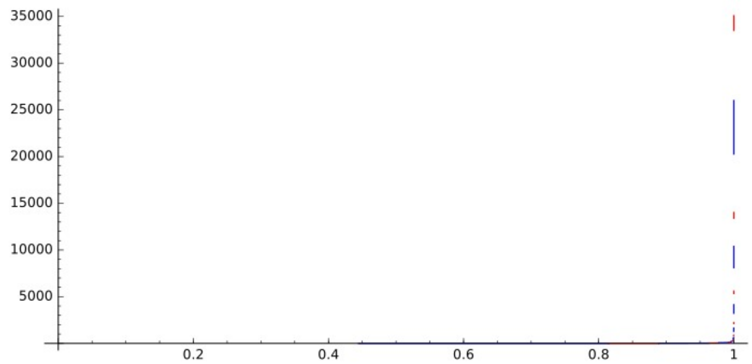

Figure 3.3: The derivative of the wage function $v : K \to \mathbf{R}$, with $N = N' = 5$, $\theta = \theta' = \frac{1}{2}$, $c = 4$, and $b_E(x) = b_L(x) = e^x$.

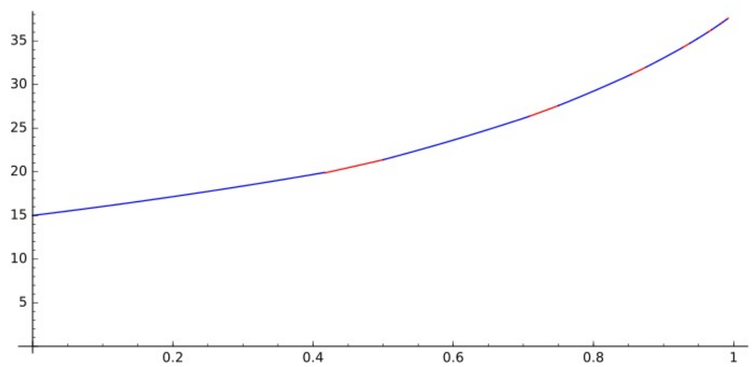

Figure 3.4: The benefit from education $u : A \to \mathbf{R}$, with $N = 2$, $N' = 5$, $\theta = \theta' = \frac{1}{3}$, $c = 12$, and $b_E(x) = b_L(x) = e^x$.

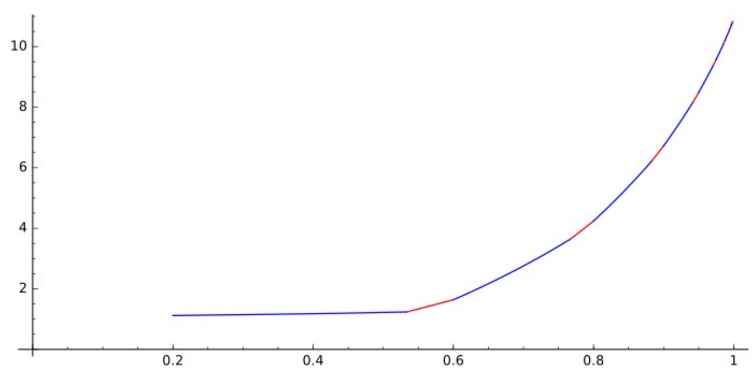

Figure 3.5: The wage function $v : K \to \mathbf{R}$, with $N = 2$, $N' = 5$, $\theta = \theta' = \frac{1}{3}$, $c = 12$, and $b_E(x) = b_L(x) = e^x$.

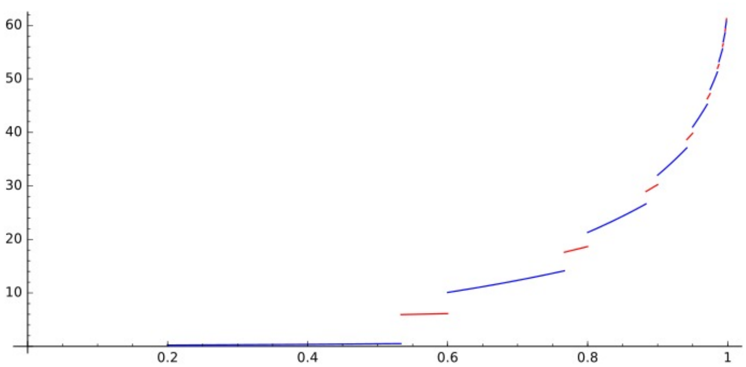

Figure 3.6: The derivative of the wage function $v : K \to \mathbf{R}$, with $N = 2$, $N' = 5$, $\theta = \theta' = \frac{1}{3}$, $c = 12$, and $b_E(x) = b_L(x) = e^x$.

Chapter 4

Analysis of a multi-sector matching problem with communication and cognitive skills

In this chapter, we study a model that couples an education and a labour market, similar to the one from Section 2.6. The novelty of the model in this chapter is that every individual is assumed to have two skills: a communication and a cognitive skill. This model was first proposed by McCann, Shi, Siow and Wolthoff in [27]. In [27], the authors explain the importance of including a non-cognitive skill (here we call it a communication skill) when we study labour and education market. Indeed, schools use non-cognitive factors (like volunteer activities and leadership positions) in their admission decisions and those non-cognitive factors are empirically correlated with professional success.

In the paper where the model is introduced [27], the authors explain the motivation for the choice of parameters and the interest of studying a model of individuals with two dimensional skills. They identify conditions to ensure that the teams in the labour market of their fictitious society are composed of one manager and several workers (and then their model forces this relationship). They study the relationship between the skills of a student and the cognitive skill of his or her teacher, and between the cognitive skill of a worker and the cognitive skill of his or her manager. In this chapter, we provide rigorous proofs for some of their observations. They also run simulations to illustrate properties of solutions.

Erlinger, McCann, Shi, Siow and Wolthoff [13] studied the analogous model in one dimension; that is, with only a cognitive skill. In this chapter, we adapt some of their results to the two dimensional case of the model. In particular, we analyze the structure of optimal solutions, we prove that there is no duality gap and that solutions exist under mild assumptions. We also discuss properties of optimal matching. The fact that there is no duality gap can be interpreted, in economics terms as the existence of competitive equilibrium.

As shown in the study of the one dimensional version of the model [13], an adult's wage is convex in his or her cognitive skill. With the assumptions of this model, for some adults, their wage will depend linearly on their communication skill, while for others, their wage is independent of their communication skill. Thus, we can't extend all uniqueness results. This implies that the properties of optimal matching when individuals' skills are assumed to be two dimensional are different from the one dimensional case.

4.1 Description of the problem

We use the modelization from [27, Section 2 and 3]. The model we use is slightly more general than the one in [27], because we include some more general notation from [13]. This is the same model as in Section 2.6, but with a communication skill.

Suppose each student has a cognitive ability $a \in [0, \bar{a}) \subset \mathbf{R}$ and each adult has a cognitive skill $k \in [0, \bar{k}) \subset \mathbf{R}$. Suppose without loss of generality that $\bar{a} = \bar{k}$. The adult cognitive skill of an individual is determined through his or her cognitive ability as student and the cognitive skill of his or her teacher when he or she was a student.

Now suppose each individual also has a communication skill $\eta \in N = (\underline{\eta}, \overline{\eta}) \subset [1, \infty)$ which is fixed for each individual for life, i.e. can't be improved by education. The skill spaces for abilities of students and skills of adults will be $A = K = N \times [0, \bar{k})$ respectively. We'll denote the cognitive ability of students by a and their communication skill of students by ι. We'll denote the cognitive skill k for generic adults and w for workers, m for managers and t for teachers and the communication skill of adults η. When there is more then one adult involve, we'll subscript η by the letter associated to the profession of the adult.

Let $\theta_E, \theta_L \in (0, 1)$. The parameter θ_E represents the role of the teacher's cognitive skill and $1 - \theta_E$ represents the role of the student's cognitive ability in the improvement of a student's future cognitive skill. The parameter θ_L represents the role of the manager's cognitive skill and $1 - \theta_L$ represents the role of a worker's cognitive skill in the outcome of a team.

Let,

$$\begin{aligned} z_E &: (\iota, a; \eta, t) \in \bar{A} \times \bar{K} && \mapsto (\iota, (1-\theta_E)a + \theta_E t) \in \bar{K} \\ z_E^1 &: (\iota, a; \eta, t) \in \bar{A} \times \bar{K} && \mapsto (1-\theta_E)a + \theta_E t \in [0, \bar{k}] \\ z_L &: (\eta_w, w; \eta_m, m) \in \bar{K} \times \bar{K} && \mapsto (1-\theta_L)w + \theta_L m \in \mathbf{R}. \end{aligned}$$

We define $z_E(\iota, a; \eta, t)$ to be the communication and cognitive ability of an adult that was a student of type (ι, a) and studied with a professor of type (η, t). The function z_E^1 gives this adult's cognitive skill. Similarly, z_L gives the efficiency attained by a collaboration between a given worker of type (η_w, w) and a manager of type (η_m, m).

With these functions, we can estimate the productivity of a student of type (ι, a) with a teacher of type (η, t) to be $cb_E(z_E(\iota, a; \eta, t))$ where $c \geq 0$ is fixed (this is the general benefit from education), and the productivity of a worker of type (η_w, w) with manager of type (η_m, m) to be $b_L(z_L(\eta_w, w; \eta_m, m))$. We assume b_E and b_L are both twice differentiable (in both variables in the case of b_E) and satisfy positive lower bounds:

$$\begin{aligned} 0 &< \min\{\, \underline{b}_L = b_L(0), \underline{b}'_L = b'_L(0), \underline{b}''_L = \inf_{k\in[0,\bar{K}]} b''_L(k) \,\} \\ 0 &< \underline{b}_E = b_E(\underline{\eta}, 0) \\ 0 &< \min\{\, \inf_{\eta,k} \tfrac{\partial}{\partial \eta} b_E(\eta,k) = \tfrac{\partial}{\partial \eta}\big|_{(\eta,k)=(\underline{\eta},0)} b_E(\eta,k), \inf_{\eta,k} \tfrac{\partial}{\partial k} b_E(\eta,k) = \tfrac{\partial}{\partial k}\big|_{(\eta,k)=(\underline{\eta},0)} b_E(\eta,k) \,\} \\ 0 &< \underline{b}'_E = \min\{\, \tfrac{\partial}{\partial \eta}\big|_{(\eta,k)=(\underline{\eta},0)} b_E(\eta,k), \tfrac{\partial}{\partial k}\big|_{(\eta,k)=(\underline{\eta},0)} b_E(\eta,k) \,\} \\ 0 &< \underline{b}''_E = \min\{\, \inf_{\eta,k} \tfrac{\partial^2}{\partial \eta^2} b_E(\eta,k), \inf_{\eta,k} \tfrac{\partial^2}{\partial k^2} b_E(\eta,k), \inf_{\eta,k} \tfrac{\partial^2}{\partial \eta \partial k} b_E(\eta,k) \,\}. \end{aligned} \tag{4.1}$$

We also assume that $b_E : \bar{A} \to \mathbf{R}$, $b_L : [0, \bar{k}] \to \mathbf{R}$, are both bounded. Let $\bar{b}_E = b_E(\bar{\eta}, \bar{k})$, $\bar{b}_L = b_L(\bar{k})$ be their upper bound.

Fix $\rho_E \geq 1$ and $\rho_L \geq 1$. Suppose that a teacher of type (η, t) can teach $\rho_E \eta \geq 1$ students and a manager of type (η, m) can manage $\rho_L\eta \geq 1$ workers. We don't need to assume $\rho_L\eta$ or $\rho_E\eta$ to be integers, because a student (resp. worker) can have more then one teacher (resp. manager).

We will suppose that the distribution of students' ability, $\alpha \in \mathcal{P}(\bar{A})$, is known and is a Borel probability measure.

4.1.1 Primal formulation

Our problem is to maximize the production of the society:

$$LP^* := \max_{\epsilon,\lambda} \left(\int_{\bar{A}\times\bar{K}} cb_E \left(z_E(\iota, a; \eta, t) \right) \epsilon(d\iota, da; d\eta, dt) \right.$$
$$\left. + \int_{\bar{K}\times\bar{K}} b_L \left(z_L(\eta_w, w; \eta_m, m) \right) \lambda(d\eta_w, dw; d\eta_m, dm) \right), \quad (4.2)$$

subject to:

$$\pi_{1\#}\epsilon = \alpha \quad \text{the distribution of students' ability is known} \quad (4.3)$$
$$\kappa := \kappa_w + \kappa_m + \kappa_t = z_{E\#}\epsilon \quad \text{the measure for adults' skill is coherent} \quad (4.4)$$
$$\epsilon, \lambda \geq 0,$$

where $\kappa_{w/m/t}$ is the distribution of workers/managers/teachers:

$$\kappa_w(\eta, k) = \left(\pi_{1\#}\lambda\right)(\eta, k) = \int_{\bar{K}} \lambda\left(\eta, k; d\eta_m, dm\right);$$
$$\kappa_m(\eta, k) = \frac{1}{\rho_L\eta}\pi_{2\#}\lambda\left(\eta, k\right) = \frac{1}{\rho_L\eta}\int_{\bar{K}} \lambda\left(d\eta_w, dw; \eta, k\right); \text{ and}$$
$$\kappa_t(\eta, k) = \frac{1}{\rho_E\eta}\pi_{2\#}\epsilon\left(\eta, k\right) = \frac{1}{\rho_E\eta}\int_{\bar{K}} \epsilon\left(d\iota, da; \eta, k\right).$$

More explicitly the constraints are $\epsilon \geq 0$, $\lambda \geq 0$,

$$\int_{\bar{K}} \epsilon(\iota, a; d\eta, dt) = \alpha(\iota, a) \quad \text{and} \quad (4.3')$$
$$\int_{\bar{K}} z_E(\eta, k; \eta_t, t)\epsilon(\eta, k; d\eta_t, dt) = \int_{\bar{K}} \lambda(\eta, k; d\eta_m, dm) \quad (4.4')$$
$$+ \int_{\bar{K}} \frac{1}{\rho_L\eta}\lambda(d\eta_w, dw; \eta, k) + \int_{\bar{A}} \frac{1}{\rho_E\eta}\epsilon(d\iota, da; \eta, k).$$

Let G_0 be the set of pairs of measures (ϵ, λ) satisfying the constraints.

4.1.2 Dual formulation

Let $u : \bar{K} \to \mathbf{R}$ and $v : \bar{K} \to \mathbf{R}$ be the Lagrange multiplier associated with (4.3) and (4.4) respectively. The Lagrangian function is:

$$\begin{aligned}\mathcal{L}(\epsilon,\lambda;u,v) = &\int_{\bar{K}\times\bar{K}} \left(b_L\left(z_L(\eta_w,w;\eta_m,m)\right) - v(\eta_w,w) - \frac{1}{\rho_L\eta_m}v(\eta_m,m)\right)\lambda(d\eta_w,dw;d\eta_m,dm)\\ &+\int_{\bar{A}\times\bar{K}}\left(cb_E\left(z_E(\iota,a;\eta,t)\right) - \frac{1}{\rho_E\eta}v(\eta,t) - u(\iota,a) + v\left(z_E(\iota,a;\eta,t)\right)\right)\epsilon(d\iota,da;d\eta,dt)\\ &+\int_{\bar{A}} u(\iota,a)\alpha(d\iota,da).\end{aligned}$$

Formally, interchanging the order of sup and inf should yield the dual problem

$$\begin{aligned}\inf_{u,v}\sup_{\epsilon,\lambda\geq 0}&\int_{\bar{K}\times\bar{K}} \left(b_L\left(z_L(\eta_w,w;\eta_m,m)\right) - v(\eta_w,w) - \frac{1}{\rho_L\eta_m}v(\eta_m,m)\right)\lambda(d\eta_w,dw;d\eta_m,dm)\\ &+\int_{\bar{A}\times\bar{K}}\left(cb_E\left(z_E(\iota,a;\eta,t)\right) - \frac{1}{\rho_E\eta}v(\eta,t) - u(\iota,a) + v\left(z_E(\iota,a;\eta,t)\right)\right)\epsilon(d\iota,da;d\eta,dt)\\ &+\int_{\bar{A}} u(\iota,a)\alpha(d\iota,da).\end{aligned}$$

If the function integrated over λ or ϵ is positive, the supremum is unbounded. So in the optimal case,

$$b_L\left(z_L(\eta_w,w;\eta_m,m)\right) - v(\eta_w,w) - \frac{1}{\rho_L\eta_m}v(\eta_m,m) \leq 0,$$

$$cb_E\left(z_E(\iota,a;\eta,t)\right) - \frac{1}{\rho_E\eta}v(\eta,t) - u(\iota,a) + v\left(z_E(\iota,a;\eta,t)\right) \leq 0$$

and their integrals with respect to λ and ϵ vanish.

Thus, we get the following problem:

$$LP_* := \inf_{u,v}\int_{\bar{A}} u(\iota,a)\alpha(d\iota,da), \tag{4.5}$$

subject to the stability constraint

$$v(\eta_w,w) + \frac{1}{\rho_L\eta_m}v(\eta_m,m) \geq b_L\left(z_L(\eta_w,w;\eta_m,m)\right) \tag{4.6}$$

for the labour market and

$$\frac{1}{\rho_E\eta}v(\eta,t) + u(\iota,a) \geq cb_E\left(z_E(\iota,a;\eta,t)\right) + v\left(z_E(\iota,a;\eta,t)\right) \tag{4.7}$$

for the education sector.

The function $u : (\eta,a) \in A \to [0,\infty)$ represents the gain of a student type (η,a) from education and $v : (\eta,k) \in K \to [0,\infty)$ is the wage of an adult of type (η,k). Let F_0 be the set of functions $(u,v) = (u_0+u_1, v_0+v_1)$ satisfying the constraints which differ from bounded continuous functions $u_0, v_0 \in \mathcal{C}(\bar{A})$ by functions $u_1, v_1 : \bar{A} \to [0,\infty]$ that are non-decreasing in the second parameter.

Note 4.1.1. *Note that for $(u,v) \in F_0$, their lower semi-continuous extensions $(lsc\ u, lsc\ v)$ (see definition 2.1.3) also belong to F_0, and the objective value at $(lsc\ u, lsc\ v)$ will be lower than the objective value at (u,v). We therefore assume u and v are lower semi-continuous from now on.*

4.2 Structure of optimal solutions

In this section, we give a general analysis of how solutions to our problems (4.2) and (4.5) behave. We study how variation in skills impacts the profession or future profession of an individual and how it impacts the matching patterns.

Let

$$\begin{aligned} v_w(\eta,k) &:= \sup_{(\eta_m,m)\in\bar{K}} b_L(z_L(\eta,k;\eta_m,m)) - \frac{1}{\rho_L\eta_m} v(\eta_m,m) && (4.8)\\ &:= \tilde{v}_w(k),\\ v_m(\eta,k) &:= \rho_L\eta\left(\sup_{(\eta_w,w)\in\bar{K}} b_L\left(z_L(\eta_w,w;\eta,k)\right) - v(\eta_w,w)\right) && (4.9)\\ &= \rho_L\eta\tilde{v}_m(k), \text{ and}\\ v_t(\eta,k) &:= \rho_E\eta\left(\sup_{(\iota,a)\in\bar{A}} cb_E\left(z_E(\iota,a;\eta,k)\right) + v(z_E(\iota,a;\eta,k)) - u(\iota,a)\right) && (4.10)\\ &= \rho_E\eta\tilde{v}_t(k). \end{aligned}$$

be the wage function of workers, managers and teachers, respectively. For the teacher's wage, we will assume $\infty - \infty = \infty$ by convention. As z_L and z_E^1 don't depend on their first and third variables, workers' wage function doesn't depend on his or her communication skill and a manager or teacher's wage is linear in its communication skill. This represents the fact that $\tilde{v}_m(k)$ (resp. $\tilde{v}_t(k)$) is the wage of a manager (resp. teacher) per worker (resp. student) and their real salary is that wage per worker (resp. student) multiplied by the number of workers (resp. students) they have.

If v is proper (see definition 2.1.1) and lower semi-continuous (see definition 2.1.2), the supremum will be attained in (4.8) and (4.9). If v is convex and non decreasing in both variables and u is proper, the supremum will be attained in (4.10).

Let

$$\bar{v} := \max\{v_w, v_m, v_t\}.$$

By (4.6), feasible wage functions, v, satisfy $v \geq v_w$, and $v \geq v_m$ and by (4.7), $v \geq v_t$ so $v \geq \bar{v}$.

We will eventually prove that $\bar{v}$ is the optimal wage a.e., i.e. the wage of an adult is the maximum of the wage he would get as a worker, manager or teacher.

We show that for a fixed cognitive skill level k, there exists a cutoff communication skill $\hat{\eta}(k)$ such that individuals with communication skill lower than $\hat{\eta}(k)$ will be workers and individuals with communication skill higher than $\hat{\eta}(k)$ won't be workers. First, we show the following:

Lemma 4.2.1. *Let $k \in [0,\bar{k}]$. If there exists an $\eta \in \bar{N}$ such that adults with skill (η,k) are necessarily managers i.e. $v_m(\eta,k) > \max\{v_w(\eta,k), v_t(\eta,k)\}$ then $\tilde{v}_m(k) > 0$.*

Similarly, if there exists an $\eta \in \bar{N}$ such that adults with skill (η,k) are necessarily teachers i.e. $v_t(\eta,k) > \max\{v_w(\eta,k), v_m(\eta,k)\}$ then $\tilde{v}_t(k) > 0$.

Proof. We prove the result of wage of managers, the proof is the same for wage of teachers. Setting $(\eta_w,w) = (\eta_m,m) = (\eta,k)$ in (4.6), we get that $v(\eta,k) > 0$ for all η_k. Therefore, if there exists an η_k such that (η_k,k) are managers and it follows from Theorem 4.5.5 that $v_m(\eta_k,k) = v(\eta_k,k) > 0$. The result then follow from the fact that $v_m(\eta_k,k) = \rho_L\eta_k\tilde{v}_m(k)$. □

Proposition 4.2.2 (c.f. [27, Proposition 3])**.** *For each cognitive skill $k \in [0, \bar{k})$, there exists a cutoff value $\hat{\eta}(k) \in N$ such that individuals with communication skill $\eta < \hat{\eta}(k)$ will be a worker and almost every individual with communication skill $\eta > \hat{\eta}(k)$ won't be a worker.*

Proof. Fix k. Suppose for all η, every individual with skills (η, k) is a worker, then we pick $\hat{\eta}(k) = \bar{\eta}$.

Suppose there exists an η such that individuals with skill (η, k) can be managers. By the envelope theorem and (4.8), as we assumed b_L to be twice differentiable, $\frac{\partial}{\partial \eta} v_w(\eta, k) = 0$. Using (4.9), we get $\frac{\partial}{\partial \eta} v_m(\eta, k) = -\rho_L \tilde{v}_m(k)$. Putting together those two derivatives and the fact that $\rho_L > 0$ and $\tilde{v}_m(k)$ is non-negative, we get

$$\frac{\partial}{\partial \eta} \left(v_w(\eta, k) - v_m(\eta, k) \right) = -\rho_L \tilde{v}_m(k) \leq 0.$$

Therefore $v_w(\eta, k) - v_m(\eta, k)$ is a decreasing function of η, so there exists $\eta_m(k)$ such that

$$\big(v_w(\eta, k) - v_m(\eta, k)\big) \left(\eta_m(k) - \eta\right) \geq 0.$$

with equality if and only if $\eta = \eta_m(k)$.

Now, if for all η every individual with skills (η, k) is a worker or a manager, we pick $\hat{\eta}(k) = \eta_m(k)$.

Suppose there exists an η such that individuals with skill (η, k) can be teachers. We can mimic the previous argument to get an $\eta_t(k)$ such that

$$\big(v_w(\eta, k) - v_t(\eta, k)\big) \left(\eta_t(k) - \eta\right) \geq 0.$$

with equality if and only if $\eta = \eta_t(k)$. If there is no individual with cognitive skill k that are manager, we pick $\hat{\eta}(k) = \eta_t(k)$. Otherwise, $\hat{\eta}(k) = \min\{\eta_m(k), \eta_t(k)\}$. □

By definition, v_w, v_m and v_t are suprema so Lemma 2.1.6 allows us to find pointwise a.e. bound for their derivatives. The next two lemmas use those bounds to prove that the optimums u and v will be convex and non-decreasing in both variables and supermodular.

Lemma 4.2.3 (Structure of the benefit from education function)**.** *Let $v : K \to \mathbf{R}$ be convex and non-decreasing in both variables with $v(\eta, \bar{k}) \geq \limsup_{k \to \bar{k}} v(\eta, k)$, Then the student payoff u defined by*

$$u(\eta, a) = \sup_{(\eta_t, t) \in \bar{K}} c b_E(z_E(\eta, a; \eta_t, t)) + v(z_E(\eta, a; \eta_t, t)) - \frac{1}{\rho_E \eta_t} v(\eta_t, t)$$

is also convex and non-decreasing in its second variable on K, supermodular and satisfies

$$\begin{aligned}
\frac{1}{1-\theta_E} \frac{\partial u(\eta, a)}{\partial a} &\geq c\underline{b}'_E + \inf_{\eta, k} \frac{\partial v(\eta, k)}{\partial k}, \\
\frac{1}{(1-\theta_E)^2} \frac{\partial^2 u(\eta, a)}{\partial a^2} &\geq c\underline{b}''_E + \inf_{\eta, k} \frac{\partial^2 v(\eta, k)}{\partial k^2}, \\
\frac{\partial u(\eta, a)}{\partial \eta} &\geq c\underline{b}'_E + \inf_{\eta, k} \frac{\partial v(\eta, k)}{\partial \eta}, \\
\frac{\partial^2 u(\eta, a)}{\partial \eta^2} &\geq c\underline{b}''_E, \qquad \textit{and} \\
\frac{\partial^2 u(\eta, a)}{\partial \eta \partial a} &\geq c\underline{b}''_E + \inf_{\eta, k} \frac{\partial^2 v(\eta, k)}{\partial \eta \partial k}
\end{aligned}$$

pointwise a.e.

Proof. The proof mimics the proof of [13, Lemma 5, second claim]. Because v is convex in k and convex and piecewise linear in η, it is twice differentiable a.e. on K. Therefore, we first assume $v \in C^2(\bar{K})$. Let

$$f(\iota, a; \eta, t) = cb_E(z_E(\iota, a; \eta, t)) + v(z_E(\iota, a; \eta, t)).$$

For each fixed ι (resp. a), and fixed η, t, f is convex and non-decreasing as a function of a (resp ι). As $\frac{\partial z_E^1}{\partial a} = 1 - \theta_E$, we have,

$$f_a(\iota, a; \eta, t) = (1 - \theta_E)\left(c\frac{\partial}{\partial k}\bigg|_{k=z_E^1(\iota,a;\eta,t)} b_E(\iota, k) + \frac{\partial}{\partial k}\bigg|_{k=z_E^1(\iota,a;\eta,t)} v(\iota, k)\right)$$

So

$$\frac{f_a(\iota, a; \eta, t)}{1 - \theta_E} \geq c\underline{b}'_E + \inf \frac{\partial v}{\partial k}(\eta, k).$$

Similarly,

$$\frac{f_{aa}(\eta, a; \eta_t, t)}{(1 - \theta_E)^2} \geq c\underline{b}''_E + \inf \frac{\partial^2 v}{\partial k^2}(\eta, k).$$

The bounds for f_ι, $f_{\iota,\iota}$ and $f_{\iota,a}$ are found similarly. It then follows from Lemma 2.1.6 that the same bounds are true for $u(\eta, a)$.

If we don't suppose v is $C^2(K)$, we can approximate it uniformly on compact subsets of K by C^2 functions v^i satisfying the same hypotheses as v. We get that

$$f^i(\eta, a; \eta_t, t) = cb_E(z_E(\eta, a; \eta_t, t)) + v^i(z_E(\eta, a; \eta_t, t))$$

converge to f uniformly on compact subsets of $\bar{A} \times \bar{K} \setminus \{ (\bar{a}, \bar{k}) \}$ and

$$u^i(\eta, a) = \sup_{(\eta_t, t) \in \bar{K}} f^i(\eta, a; \eta_t, t) - \frac{1}{\rho_E \eta_t} v(\eta_t, t)$$

converge uniformly to u on compact subsets of A. So u inherits the same Lipschitz and local semi-convexity bounds as u^i in the distributional (and hence pointwise a.e.) sense. □

Lemma 4.2.4 (Structure of wage functions)**.** *Let $v : K \to \mathbf{R}$ be convex and non-decreasing in its second variable, and non-decreasing, convex and piecewise linear in its first variable and supermodular with $v(\eta, \bar{k}) \geq \limsup_{k \to \bar{k}} v(\eta, k)$. Then $v_{w/m/t}$ are positive, non-decreasing and convex in the second variable on the subset of $\bar{K}$ where they represent individuals. Their maximum*

$$\bar{v} := \max \{ v_w, v_m, v_t \}$$

is non-decreasing and convex in the second variable on $\bar{K}$, piecewise linear in their first variable, and *satisfy*

$$\begin{aligned}
\frac{\partial \bar{v}(\eta,k)}{\partial \eta} &\geq 0 \\
\frac{\partial^2 \bar{v}(\eta,k)}{\partial \eta^2} &= 0 \\
\frac{\partial \bar{v}(\eta,k)}{\partial k} &\geq \min\left\{(1-\theta_E)\underline{b}_L', \eta\rho_L\theta_L\underline{b}_L', \rho_E\eta_t\theta_E\left(c\underline{b}_E' + \inf_{(\eta,k)}\frac{\partial}{\partial k}v(\eta,k)\right)\right\} \\
\frac{\partial^2 \bar{v}(\eta,k)}{\partial k^2} &\geq \min\left\{(1-\theta_E)^2\underline{b}_L'', (\theta_L)^2\eta\rho_L\underline{b}_L'', \rho_E\eta_t\theta_E^2\left(c\underline{b}_E'' + \inf_{(\eta,k)}\frac{\partial^2}{\partial k^2}v(\eta,k)\right)\right\} \qquad \textit{and} \\
\frac{\partial^2 \bar{v}(\eta,k)}{\partial \eta \partial k} &\geq \min\left\{0, \rho_L\theta_L\underline{b}_L', \theta_E\rho_E\left(c\underline{b}_E' + \inf_{(\eta,k)}\frac{\partial}{\partial k}v(\eta,k)\right)\right\}
\end{aligned}$$

pointwise a.e.

Proof. To prove this claim, we follow the idea of the proof of [13, Lemma 5, third claim]; that is we will prove that, on the subset of $\bar{K}$ where they represent individuals, $v_{w/m/t}$ are piecewise linear in their first variable, non-decreasing, convex in their second variable, and supermodular, by finding lower bounds on their derivative.

Let's begin with v_t. Because v is convex in k and convex and piecewise linear in η, it is twice differentiable a.e. on K. First we suppose $v \in C^2(\bar{A})$. As in Lemma 4.2.3, we set

$$f(\iota, a; \eta_t, t) = cb_E(z_E(\iota, a; \eta_t, t)) + v(z_E(\iota, a; \eta_t, t)).$$

Then we have

$$v_t(\eta_t, t) = \rho_E\eta_t\left(\sup_{(\iota,a)\in\bar{A}} f(\iota, a; \eta_t, t) - u(\iota, a)\right).$$

Note that

$$\begin{aligned}
f_{\eta_t} &= 0 \\
f_t &= \theta_E\left(c\frac{\partial}{\partial k}b_E(\eta,k)\bigg|_{(\eta,k)=z_E(\iota,a;\eta_t,t)} + \frac{\partial v}{\partial k}\left(z_E(\iota,a;\eta_t,t)\right)\right) \\
&\geq \theta_E\left(c\underline{b}_E' + \inf_{(\eta,k)}\frac{\partial v}{\partial k}(\eta,k)\right) \\
f_{tt} &= \theta_E^2\left(c\frac{\partial^2}{\partial k^2}b_E(\eta,k)\bigg|_{(\eta,k)=z_E(\iota,a;\eta_t,t)} + \frac{\partial^2 v}{\partial k^2}\left(z_E(\iota,a;\eta_t,t)\right)\right) \\
&\geq \theta_E^2\left(c\underline{b}_E'' + \inf_{(\eta,k)}\frac{\partial^2 v}{\partial k^2}(\eta,k)\right).
\end{aligned}$$

So again by Lemma 2.1.6, we have that:

$$\begin{aligned}\frac{\partial v_t(\eta_t,t)}{\partial t} &\geq \theta_E\rho_E\eta_t\left(c\underline{b}'_E+\inf_{(\eta,k)}\frac{\partial v}{\partial k}(\eta,k)\right),\\ \frac{\partial^2 v_t(\eta_t,t)}{\partial\eta_t\partial t} &\geq \theta_E\rho_E\left(c\underline{b}'_E+\inf_{(\eta,k)}\frac{\partial v}{\partial k}(\eta,k)\right), \quad\text{and}\\ \frac{\partial^2 v_t(\eta_t,t)}{\partial t^2} &\geq \theta_E^2\rho_E\eta_t\left(c\underline{b}''_E+\inf_{(\eta,k)}\frac{\partial^2 v}{\partial k^2}(\eta,k)\right)\end{aligned}$$

a.e. Finally,

$$\begin{aligned}\frac{\partial v_t(\eta_t,t)}{\partial\eta_t} &= \rho_E\left(\sup_{(\iota,a)\in\bar{A}} f(\iota,a;\eta_t,t)-u(\iota,a)\right)\\ &= \frac{v_t(\eta_t,t)}{\eta_t}\end{aligned}$$

which is positive if we are on the subdomain of $\bar{K}$ where individual can be teachers because in that case, by Lemma 4.2.1, $v_t > 0$.

If v is not twice differenciable, we can approximate it uniformly on compact subsets of K by C^2 function v^i satisfying the same hypotheses as v. We get that

$$f^i(\iota,a;\eta,t) = cb_E\big(z_E(\iota,a;\eta_t,t)\big)+v^i\big(z_E(\iota,a;\eta_t,t)\big)$$

converge uniformly to f on compact subsets of $\bar{A}\times\bar{K}\setminus\big\{\big(\bar{a},\bar{k}\big)\big\}$ and we can define from it v_t^i that will converge uniformly to v_t on compact subsets of A. So v_t inherits the same Lipschitz and local semi-converxity bounds as v_t^i in the distributional (and hence pointwise a.e.) sense.

Now consider v_w and v_m. Let

$$f(\eta_w,w;\eta_m,m) = b_L(z_L(\eta_w,w;\eta_m,m))$$

and rewrite

$$\begin{aligned}v_w(\eta_w,w) &= \sup_{(\eta_m,m)} f(\eta_w,w;\eta_m,m)-\frac{1}{\eta_m\rho_L}v(\eta_m,m)\\ v_m(\eta_m,m) &= \eta_m\rho_L\left(\sup_{(\eta_w,w)} f(\eta_w,w;\eta_m,m)-v(\eta_w,w)\right).\end{aligned}$$

As

$$\begin{aligned} f_{\eta_w} &= 0 = f_{\eta_m} \\ f_w &= (1-\theta_L)b_L'(z_L(\eta_w, w; \eta_m, m)) \geq (1-\theta_L)\underline{b}_L' \\ f_{ww} &= (1-\theta_L)^2 b_L''(z_L(\eta_w, w; \eta_m, m)) \geq (1-\theta_L)^2\underline{b}_L'' \\ f_m &= \theta_L b_L'(z_L(\eta_w, w; \eta_m, m)) \geq \theta_L\underline{b}_L' \\ f_{mm} &= \theta_L^2 b_L''(z_L(\eta_w, w; \eta_m, m)) \geq \theta_L^2\underline{b}_L'', \end{aligned}$$

using Lemma 2.1.6, we get

$$\begin{aligned} \frac{\partial v_w}{\partial \eta}(\eta, w) &= 0 \\ \frac{\partial v_m}{\partial \eta}(\eta, m) &= \rho_L\left(\sup_{(\eta_w, w)} f(\eta_w, w; \eta, m) - v(\eta_w, w)\right) = \frac{v_m(\eta, m)}{\eta} \end{aligned}$$

which is positive if we are on the subdomain of $\bar{K}$ where individual can be managers because in that case, by Lemma 4.2.1, $v_m > 0$ and $\frac{\partial v_m}{\partial \eta}(\eta, m) > 0$. Taking more derivatives, we have:

$$\begin{aligned} \frac{\partial^2 v_m}{\partial \eta_m^2}(\eta_m, m) &= 0, & \frac{\partial v_w}{\partial w}(\eta_w, w) &\geq (1-\theta_L)\underline{b}_L', \\ \frac{\partial^2 v_w}{\partial w^2}(\eta_w, w) &\geq (1-\theta_L)^2\underline{b}_L'', & \frac{\partial v_m}{\partial m}(\eta_m, m) &\geq \theta_L\eta_m\rho_L\underline{b}_L', \\ \frac{\partial^2 v_m}{\partial \eta_m \partial m}(\eta_m, m) &\geq \theta_L\rho_L\underline{b}_L', & \frac{\partial^2 v_m}{\partial m^2}(\eta_m, m) &\geq \theta_L^2\eta_m\rho_L\underline{b}_L''. \end{aligned}$$

Thus, $\bar{v}$ is increasing and convex in both variables and supermodular and its partial derivatives satisfy the appropriate bounds.

□

Note that by being piecewise linear in η, $\bar{v}$ is also piecewise convex in η. We will prove later that $\bar{v}$ is in fact globally convex in η.

In Proposition 4.2.2 we strengthen the argument from [27], which give, for fixed cognitive skill, conditions on the communication skill to be a worker or not. The next proposition studies, for fixed communication skill, the relationship between an individual's occupation and his or her cognitive skill. Note that as we fixed communication skill, this is equivalent then studying the one dimensional problem [13]. Therefore, this proposition is equivalent to [13, Proposition 7].

Proposition 4.2.5 (Specialization by type; the educational pyramid)**.** *Suppose $u : A \to \mathbf{R}$, $v : K \to \mathbf{R}$ are convex and non-decreasing in k and satisfy $v = \max\{v_w, v_m, v_t\}$ where $v_{w/m/t}$ are defined by* (4.8), (4.9) *and* (4.10).

1. *If*

$$\rho_L\underline{\eta}\theta_L \geq (1-\theta_L)\sup_{k\in K}\frac{b_L'\left((1-\theta_L)k + \theta_L\bar{k}^-\right)}{b_L'(\theta_E k)} \tag{b}$$

then for fixed communication skill, the cognitive ability of all workers' types is weakly below the one of all managers' types.

2. *If*

$$\rho_E \underline{\eta} \theta_E \underline{b}'_E \geq \overline{b}'_L \max\{\, \rho_L \overline{\eta} \theta_L, 1 - \theta_L \,\} \tag{a}$$

then for fixed communication skill, the cognitive ability of all teacher types is weakly above the cognitive ability of all manager and worker types.

3. *Suppose*

$$\rho_E \underline{\eta} \theta_E \geq \rho_L \overline{\eta} \theta_L \sup_{0 \leq \frac{z}{\eta} \leq k} \frac{b'_L\left((1-\theta_L)z^- + \theta_L \bar{k}\right)}{\rho_L \underline{\eta} \theta_L b'_L(\theta_L z^+) + c\frac{\partial}{\partial z} b_E(\eta, z^+)} \tag{c}$$

and (b) *holds, and*

$$f(\iota, a; \eta_t, t) := u(\iota, a) + \frac{1}{\rho_E \eta_t} v(\eta_t, t) - c b_E(z_E(\iota, a; \eta_t, t) - v\left(z_E(\iota, a; \eta_t, t)\right)$$

vanishes at some $(\iota, a; \eta_t, t) \in \bar{A} \times \bar{K}$ *where* $v\left(z_E(\iota, a; \eta_t, t)\right) = v_m\left(z_E^1(\iota, a; \eta_t, t)\right)$. *Then for fixed* η_t,

$$v > v_m \text{ for } (\eta_t, k) \text{ where } k > t.$$

In words, no manager (or worker) can have a higher salary than a teacher of managers.

4. *If*

$$\underline{\eta} \rho_E \theta_E \geq 1, \tag{d}$$

then any student's cognitive ability will be weakly below his or her teacher's cognitive skill and strictly below if

$$\text{either } c > 0 \text{ or } \underline{\eta} \rho_E \theta_E > 1 \tag{e}$$

and (d) *holds.*

5. *If*

$$\text{either } c > 0 \text{ or } \frac{\partial}{\partial k} v(\cdot, 0^+) > 0 \tag{f}$$

then conditions (d) *and* (e) *imply that any teacher of type* (η, k) *will have at most finitely many academic descendants* $d(\eta, k)$, *i.e. a teacher of teachers of teachers... will eventually reach a teacher of managers or a teacher of workers which will end the teaching chain, except if* v *is not differentiable in* k *at* (η, k). *Moreover, we have*

$$\begin{aligned}
\frac{\partial}{\partial k} v_t(\eta_0, k_0) =& \rho_E \eta_0 \theta_E \left(c \frac{\partial}{\partial k} b_E(\eta_1, k_1) + \rho_E \eta_1 \theta_E \left(c \frac{\partial}{\partial k} b_E(\eta_2, k_2) + \rho_E \eta_2 \theta_E \Bigg(\cdots \right.\right. \\
& \left.\left. \cdots + \rho_E \eta_{d-1} \theta_E \left(c \frac{\partial}{\partial k} b_E(\eta_d, k_d) + \frac{\partial}{\partial k} v(\eta_d, k_d) \right) \Bigg) \right) \right) \\
\geq & \begin{cases} \frac{1 - (\rho_E \underline{\eta} \theta_E)^d}{1 - \rho_E \underline{\eta} \theta_E} \rho_E \underline{\eta} \theta_E c \frac{\partial}{\partial k} b_E(\underline{\eta}, \theta_E^d k_0) + \left(\rho_E \underline{\eta} \theta_E\right)^d \frac{\partial}{\partial k} v(\underline{\eta}, \theta_E^d k_0) & \text{if } \rho_E \underline{\eta} \theta_E \neq 1 \\ d c \frac{\partial}{\partial k} b_E(\underline{\eta}, \theta_E^d k_0) + \frac{\partial}{\partial k} v(\underline{\eta}, \theta_E^d k_0) & \text{if } \rho_E \underline{\eta} \theta_E = 1 \end{cases}
\end{aligned}$$

Note 4.2.6. *Condition* (a) *requires that the impact of teachers (which depends on* ρ_E, θ_E, β_E *and* η*)*

is greater than the impact of workers (which depends on θ_L and b_L) and than the impact of managers (which depends on ρ_L, θ_L, β_L and η).

Condition (b) *forces the best worker to have a lesser impact on the production of the labour market than the worst manager.*

Condition (c) *is a weaker condition than* (a) *which ensures that some teachers have more impact than everyone in the labour market.*

Condition (d) *ensures that a teacher teaches at least one student and condition* (e) *makes this strictly more than one student or puts the benefit from education b_E in the objective functional.*

Condition (f) *ensures that the benefit from education b_E is in the objective functional or that the wage function of adults is strictly increasing at cognitive skill* 0 *for any communication skill.*

Proof of Proposition 4.2.5. The proof mimics the proof of [13, Proposition 7]. By Lemma 4.2.4, $v_{w/m/t}$ are convex in k, hence one-sided differentiable and two-sided differentiable a.e. At points where they are differentiable, Lemma 2.1.6 allows the following estimate:

$$\frac{\partial}{\partial k}v_w(\eta,k) = (1-\theta_L)b_L'\left((1-\theta_L)k+\theta_L m\right) \tag{4.11}$$

$$\frac{\partial}{\partial k}v_m(\eta,k) = \rho_L\eta\theta_L b_L'\left((1-\theta_L)w+\theta_L k\right) \tag{4.12}$$

$$\frac{\partial}{\partial k}v_t(\eta,k) = \rho_E\eta\theta_E\left(c\frac{\partial}{\partial k}\Big|_{k=(1-\theta_E)a+\theta_E k} b_E(\iota,k) + \frac{\partial}{\partial k}\Big|_{k=(1-\theta_E)a+\theta_E k} v(\iota,k)\right)$$

where $m, w \in [0,\bar{k}]$ and $(\iota, a) \in A$ are the maximisers of (4.8), (4.9) and (4.10) respectively.

1. To show that a worker's cognitive skill lies weakly below any manager's cognitive skill, we prove that $v_m - v_w$ is strictly increasing in its second variable (cognitive skill). As this function is non-negative for all (η_m, m) managers and non-positive for all (η_w, w) workers this will show the claim.

   We have that $\frac{\partial}{\partial k}v_m(\eta,k)$ is minimal when $w = 0$ in (4.12) so its minimal value is

   $$\rho_L\eta\theta_L b_L'(\theta_L k),$$

   and $\frac{\partial}{\partial k}v_w$ is maximal when $m = \bar{k}$ in (4.11) so it maximal value is

   $$(1-\theta_L)b_L'\left((1-\theta_L)\bar{k}^- + \theta_L k\right).$$

   Now because of assumption (b) we have

   $$\begin{aligned}\frac{\partial}{\partial k}v_m(\eta,k) &\geq \rho_L\eta\theta_L b_L'(\theta_L k)\\ &\geq \rho_L\underline{\eta}\theta_L b_L'(\theta_L k)\\ &\geq (1-\theta_L)\sup_{k\in K} b_L'\left((1-\theta_L)k\theta_E;\bar{k}^-\right) && \text{by (b)}\\ &\geq \frac{\partial}{\partial k}v_w(\eta,k).\end{aligned}$$

2. We will show that

   $$\frac{\partial}{\partial k}v_t(\eta,k) \geq \max\left\{\frac{\partial}{\partial k}v_w, \frac{\partial}{\partial k}v_m\right\}.$$

It is harder to get an estimate of $\frac{\partial}{\partial k}v_t$ because of its dependence on v. If a student of cognitive skill a_1 is taught by a teacher of cognitive skill k_0, it will have cognitive skill

$$k_1 = (1-\theta_E)a_1 + \theta_E k_0$$

as an adult.

So,

$$\begin{aligned}\frac{\partial}{\partial k}v_t(\eta, k_0) &= \rho_E \eta \theta_E \left(c\frac{\partial}{\partial k_1} b_E(\eta_1, k_1) + \frac{\partial}{\partial k_1} v(\eta_1, k_1)\right) && (4.13)\\ &\geq \rho_E \underline{\eta} \theta_E c \underline{b}'_E && \text{as } v \text{ is increasing in its second variable}\\ &\geq \bar{b}'_L \max\{\rho_L \bar{\eta}\theta_L, 1-\theta_L\} && \text{by (a)}\\ &\geq \max\left\{\frac{\partial}{\partial k}v_m(\eta_t, k_0), \frac{\partial}{\partial k}v_w(\eta_t, k_0)\right\}.\end{aligned}$$

3. Assume that a teacher of type (η_t, t) teaches a student of type (ι, k) who becomes a manager of type $(\eta_m = \iota, m) = z_E(\iota, a; \eta_t, t)$.

   We have that $v \geq v_m$ with equality at (η_m, m), thus,

$$\frac{\partial}{\partial k} v\left(\eta_m, m^+\right) \geq \frac{\partial}{\partial k} v_m\left(v_m, m^+\right).$$

   Now we have:

$$\begin{aligned}\frac{\partial}{\partial k}v_t(\eta_t, t^+) &\geq \rho_E \eta_t \theta_E \left(c\frac{\partial}{\partial k} b_E\left(\eta_m, m^+\right) + \frac{\partial}{\partial k} v_m\left(\eta_m, m^+\right)\right)\\ &\geq \rho_E \underline{\eta} \theta_E \left(c\frac{\partial}{\partial k} b_E\left(\eta_m, m^+\right) + \rho_L \eta_m \theta_L b'_L\left(z_E(w, m^+)\right)\right) && \text{for a } w \in [0, \bar{k})\\ &\geq \rho_E \underline{\eta} \theta_E \left(c\frac{\partial}{\partial k} b_E\left(\eta_m, m^+\right) + \rho_L \eta_m \theta_L b'_L\left(\theta_L m^+\right)\right)\\ &\geq \rho_L \overline{\eta} \theta_L b'_L\left((1-\theta_L)m^- + \theta_L \overline{k}\right) && \text{by (c).}\end{aligned}$$

   Now (b) implies that for a fixed η_t, m is weakly above the cognitive skill of any worker. Thus,

$$\begin{aligned}\frac{\partial}{\partial k}v_m(\eta_t, t) &\leq \rho_L \eta_t \theta_L b'_L\left((1-\theta_L)m^- + \theta_L t\right)\\ &\leq \rho_L \overline{\eta} \theta_L b'_L\left((1-\theta_L)m^- + \theta_L t\right),\end{aligned}$$

   which gives us

$$\frac{\partial}{\partial k}v_t(\eta_t, t^+) \geq \frac{\partial}{\partial k}v_m(\eta_t, t)$$

   and the convexity of v_t and strict convexity of v_m (see proof of Lemma 4.2.4) imply that

$$\frac{\partial}{\partial k}v_t(\eta_t, k) \geq \frac{\partial}{\partial k}v_m(\eta_t, k)$$

   for $k > t$.

So $v_t(\eta_t,\cdot) > v_m(\eta_t,\cdot)$ on $(t,\bar{k})$.

4. By (4.13) and (d) for a student (ι,a) who will become an adult of type (ι,k) with teacher (η_t,t) we have:

$$\begin{aligned}\frac{\partial}{\partial k}v_t(\eta_t,t) &\geq c\frac{\partial}{\partial k}b_E(\iota,k)+\frac{\partial}{\partial k}v(\iota,k)\\ &\geq \frac{\partial}{\partial k}v(\eta_a,k).\end{aligned}$$

If $\underline{\eta}\rho_E\theta_E > 1$ the first inequality is strict and if $c>0$ the second inequality is strict, so we conclude that any student's cognitive ability will be below his or her teacher's cognitive skill.

5. Start with a teacher (η_0,k_0) with students type (ι_1,a_1) who will become adults of type $(\eta_1,k_1 = z^1_E(\iota_1,a_1;\eta_0,k_0))$. If a student decides to be a worker or a manager, there is nothing to prove.

If a student becomes a teacher, we'll denote (η_2,a_2) the type of his or her students. Similarly, as long as the next generation are teachers they have type $(\eta_n,k_n = z^1_E(\iota_n,a_n;\eta_{n-1},k_{n-1}))$ and their students have type (ι_{n+1},a_{n+1}).

We claim that there exists a finite d such that the students of type (ι_d,a_d) choose to become workers or managers. Suppose not to get a contradiction. Then the teachers' cognitive skills $k_{i+1}<k_i$ converge to some $k_\infty\in K$. Their communication skills η_i form a positive sequence. Then, taking the limit in (4.13) gives:

$$\lim_{i\to\infty}\frac{\partial}{\partial k}v_t\left(\eta_i,k_i\right)=\lim_{i\to\infty}\rho_E\eta_i\theta_E\left(c\frac{\partial}{\partial k}b_E\left(\eta_i,k_i\right)+\frac{\partial}{\partial k}v\left(\eta_i,k_i\right)\right).$$

Rearranging the terms, and noting that $v=v_t$ in this case, we get:

$$\lim_{i\to\infty}\left(1-\rho_E\eta_i\theta_E\right)\frac{\partial}{\partial k}v_t\left(\eta_i,k_i\right)=\lim_{i\to\infty}\rho_E\eta_i\theta_E c\frac{\partial}{\partial k}b_E\left(\eta_i,k_i\right).$$

If (d) holds, the left hand side is ≤ 0 and the right hand side is ≥ 0. Adding condition (e) and (f) gives that one of those inequalities is strict. As v_t and b_E are strictly positive, the strict inequality still holds in the limit. This is a contradiction and we conclude that the sequence (η_i,k_i) terminates at a finite d. By iterating, we get:

$$\begin{aligned}\frac{\partial}{\partial k}v_t(\eta_0,k_0) =&\rho_E\eta_0\theta_E\left(c\frac{\partial}{\partial k}b_E(\eta_1,k_1)+\rho_E\eta_1\theta_E\left(c\frac{\partial}{\partial k}b_E(\eta_2,k_2)+\rho_E\eta_2\theta_E\left(\cdots\right.\right.\right.\\ &\left.\left.\left.\cdots+\rho_E\eta_{d-1}\theta_E\left(c\frac{\partial}{\partial k}b_E(\eta_d,k_d)+\frac{\partial}{\partial k}v(\eta_d,k_d)\right)\right)\right)\right)\\ \geq&\begin{cases}\frac{1-(\rho_E\underline{\eta}\theta_E)^d}{1-\rho_E\underline{\eta}\theta_E}\rho_E\underline{\eta}\theta_E c\frac{\partial}{\partial k}b_E(\underline{\eta},\theta_E^d k_0)+\left(\rho_E\underline{\eta}\theta_E\right)^d\frac{\partial}{\partial k}v(\underline{\eta},\theta_E^d k_0) & \text{if } \rho_E\underline{\eta}\theta_E\neq 1\\ dc\frac{\partial}{\partial k}b_E(\underline{\eta},\theta_E^d k_0)+\frac{\partial}{\partial k}v(\underline{\eta},\theta_E^d k_0) & \text{if } \rho_E\underline{\eta}\theta_E=1\end{cases}\end{aligned}$$

□

4.3 Perturbed problem

In section 4.5, we'll try to find conditions for problems (4.2) and (4.5) to have (unique) solutions. To do so, we will minimize the dual problem LP_* under the additional hypothesis that v is convex and non-decreasing in both variables and supermodular, following the strategy of Erlinger et al [13]. These hypotheses and lemma 4.2.3 imply that u is strictly non-decreasing, strictly convex in both variables and strictly supermodular[1]. Having non-decreasing wage functions in both variables is a natural hypothesis as it means that an increase in a skill variable never corresponds to a decrease in wage. Supermodularity means that an increase in one variable increases the impact of the other variable. That is, the wage of an individual with higher cognitive skill is augmented more with an increase in his or her communication skill than that of an individual with lower cognitive skill.

The technical reason to assume these hypotheses is to force the feasible functions to be in a compact space, allowing us to extract limits from minimizing sequences. We then show that these additional constraints on v are non-binding for our problem.

Note that if u and v are convex in each of their variables, they can only be discontinuous by having a decreasing jump on the bottom or the left side of K $\big($on $(0,\eta)$ for $\eta \in N$ or on $(k,\underline{\eta})$ for $k \in [0,\bar{k})\big)$ or an increasing jump on the top or the right side of K $\big($on $(\bar{k},\eta)$ for $\eta \in N$ or on $(k,\bar{\eta})$ for $k \in [0,\bar{k})\big)$. If u and v are also increasing the first discontinuity is impossible and if they are lower semi-continuous, the second discontinuity is impossible.

To prove that the extra conditions stating that v is convex and non-decreasing in both variables and supermodular are non-binding, it is necessary to control our functions u, v on the full domain A, K not only on spt α and spt κ (where $\kappa = z_{E\#}\epsilon$ is the unknown distribution of adults). Our original problem doesn't control the value of u and v outside the support of α and κ. Moreover, as v is non-strictly supermodular and non-strictly convex in its first variable (in fact it is piecewise linear), we can't prove directly that it is supermodular and convex in its first variable.

To solve these problems, we introduce a perturbed version of the dual problem.

$$LP_*(\delta) := \inf_{(u,v)\in F_\delta} \delta \left\langle u+v \right\rangle_A + \int_{\bar{A}} u(\iota,a)\alpha(d\iota,da) \tag{4.14}$$

where

$$\left\langle v \right\rangle_A := \frac{1}{\mathcal{L}(A)} \int_A v d\mathcal{L}$$

denotes the Lebesgue average of v over A, and F_δ is the set of $(u,v) = (u_0+u_1, v_0+v_1)$ which differ from bounded continuous functions u_0, v_0 by functions u_1, v_1 that are non-decreasing in the second variable, such that

$$\begin{cases} v(\eta_w,w) + \dfrac{1}{\rho_L \eta_m} v(\eta_m,m) \geq b_L\left(z_L(\eta_w,w;\eta_m,m)\right) + \delta\left(\eta_w w + \eta_w^2 + \eta_m\right) & \text{(4.15a)} \\ \dfrac{1}{\rho_E \eta_t} v(\eta_t,t) + u(\iota,a) \geq c_\delta b_E\left(z_E(\iota,a;\eta_t,t)\right) + v(z_E(\iota,a;\eta_t,t)) + \delta\eta_t, & \text{(4.15b)} \end{cases}$$

similar to the non-perturbed problem[2].

[1] Note that u and v don't have to be jointly convex.

[2] Note that c_δ might depend on δ if we want. For example, $c = 0$ can be approximated with $c_\delta \to 0$, when $\delta \to 0$.

These problems admit primal problems:

$$LP^*(\delta) := \max_{(\epsilon,\lambda)\in G_\delta} \int_{\bar{A}\times\bar{K}} \left(c_\delta b_E\left(z_E(\iota, a; \eta_t, t)\right) + \delta\eta_t\right)\epsilon(d\iota, da; d\eta_t, dt) \tag{4.16}$$
$$+ \int_{\bar{K}\times\bar{K}} b_L\left(z_L(\eta_w, w; \eta_m, m)\right) + \delta\left(\eta_w w + \eta_w^2 + \eta_m\right)\lambda(d\eta_w, dw; d\eta_m, dm)$$

where G_δ is the set of pairs $(\epsilon, \lambda) \geq 0$ such that:

$$\begin{cases} \pi_{1\#}\lambda + \pi_{2\#}\left(\dfrac{\lambda}{\rho_L\eta_2}\right) + \pi_{2\#}\left(\dfrac{\epsilon}{\rho_E\eta_2}\right) = z_{E\#}\epsilon + \dfrac{\delta}{|K|}\mathcal{L}|_K & (4.17a) \\ \pi_{1\#}\epsilon = \alpha + \dfrac{\delta}{|A|}\mathcal{L}|_A. & (4.17b) \end{cases}$$

We'll use $f = f(\eta_w, w, \eta_m, m) = \eta_w w + \eta_w^2 + \eta_m$ for the function that multiplies δ in $LP^*(\delta)$.

4.4 Strong duality

We now want to show that $LP_*(\delta) = LP^*(\delta)$. We first need the following lemma:

Lemma 4.4.1 (Endogenous distribution of adult skills)**.** *Fix a Borel measure $\alpha \geq 0$ on $\bar{A}$. If $\epsilon \geq 0$ on $\bar{A}\times\bar{K}$ is such that $\pi_{1\#}\epsilon = \alpha$, then for each $\bar{\eta} - \Delta\eta \in (\underline{\eta}, \bar{\eta})$ and $\bar{k} - \Delta k \in (0, \bar{k})$, the measure $\kappa = z_{E\#}\epsilon$, where z_E is defined in section 4.1, satisfies:*

$$\int_{[\bar{\eta}-\Delta\eta,\eta]\times[\bar{k}-\Delta k,\bar{k}]} \kappa(d\eta_k, dk) \leq \int_{[\bar{\eta}-\Delta\eta,\bar{\eta}]\times[\bar{a}-\frac{1}{1-\theta_E}\Delta k,\bar{a}]} \alpha(d\eta_a, da).$$

Proof. The proof mimics the proof of [13, Lemma 14, first claim]. First note that

$$\kappa\left([\bar{\eta}-\Delta\eta, \eta] \times [\bar{k}-\Delta k, \bar{k}]\right) = \epsilon\left(z_E^{-1}\left([\bar{\eta}-\Delta\eta, \eta] \times [\bar{k}-\Delta k, \bar{k}]\right)\right).$$

We have that $(\eta_a, a; \eta_t, t)$ is in $z_E^{-1}\left([\bar{\eta}-\Delta\eta, \eta] \times [\bar{k}-\Delta k, \bar{k}]\right)$ if and only if

$$\eta_a \in [\bar{\eta}-\Delta\eta, \eta]$$

and

$$\begin{aligned} \bar{k} - \Delta k &\leq z_E^1(\eta_a, a; \eta_k, k) \\ &= (1-\theta_E)a + \theta_E k \\ &\leq (1-\theta_E)a + \theta_E \bar{k} \\ \iff a &\geq \bar{k} - \frac{1}{1-\theta_E}\Delta k. \end{aligned}$$

So

$$\kappa\left([\bar\eta-\Delta\eta,\eta]\times[\bar k-\Delta k,\bar k]\right)=\epsilon\left(b^{-1}\left([\bar\eta-\Delta\eta,\eta]\times\left[\bar k-\Delta k,\bar k\right]\right)\right)$$
$$\leq\epsilon\left(\left[[\bar\eta-\Delta\eta,\eta]\times\left[\bar k-\frac{1}{1-\theta_E}\Delta k,\bar k\right]\right]\times\bar K\right)$$
$$=\alpha\left([\bar\eta-\Delta\eta,\eta]\times\left[\bar k-\frac{1}{1-\theta_E}\Delta k,\bar k\right]\right).$$

□

Lemma 4.4.2 (Endogenous distribution of adult skills with disintegration theorem)**.** *Fix a Borel measure $\alpha\geq 0$ on $\bar A$. Let α_η be the disintegration of α with respect to η. If $\epsilon\geq 0$ on $\bar A\times\bar K$ is such that $\pi_{1\#}\epsilon=\alpha$, then for each $\bar k-\Delta k\in(0,\bar k)$, $\kappa=z_{E\#}\epsilon$ we have:*

$$\int_{[\bar k-\Delta k,\bar k]}\kappa_\eta(dk)\leq\int_{[\bar a-\frac{1}{1-\theta_E}\Delta k,\bar a]}\alpha_\eta(da)$$

where κ_η is the disintegration of κ with respect to η.

Proof. The proof follows the proof of [13, Lemma 14, first claim]. The disintegration of a measure is the conditional expectation, so

$$\int_{[\bar k-\Delta k,\bar k]}\kappa_\eta(dk)=\int_{[\bar k-\Delta k,\bar k]}\frac{\kappa(\eta,dk)}{\int_{\bar K}\kappa(\eta,dk')}$$

As κ is a probability measure, so are its marginals. So $\int_{\bar K}\kappa(\eta,dk')=1$ and the result can be proven the same way as Lemma 4.4.1. □

We can now prove one side of the duality for the perturbed problems. This inequality is usually straightforward to prove, but in this case, we need to show that $v\in L^1(\bar A,z_{E\#}\epsilon)$.

Proposition 4.4.3 (Easy direction of duality for unbounded functions)**.** *If Borel measures $(\epsilon,\lambda)\in G_\delta$ and Borel functions $(u,v)\in F_\delta$ are such that $u\in L^1(\bar A,\alpha)$ and $u\delta,v\delta\in L^1(A,\mathcal{L})$, then*

$$\alpha(u)+\delta\left\langle u+v\right\rangle_A\geq\epsilon(c_\delta b_E(b)+\delta\eta_t)+\lambda\left(b_L(z_L)+\delta f\right)$$

provided $v\in L^1(\bar A,z_{E\#}\epsilon)$.

If α satisfies the doubling condition

$$\int_{[\bar a-\Delta a,\bar a]}\alpha(\eta,da)\leq C\int_{\left[\bar a-\frac{1}{2}\Delta a,\bar a\right]}\alpha(\eta,da)\qquad\forall\eta\in[\underline{\eta},\bar\eta]\tag{4.18}$$

then $v\in L^1(\bar A,z_{E\#}\epsilon)$.

Note that it makes sense to assume that $u\in L^1(\bar A,\alpha)$ and $u\delta,v\delta\in L^1(A,\mathcal{L})$ in order to ensure that $LP_*(\delta)$ is finite. Note also that Proposition 4.4.3 implies that

$$LP_*(\delta)\geq LP^*(\delta).$$

Proof of Proposition 4.4.3. The proof follows the proof of [13, Proposition 8]. Take two pairs (ϵ, λ) and (u, v) satisfying the hypotheses. The stability constraint in education sector (4.7) implies,

$$u(\iota, a) - c_\delta b_E(z_E(\iota, a; \eta_t, t)) - \delta\eta_t \geq v(z_E(\iota, a; \eta_t, t)) - \frac{1}{\rho_E \eta_t} v(\eta_t, t)$$

on $\bar{A} \times \bar{K}$. Integrating over ϵ, we get

$$\begin{aligned}
&\int_{\bar{A}\times\bar{K}} \left(v(z_E(\iota, a; \eta_t, t)) - \frac{1}{\rho_E \eta_t} v(\eta_t, t)\right) \epsilon(d\iota, da; d\eta_t, dt) \\
\leq &\int_{\bar{A}\times\bar{K}} \big(u(\iota, a) - c_\delta b_E(z_E(\iota, a; \eta_t, t)) - \delta\eta_t\big) \epsilon(d\iota, da; d\eta_t, dt) \qquad (4.19)\\
= &\int_{\bar{A}} u(\iota, a) \pi_{1\#}\epsilon(d\iota, da) - \epsilon\big(c_\delta b_E(z_E) + \delta\eta_t\big) \\
= &\int_{\bar{A}} u(\iota, a) \left(\alpha + \frac{\delta}{|A|} \mathcal{L}|_A\right) - \epsilon\big(c_\delta b_E(z_E) + \delta\eta_t\big) \qquad \text{by (4.17b)} \\
= &\alpha(u) + \langle \delta u \rangle_A - \epsilon\big(c_\delta b_E(z_E) + \delta\eta_t\big).
\end{aligned}$$

Adding $\langle \delta v \rangle_K$ (and noting that $K = A$) we get:

$$\begin{aligned}
&\alpha(u) - \epsilon\big(c_\delta b_E(z_E) + \delta\eta_t\big) + \delta \langle u + v \rangle_A \\
\geq &\langle \delta v \rangle_K + \int_{\bar{A}\times\bar{K}} \left(v(z_E(\iota, a; \eta_t, t)) - \frac{1}{\rho_E \eta_t} v(\eta_t, t)\right) \epsilon(d\iota, da; d\eta_t, dt). \qquad (4.20)
\end{aligned}$$

Also, by integrating v over (4.17a) we have:

$$\begin{aligned}
&\langle \delta v \rangle_K + \int_{\bar{K}} v(\eta, k) z_{E\#}\epsilon(d\eta, dk) - \int_{\bar{K}\times\bar{K}} \frac{v(\eta_k, k)}{\rho_E \eta_k} \epsilon(d\iota, da; d\eta_k, dk) \qquad (4.21)\\
= &\int_{\bar{K}\times\bar{K}} v(\eta_k, k) \lambda(d\eta_k, dk; d\eta_m, dm) + \int_{\bar{K}\times\bar{K}} \frac{1}{\rho_L \eta_k} v(\eta_k, k) \lambda(d\eta_w, dw; d\eta_k, dk) \\
= &\int_{\bar{K}\times\bar{K}} \left(v(\eta_w, w) + \frac{1}{\rho_L \eta_m} v(\eta_m, m)\right) \lambda(d\eta_w, w; d\eta_m, m) \\
\geq &\int_{\bar{K}\times\bar{K}} \Big(b_L(z_L(\eta_w, w; \eta_m, m)) + \delta f\Big) \lambda(d\eta_w, w; d\eta_m, m) \qquad \text{by (4.15a).} \qquad (4.22)
\end{aligned}$$

If $v \in L^1(\bar{A}, z_{E\#}\epsilon)$, (4.20) equals (4.21) and we conclude that

$$\alpha(u) + \delta \langle u + v \rangle_A \geq \epsilon(c_\delta b_E \circ z_E + \delta\eta_t) + \lambda\left(b_L \circ z_L + \delta f\right).$$

Now we must show that (4.18) implies that $v \in L^1(\bar{A}, z_{E\#}\epsilon)$. We supposed that $(u, v) = (u_0 + u_1, v_0 + v_1)$ differ from bounded continuous functions $u_0 \in C(\bar{A})$, $v_0 \in C(\bar{K})$ by functions that are non-decreasing in the second parameter u_1, v_1. Suppose, without loss of generality, that v_1 is strictly increasing in the second parameter.

Suppose that α satisfies (4.18). Using the layer cake representation, we have:

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

Let $v_1^{-1}(y,\infty)_1$ be the lowest η such that there exists a k such that $v_1(\eta,k) > y$ and let $k = v_1^{-1}(y,\infty)(\eta)$ be the lowest k such that $v_1(\eta,k) > y$. We have

$$\begin{aligned}
\kappa\left[v_1^{-1}(y,\infty)\right] &= \int_{\left[(v_1^{-1}(y,\infty)_1,\bar{\eta}\right]} \int_{\left[(v_1^{-1}(y,\infty)(\eta),\bar{k}\right]} \kappa(d\eta_k, dk) \\
&= \int_{\left[(v_1^{-1}(y,\infty)_1,\bar{\eta}\right]} \int_{\left[(v_1^{-1}(y,\infty)(\eta),\bar{k}\right]} \kappa_\eta(dk)\pi_{1\#}\kappa(d\eta) \\
&= \int_{\left[(v_1^{-1}(y,\infty)_1,\bar{\eta}\right]} \int_{\left[\bar{k}-\left(\bar{k}-v_1^{-1}(y,\infty)(\eta)\right),\bar{k}\right]} \kappa_\eta(dk)\pi_{1\#}\kappa(d\eta) \\
&\le \int_{\left[(v_1^{-1}(y,\infty)_1,\bar{\eta}\right]} \int_{\left[\bar{a}-\frac{1}{1-\theta_E}\left(\bar{k}-v_1^{-1}(y,\infty)(\eta)\right),\bar{k}\right]} \alpha_\eta(dk)\pi_{1\#}\kappa(d\eta) \\
&\le \int_{\left[(v_1^{-1}(y,\infty)_1,\bar{\eta}\right]} \int_{\left[\bar{a}-\frac{1}{1-\theta_E}\left(\bar{k}-v_1^{-1}(y,\infty)(\eta)\right),\bar{k}\right]} \alpha_\eta(dk)\pi_{1\#}\alpha(d\eta) \\
&\le C^d \int_{\left[(v_1^{-1}(y,\infty)_1,\bar{\eta}\right]} \int_{\left[\bar{a}-\frac{1}{2^d}\frac{1}{1-\theta_E}\left(\bar{k}-v_1^{-1}(y,\infty)(\eta)\right),\bar{k}\right]} \alpha_\eta(dk)\pi_{1\#}\alpha(d\eta)
\end{aligned}$$

where the first inequality follows from Lemma 4.4.2, the second from the fact that $\pi_{1\#}\kappa = \pi_{1\#}\alpha$, and the third from (4.18).

Letting $d = \frac{-\ln(1-\theta_E)}{\ln 2} > 0$, we get

$$\begin{aligned}
\kappa\left[v_1^{-1}(y,\infty)\right] &\le C^{\frac{-\ln(1-\theta_E)}{\ln 2}} \int_{\left[(v_1^{-1}(y,\infty)_1,\bar{\eta}\right]\left[(v_1^{-1}(y,\infty)(\eta),\bar{a}\right]} \alpha_\eta(dk)\pi_{1\#}\alpha(d\eta) \\
&= C^{\frac{-\ln(1-\theta_E)}{\ln 2}} \alpha\left[v_1^{-1}(y,\infty)\right].
\end{aligned}$$

Setting $s = t$ in (4.15b), we have that

$$u(\eta,k) \ge c_\delta b_E(\eta,k) + \delta\eta_t + \frac{\rho_E\eta - 1}{\rho_E\eta} v(\eta,k),$$

so

$$\begin{aligned}
u(\eta,k) &\ge \frac{\rho_E\eta - 1}{\rho_E\eta} v(\eta,k) \\
&\ge \frac{\rho_E\underline{\eta} - 1}{\rho_E\underline{\eta}} v(\eta,k).
\end{aligned}$$

Thus, considering u_1 and v_1, we have

$$u_1(\eta,k) \ge \frac{\rho_E\underline{\eta} - 1}{\rho_E\underline{\eta}} \left(v_1(\eta,k) + v_0(\eta,k)\right) - u_0(\eta,k).$$

As v_0 and u_0 are bounded, we conclude that there exists a constant c such that

$$u_1(\eta,k) \ge \frac{\rho_E\underline{\eta} - 1}{\rho_E\underline{\eta}} \left(v_1(\eta,k) + c\right).$$

Setting

$$\frac{\rho_E\underline{\eta} - 1}{\rho_E\underline{\eta}} = c',$$

we have that

$$u_1(\eta, k) \geq c' \left(v_1(\eta, k) + c\right).$$

So

$$\begin{aligned}
\int_{\bar{K}} v_1(\eta_k, k)\kappa(d\eta_k, dk) &\leq C^{\frac{-\ln(1-\theta_E)}{\ln 2}} \int_0^\infty \alpha[v_1^{-1}(y, \infty)]dy \\
&= C^{\frac{-\ln(1-\theta_E)}{\ln 2}} \int_{\bar{A}} v_1(\eta_a, a)\alpha(d\eta_a, da) \\
&\leq C^{\frac{-\ln(1-\theta_E)}{\ln 2}} \int_{\bar{A}} \frac{1}{c'} u_1(\eta_a, a) - c\alpha(d\eta_a, da) \\
&= C^{\frac{-\ln(1-\theta_E)}{\ln 2}} \left(\frac{1}{c'} \int_{\bar{A}} u_1(\eta_a, a)\alpha(d\eta_a, da) - c\alpha(\bar{A})\right) \\
&< \infty
\end{aligned}$$

because $u_1 \in L^1(\bar{A}, \alpha)$ and $\alpha(\bar{A}) < \infty$. □

To prove the other side of the duality, we will use the duality theorem Lemma 2.4.1.

Theorem 4.4.4 (No duality gap)**.** *Fix $\delta, c_\delta > 0$, $0 < \theta_E < 1$, $0 < \theta_L < 1$, $\bar{\eta} > \underline{\eta} > 1$ and $\bar{a} = \bar{k} > 0$. Let $A = K = [\underline{\eta}, \bar{\eta}) \times [0, \bar{k})$ and let α be a Borel probability measure on $\bar{A}$ satisfying the doubling condition* (4.18) *at $\bar{a}$. Define z_E, z_L, b_E, b_L as in section 4.1. Then the optimal values of the primal* (4.14) *and the dual* (4.16) *problem are the same;*

$$LP^*(\delta) = LP_*(\delta).$$

Proof. Following the proof of [13, Theorem 18], we will apply Lemma 2.4.1 with

$$\begin{aligned}
A &= C(\bar{A}) \oplus C(\bar{K}) \\
B &= C(\bar{A} \times \bar{K}) \oplus C(\bar{K} \times \bar{K}).
\end{aligned}$$

So

$$\begin{aligned}
A^* &= \Gamma(\bar{A}) \oplus \Gamma(\bar{K}) \\
B^* &= \Gamma(\bar{A} \times \bar{K}) \oplus \Gamma(\bar{K} \times \bar{K}),
\end{aligned}$$

where $\Gamma(X)$ is the set of Borel measures on X. Define,

$$\begin{aligned}
\varphi_\delta : C(\bar{A}) \oplus C(\bar{K}) &\to \mathbf{R} \cup \{\infty\} \\
(u, v) &\mapsto \delta \langle u + v \rangle_A + \int_{\bar{A}} u(\eta_a, a)\alpha(d\eta_a, da), \\
\varphi_\delta^* : \Gamma(\bar{A}) \oplus \Gamma(\bar{K}) &\to \mathbf{R} \cup \{\infty\} \\
(\mu, \nu) &\mapsto \begin{cases} 0 & \text{if } (\mu, \nu) = \left(\alpha + \frac{\delta}{|A|}\mathcal{L}|_A, \frac{\delta}{|K|}\mathcal{L}|_K\right) \\ +\infty & \text{otherwise,} \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
\phi_\delta : C(\bar{A} \times \bar{K}) \oplus C(\bar{K} \times \bar{K}) &\to \mathbf{R} \cup \{\infty\} \\
(\tilde{u}, \tilde{v}) &\mapsto \begin{cases} 0 & \text{if } \tilde{u}(\eta_a, a; \eta_k, k) \geq c_\delta b_E(z_E(\eta_a, a; \eta_k, k)) + \delta\eta_t \\ & \text{and } \tilde{v}(\eta_w, w; \eta_m, m)) \geq b_L(z_L(\eta_w, w; \eta_m, m)) + \delta f \\ +\infty & \text{otherwise,} \end{cases} \\
\phi^*_\delta : \Gamma(\bar{A} \times \bar{K}) \oplus \Gamma(\bar{K} \times \bar{K}) &\to \mathbf{R} \cup \{\infty\} \\
(\epsilon, \lambda) &\mapsto \begin{cases} \epsilon(c_\delta b_E(b) + \delta\eta_t) + \lambda\,(b_L(z_L) + \delta f) & \text{if } \epsilon \leq 0 \text{ and } \lambda \leq 0 \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

It is easy to check that φ^*_δ and ϕ^*_δ are indeed the Legendre transforms of φ_δ and ϕ_δ.

Now let

$$H\begin{pmatrix} u \\ v \end{pmatrix}((\iota, a; \eta_t, t), (\eta_w, w; \eta_m, m)) = \begin{pmatrix} u(\iota, a) + \frac{1}{\rho_E \eta_t} v(\eta_t, t) - v(z_E(\iota, a; \eta_t, t)) \\ v(\eta_w, w) + \frac{1}{\rho_L \eta_m} v(\eta_m, m) \end{pmatrix}$$

$$H^*\begin{pmatrix} \epsilon \\ \lambda \end{pmatrix} = \begin{pmatrix} \pi_{1\#}\epsilon \\ \lambda^1 + \left(\frac{\lambda}{\rho_L \eta_2}\right)^2 + \left(\frac{\epsilon}{\rho_E \eta_2}\right)^2 - z_{E\#}\epsilon \end{pmatrix}.$$

Once again, it is easy to show that

$$\left\langle H\begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} \epsilon \\ \lambda \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} u \\ v \end{pmatrix}, H^*\begin{pmatrix} \epsilon \\ \lambda \end{pmatrix} \right\rangle.$$

To apply Lemma 2.4.1, we need to show that ϕ_δ is continuous and real-valued at some point in $H(\mathrm{Dom}\ \varphi_\delta)$. Let $(u, v) \in \mathrm{Dom}\ \varphi_\delta$. We have that $u, v \in L^1(A, \mathcal{L}_A)$ and $u \in L^1(\bar{A}, \alpha)$, which means that $LP_*(\delta)$ is finite. Moreover, $\phi_\delta(H(u, v)) = 0$ if (u, v) are feasible for $LP_*(\delta)$. So the conditions are achieved because the conditions of being feasible and finite for $LP_*(\delta)$ are open and there exists such a pair; for example,

$$(u, v) = (c_\delta b_E(\bar{\eta}, \bar{k}) + b_L(\bar{k}) + \delta\left(\bar{\eta}\bar{k} + \bar{\eta}^2 + 2\bar{\eta}\right), b_L(\bar{k}) + \delta\left(\bar{\eta}\bar{k} + \bar{\eta}^2 + \bar{\eta}\right)).$$

Now Lemma 2.4.1 tells us that

$$\begin{aligned}
&\inf_{(u,v) \in C(\bar{A}) \oplus C(\bar{K})} \varphi_\delta(u, v) + \phi_\delta(H(u, v)) \\
=& \max_{(\epsilon,\lambda) \in \Gamma(\bar{A} \times \bar{K}) \oplus \Gamma(\bar{K} \times \bar{K})} -\varphi^*_\delta(H^*(\epsilon, \lambda)) - \phi^*_\delta(-\epsilon, -\lambda).
\end{aligned}$$

By definition of φ_δ and ϕ_δ, we have that

$$\begin{aligned}
&\inf_{(u,v) \in C(\bar{A}) \oplus C(\bar{K})} \varphi_\delta(u, v) + \phi_\delta(H(u, v)) \\
=& \inf_{\substack{(u,v) \in C(\bar{A}) \oplus C(\bar{K}) \\ H(u,v)_1 \geq c_\delta b_E(b) + \delta\eta_t \\ H(u,v)_2 \geq b_L(z_L) + \delta f}} \delta\left\langle u + v \right\rangle_A + \int_{\bar{A}} u(\eta_a, a)\alpha(d\eta_a, da).
\end{aligned}$$

Then using the definition of H we get that

$$\begin{aligned}
&\inf_{(u,v)\in C(\bar{A})\oplus C(\bar{K})} \varphi_\delta(u,v) + \phi_\delta(H(u,v)) \\
=&\inf_{\substack{(u,v)\in C(\bar{A})\oplus C(\bar{K}) \\ u(\iota,a)+\frac{1}{\rho_E\eta_t}v(\eta_t,t)-v(z_E(\iota,a;\eta_t,t))\geq c_\delta b_E(b)+\delta\eta_t \\ v(\eta_w,w)+\frac{1}{\rho_L\eta_m}v(\eta_m,m)\geq b_L(z_L)+\delta f}} \delta\left\langle u+v\right\rangle_A + \int_{\bar{A}} u(\eta_a,a)\alpha(d\eta_a,da) \\
&\geq LP_*(\delta)
\end{aligned}$$

the inf in $LP_*(\delta)$ is taken over a larger class of functions.

Simillary, we have that:

$$\begin{aligned}
&\max_{(\epsilon,\lambda)\in\Gamma(\bar{A}\times\bar{K})\oplus\Gamma(\bar{K}\times\bar{K})} -\varphi^*_\delta(H^*(\epsilon,\lambda)) - \phi^*_\delta(-\epsilon,-\lambda) \\
=&LP^*(\delta).
\end{aligned}$$

So $LP^*(\delta)\geq LP_*(\delta)$, and equality follows from Proposition 4.4.3.

□

We now use duality to identify conditions for optimality.

Corollary 4.4.5 (Characterizations of optimality)**.** *A pair of non-negative feasible measures* $(\epsilon,\lambda)\in G_\delta$ *maximizes the primal problem* $LP^*(\delta)$ *if there exists feasible* $(u,v)\in F_\delta$ *satisfying the hypothesis of Proposition 4.4.3 such that*

$$\alpha(u)+\delta\left\langle u+v\right\rangle_A = \epsilon(c_\delta b_E(b)+\delta\eta_t)+\lambda\left(b_L(z_L)+\delta f\right).$$

Proof. The proof follows the proof of [13, Corollary 9, first claim]. Let $\epsilon,\lambda\geq 0$ be a pair of feasible measures on $\bar{A}^2$. Let (u,v) be feasible be such that

$$\alpha(u)+\delta\left\langle u+v\right\rangle_A = \epsilon\left(c_\delta b_E(b)+\delta\eta_t\right)+\lambda\left(b_L(z_L)+\delta f\right). \tag{4.23}$$

By Proposition 4.4.3,

$$\epsilon\left(c_\delta b_E(b)+\delta\eta_t\right)+\lambda\left(b_L(z_L)+\delta f\right)\leq LP^*(\delta)\leq LP_*(\delta)\leq \alpha(u)+\delta\left\langle u+v\right\rangle_A.$$

So by (4.23), we have equality everywhere and (ϵ,λ) are optimal for $LP^*(\delta)$, while (u,v) are optimal for $LP_*(\delta)$. □

Corollary 4.4.6 (Characterizations of optimality)**.** *Let* $(u,v)\in F_\delta$*, satisfying the hypothesis of Proposition 4.4.3. The pair* (u,v) *minimize the* $LP_*(\delta)$ *if and only if there exists non-negative* $(\epsilon,\lambda)\in G_\delta$ *such that*

$$\alpha(u)+\delta\left\langle u+v\right\rangle_A = \epsilon(c_\delta b_E(b)+\delta\eta_t)+\lambda\left(b_L(z_L)+\delta f\right).$$

Proof. The proof follows the proof of [13, Corollary 9, second claim]. The proof for the necessary condition, for (u,v) to be minimizers, is identical then the proof of Corollary 4.4.5.

For the sufficient condition, we know from Theorem 4.4.4 that $LP^*(\delta) = LP_*(\delta)$. Suppose (u, v) is optimal, so $\alpha(u) + \delta \langle u + v \rangle_A = LP_*(\delta)$. Then the result follows from Lemma 4.5.4 (existence of optimal measures). □

The next corollary shows that the constraints attain equality a.e.

Corollary 4.4.7 (Characterizations of optimality). *Suppose $(\epsilon, \lambda) \in G_\delta$ and $(u, v) \in F_\delta$, then we have that*

$$\alpha(u) + \delta \langle u + v \rangle_A = \epsilon(c_\delta b_E(b) + \delta\eta_t) + \lambda (b_L(z_L) + \delta f)$$

if and only if $\epsilon(f) = \lambda(g) = 0$, where

$$f(\iota, a; \eta_t, t) = u(\iota, a) + \frac{v(\eta_t, t)}{\rho_E \eta_t} - c_\delta b_E(z_E(\iota, a; , \eta_t, t)) + \delta\eta_t - v(z_E(\iota, a; \eta_t, t)) \geq 0$$

and

$$g(\eta_w, w; \eta_m, m) = v(\eta_w, w) + \frac{v(\eta_m, m)}{\rho_L \eta_m} - b_L(z_L(\eta_w, w; \eta_m, m)) - \delta f \geq 0.$$

Proof. This proof follows the proof of [13, Corollary 9, third claim]. We note that $\epsilon(f) = 0$ is equivalent to the equality in (4.19) from Proposition 4.4.3, and $\lambda(g) = 0$ is equivalent to equality in (4.22) from Proposition 4.4.3. So,

$$\alpha(u) + \delta \langle u + v \rangle_A = \epsilon(c_\delta b_E(b) + \delta\eta_t) + \lambda (b_L(z_L) + \delta f)$$

is equivalent to both of those inequalities being in fact equalities, and the result follows. □

4.5 Existence of solution

The following lemma allows us to take converging subsequences, which will be essential when we prove that the infimum is attained.

Lemma 4.5.1 (Compactness for wage functions – adults). *Let $v_i : K \to [0, \infty)$ be a sequence of convex, non-decreasing in both variables and supermodular functions satisfying for all i:*

$$\text{for a.e. } \eta \in N,\ \exists g_\eta \text{ such that } \frac{\partial^2}{\partial k^2} v_i(\eta, k) \geq g_\eta(k) \text{ for a.e. } k \in [0, \bar{k});$$
$$\text{for a.e. } k \in [0, \bar{k}),\ \exists g_k \text{ such that } \frac{\partial^2}{\partial \eta^2} v_i(\eta, k) \geq g_k(\eta) \text{ for a.e. } \eta \in N;$$
$$\text{and } \frac{\partial^2}{\partial \eta \partial k} v_i(\eta, k) \geq g \text{ for a.e. } (k, \eta) \in K$$

where $0 \leq g, g_{\eta/k}$ and $g_k \in L^1_{loc}([\underline{\eta}, \bar{\eta}])$, $g_\eta \in L^1_{loc}([0, \bar{k}))$. Then

- *for all η, $\{ v_i(\eta, \cdot) \}$ admits a subsequence $\{ v_{i_l}(\eta, \cdot) \}$ which converges pointwise to a limit $v_\eta(\cdot)$:*

$$\forall k \in [0, \bar{k}) \ \lim_{l \to \infty} v_{i_l}(\eta, k) = v_\eta(k)$$

which is real valued on $[0, \tilde{k}_\eta)$, and infinity on $(\tilde{k}_\eta, \bar{k})$, and

- *for all k, $\{ v_i(\cdot, k) \}$ admits a subsequence $\{ v_{i_l}(\cdot, k) \}$ which converges pointwise to a limit $v_k(\cdot)$:*

$$\forall \eta \in N \ \lim_{l \to \infty} v_{i_l}(\eta, k) = v_k(\eta)$$

which is real valued on $[0, \tilde{\eta}_k)$, and infinity on $(\tilde{\eta}_k, \bar{\eta})$.

Both limit functions are in fact the same:

$$v_\eta(k) = v_k(\eta) = v_\infty(\eta, k).$$

Convergence is uniform on compact subsets of

$$[0, \inf_k \tilde{\eta}_k) \times [0, \inf_\eta \tilde{k}_\eta).$$

Moreover, the bounds

$$\frac{\partial^2}{\partial k^2} v_\infty(\eta, k) \geq g_\eta(k)$$
$$\frac{\partial^2}{\partial \eta^2} v_\infty(\eta, k) \geq g_k(\eta)$$
$$\frac{\partial^2}{\partial \eta \partial k} v_\infty(\eta, k) \geq g$$

still hold in the sense of distribution.

As v is convex in both variables, but is not necessarily jointly convex, we need to fix one variable and take the limit for the proof.

Proof of Lemma 4.5.1. The proof follows the proof of [13, Lemma 11, First and Second Claim]. We prove the first bullet point. The proof of the second is identical.

Fix η. By the fundamental theorem of calculus, we have:

$$v_i(\eta, k') = v_i(\eta, 0) + \int_0^{k'} \frac{\partial}{\partial k} v_i(\eta, k) dk, \tag{4.24}$$

as v_i are convex in k, and $\frac{\partial}{\partial k} v_i(\eta, k)$ are non-decreasing in k for all i. Thus, we can use Helly's selection theorem to get a subsequence with converging non-decreasing limit $\frac{\partial}{\partial k} v_\eta(k)$.

By choosing a further subsequence, we prove convergence of $v_i(\eta, 0)$ towards $v_\eta(0)$. If no such subsequence exists, we let $v_\eta(0) = \infty$ and we can conclude with $\tilde{k}_\eta = 0$. Otherwise, we can choose $\tilde{k}_\eta \in [0, \bar{k}]$ such that $\frac{\partial}{\partial k} v_\eta(k) < \infty$ if $k < \tilde{k}_\eta$ and $\frac{\partial}{\partial k} v_\eta(k) = \infty$ for $k > \tilde{k}_\eta$.

For $k' < \tilde{k}_\eta$, Lebesgue's dominated theorem with $\frac{\partial}{\partial k} v_\eta(k')$ as the dominant allows us to pass to the limit in (4.24) to obtain a limit $v_\eta(k')$.

As mentioned above, the proof of the second bullet point is the same.

The fact that $v_\eta(k) = v_k(\eta) = v_\infty(\eta, k)$ follows from fixing (η, k) and taking a subsequence to get convergence for v_η and v_k. The limit has to be the same.

To see that convergence is uniform on compact subsets of

$$[0, \inf_k \tilde{\eta}_k) \times [0, \inf_\eta \tilde{k}_\eta)$$

we need to apply the fundamental theorem of calculus twice to write

$$v_i(\eta', k') = v_i(0,0) + \int_0^{\eta'} \frac{\partial}{\partial \eta} \left(v_i(\eta, 0) + \int_0^{k'} \frac{\partial}{\partial k} v_i(\eta, k) dk \right) d\eta$$

and apply Lebesgue dominated convergence twice.

Now we prove that the bound

$$\frac{\partial^2}{\partial k^2} v_\infty(\eta, k) \geq g_\eta(k)$$

holds. Note that the bound

$$\frac{\partial^2}{\partial k^2} v_i(\eta, k) \geq g_\eta(k)$$

holds in the distributional sense, i.e.

$$\int_0^{k_\eta} \left(f''(k) v_i(\eta, k) - f(k) g_\eta(k)\right) dk \geq 0$$

for each smooth compactly supported test function $0 \leq f \in C_c^\infty([0, k_\eta])$. Since f, f'' are bounded, $v_i(\eta, \cdot)$ is also bounded on $[0, k_\eta)$. Therefore, we can use Lebesgue's dominated convergence theorem again to conclude that the inequality survives the limit.

The bound on the second derivative with respect to η is proven the same way.

For the cross derivative bound, the proof is the same, but the integral is over K and the function $f : K \to \mathbf{R}$. □

Lemma 4.5.2 (Compactness for wage functions – students)**.** *Under the same hypothesis as Lemma 4.5.1, if $k_\eta > 0$ then for fixed η, the function:*

$$u_i(\eta, a) = \max_{(\eta_t, t) \in K} cb_E\left(z_E(\eta, a; \eta_t, t)\right) + v_i\left(z_E(\eta, a; \eta_t, t)\right) - \frac{1}{\rho_E \eta_t} v_i(\eta_t, t)$$

diverges for $a > k_\eta$ as $i \to \infty$ along the converging subsequence of Lemma 4.5.1.

Proof. The proof follows the proof of [13, Lemma 11, third claim]. If $a > k_\eta$, there exits a $k < k_\eta$ such that $z_E^1(\eta, a; \eta_t, k) > k_\eta$. For this k, $cb_E(z_E(\eta, a; \eta_t, k)) < \infty$, $v_\infty(z_E(\eta, a; \eta_t, k)) = \infty$ and $v_\infty(\eta_t, k) < \infty$, so $u_0(\eta, a) = \infty$. □

Corollary 4.5.3 (Convergence uniform from below)**.** *Suppose we have a sequence of functions $v_i : K \to [0, \infty)$ satisfying the hypotheses of Lemma 4.5.1 that converge to v_∞.*

Suppose that for a fixed η, $v_\infty(\eta, \cdot)$ is real valued for $k \in [0, k_\eta)$ and infinite for $k \in (k_\eta, \bar{k})$. If

$$\lim_{\substack{k \to k_\eta \\ k < k_\eta}} v_\infty(\eta, k) < \infty$$

then

$$0 \leq \liminf_{i \to \infty} \inf_{k \in [0, k_\eta)} v_i(\eta, k) - v_\infty(\eta, k).$$

If

$$\lim_{\substack{k \to k_\eta \\ k < k_\eta}} v_\infty(\eta, k) = \infty$$

the sequence grows uniformly, i.e. for all $c < \infty$, $i' < \infty$ large enough implies

$$v_i(\eta, k) \geq c$$

for all $k > k_0 - \frac{1}{i'}$ and $i > i'$.

Proof. This proof follows the proof of [13, Corollary 12]. Fix δ and take $k' < k_\eta$ such that $v_\infty(\eta, k') > v_\infty(\eta, k_\eta^-) - \delta/2$. Take i to be large enough, so that

$$\begin{aligned} v_i(\eta, k') &> v_\infty(\eta, k') - \delta/2 \\ &> v_\infty(\eta, k_\eta^-) - \delta. \end{aligned}$$

So for all $k \in [k', k_\eta)$,

$$v_i(\eta, k) > v_\infty(\eta, k) - \delta$$

because $v_i(\eta, \cdot)$ is monotone. Thus the inequality $\liminf_{i \to \infty} v_i(\eta, k) - v_\infty(\eta, k) \geq 0$ holds for $k \in [k', k_\eta)$. For $k \in [0, k']$ the conclusion follows from the fact that convergence is uniform for compact subset of $[0, k_\eta)$.

If

$$\lim_{\substack{k \to k_\eta \\ k < k_\eta}} v_\infty(\eta, k) = \infty,$$

for $c < \infty$, take i' large enough that $v_\infty(k_0 - \frac{1}{i'}) > c$ and then take i' larger if necessary to ensure $v_i\left(k_0 - \frac{1}{i'}\right) > c$ for all $i > i'$. Monotonicity of v_i for all i conclude the proof. □

Lemma 4.5.4 (Existence of optimal measure)**.** *For fixed δ, c_δ non-negative and θ_E, θ_L, ρ_E, ρ_L positive, $\theta_E, \theta_L \in (0, 1)$, α a Borel probability measure on $\bar{A}$, then there exists feasible measures $\epsilon_\delta \geq 0$ on $\bar{A} \times \bar{K}$ and $\lambda_\delta \geq 0$ on $\bar{K} \times \bar{K}$ maximizing the perturbed problem* (4.16)*.*

Proof. This proof follows the proof of [13, Lemma 17]. To prove the existence of maximizer, we show that the objective functional is continuous and the feasible set is compact and non-empty.

- Since $b_E \circ z_E$ and $b_L \circ z_L$ are continuous, the linear functional to optimize (4.16) is continuous.
- The continuous functions $C(\bar{A} \times \bar{K})$ on the compact square $\bar{A} \times \bar{K}$ equipped with the supremum norm $\|\cdot\|_\infty$ form a Banach space. Borel positive measures on $\bar{A} \times \bar{K}$ are equivalent to non-negative linear functional on $\bar{A} \times \bar{K}$ by the Riesz-Markov-Kakutani representation theorem. These functionals are in the dual of $C(\bar{A} \times \bar{K})$. As the ϵ's are probability measures, they have unitary norm. As the λ's are positive measures with norm between $\rho\underline{\eta}$ and $\rho\bar{\eta}$ they are contained into a cylinder. By the Banach-Alaoglu theorem, they form a weak-$*$ subset in the dual of $C(\bar{A} \times \bar{K})$.

By definition of weak-$*$ topology, a sequence ϵ_i converge to ϵ_∞ if and only if the integral $\epsilon_i(f)$ converge to $\epsilon_\infty(f)$ for all $f \in C(\bar{A} \times \bar{K})$.

Suppose $(\epsilon, \lambda) \in G_\delta$. Then we have

$$\int_{\bar{K}^2} \left(f(\eta_w, w) + \frac{f(\eta_m, m)}{\rho_L \eta_m} \right) \lambda(\eta_w, w; \eta_m, m)$$
$$= \langle \delta f \rangle_K + \int_{\bar{A} \times \bar{K}} \left(f(z_E(\iota, a; \eta_t, t)) - \frac{f(\eta_t, t)}{\rho_E \eta_t} \right) \epsilon(\iota, a; \eta_t, t) \qquad \text{and} \tag{4.17a}$$
$$\int_{\bar{A} \times \bar{K}} f(\iota, a) \epsilon(d\iota, da; d\eta_t, dt) = \int_{\bar{A}} f(\iota, a) \alpha(d\iota, a) + \langle f \delta \rangle_A \tag{4.17b}$$

for each $f \in C(\bar{A})$. Thus, G_δ is compact.

- We need to prove that G_δ is non-empty. Suppose ϵ concentrates on the diagonal as follows:

$$\epsilon = (\mathbb{1} \times \mathbb{1})_\# \left(\alpha + \frac{\delta}{\mathcal{L}(A)} \mathcal{L}|_A \right).$$

Then the marginals of ϵ are equal to $z_{E\#}\epsilon$ because $z_E(\iota, a; \iota, a) = (\iota, a)$:

$$\pi_{1\#}\epsilon = \pi_{2\#}\epsilon = \kappa = z_{E\#}\epsilon = \alpha + \frac{\delta}{\mathcal{L}(A)} \mathcal{L}|_A).$$

In this case, it is obvious that (4.17b) is satisfied. Taking

$$\lambda(\eta_1, k_1; \eta_2; k_2) = \frac{1 - \frac{1}{\rho_E \eta_1}}{1 + \frac{1}{\rho_L \eta_1}} \epsilon(\eta_1, k_1; \eta_2, k_2) + \frac{1}{1 + \frac{1}{\rho_L \eta_1}} (\mathbb{1} \times \mathbb{1})_\# \left(\frac{\delta}{\mathcal{L}(A)} \mathcal{L}|_A \right)$$

ensures that (4.17a) is satisfied too. Thus G_δ is non-empty. □

Theorem 4.5.5 (Existence of minimizing wages)**.** *For fixed $A = K$, θ_E, θ_L, satisfying conditions of section 4.1, α satisfying* (4.18) *and null on sets of Hausdorff dimension one, the infimum of the dual problem* (4.5) *is attained by functions (u, v) satisfying $v = \max\{v_w, v_m, v_t\}$ on $\bar{K}$, and*

$$u(\iota, a) = \sup_{(\eta_t, t) \in \bar{K}} cb_E(b(\iota, a; \eta_t, t)) + v(b(\iota, a; \eta_t, t)) - \frac{1}{\rho_E \eta_t} v(\eta_t, t)$$

on $\bar{A}$, where $v_{w/m/t}$ are defined in section 4.2. Moreover, u, v are continuous, convex in their second variable, non-decreasing in each variable, supermodular and, real-valued on $N \times [0, \bar{a})$.

Proof. The proof follows the proof of [13, Theorem 13].

Existence of optimizer for fixed $\delta > 0$, $c_\delta > 0$.

We will first study the perturbed dual problem (4.14), for $0 < \delta < 1$ and $c_\delta := c > 0$ (or $c_\delta = \delta$ if $c = 0$) under the constraints (4.15a) and (4.15b). We will add the artificial constraint that v is convex and non-decreasing in both variables and supermodular. First, we want to prove that there exists a minimizer.

Note that, as pointed out in the proof of Theorem 4.4.4, the pair of constant functions

$$(u, v) = (c_\delta b_E(\bar{\eta}, \bar{k}) + b_L(\bar{k}) + \delta \left(\bar{\eta}\bar{k} + \bar{\eta}^2 + 2\bar{\eta} \right), b_L(\bar{k}) + \delta \left(\bar{\eta}\bar{k} + \bar{\eta}^2 + \bar{\eta} \right))$$

is feasible, i.e. satisfies (4.15a) and (4.15b). Thus, we have an upper bound of

$$\delta\left\langle 2\bar{b}_L + c_\delta \bar{b}_E + \delta\left(2\bar{\eta}\bar{k} + 2\bar{\eta}^2 + 4\bar{\eta}\right)\right\rangle_A + \int_{\bar{A}} \left(c_\delta \bar{b}_E + \bar{b}_L + \delta\left(\bar{\eta}\bar{k} + \bar{\eta}^2 + 2\eta\right)\right)\alpha(d\iota, da)$$

for the infimum.

Let (u^i_δ, v^i_δ) be a minimizing sequence for (4.14). Lemma 4.5.1 allows us to find a converging subsequence for both functions. The function v_δ that is the limit is non-decreasing and convex in both variables.

The limit (u_δ, v_δ) satisfies the constraints (4.15a) and (4.15b) because pointwise limits conserve inequalities.

The limit (u_δ, v_δ) minimize (4.14) by Fatou's lemma as follows:

$$\begin{aligned}\delta\left\langle u_\delta + v_\delta\right\rangle_A + \int_{\bar{A}} u_\delta(\iota, a)\alpha(d\iota, a) &\le \liminf_{i\to\infty} \delta\left\langle u^i_\delta + v^i_\delta\right\rangle_A + \int_{\bar{A}} u^i_\delta(\iota, a)\alpha(d\iota, a) \\ &= LP_*(\delta).\end{aligned}$$

Properties of the optimizer (u_δ, v_δ)**.**

- By Lemma 4.5.1, for each fixed η, there exists a k_η such that the limit v_δ is finite for $k < k_\eta$ and infinite for $k > k_\eta$. Because $LP_*(\delta)$ is bounded, $k_\eta = \bar{k}$. Similarly, we can prove that v_δ is finite for $\eta < \bar{\eta}$ for all $k \in [0, \bar{k})$.

- Note that

$$u_\delta(\eta, a) \ge \max_{(\eta_t, t)\in K} cb_E(z_E(\eta, a; \eta_t, t)) + v_\delta(z_E(\eta, a; \eta_t, t)) - \frac{1}{\rho_E \eta_t} v_\delta(\eta_t, t) \tag{4.26}$$

is in fact an equality. If is isn't, we can improve the objective functional by replacing u_δ by the right hand side of (4.26).

- Using $(u, v) = (u_\delta, v_\delta)$, we define $(v^w_\delta, v^m_\delta, v^t_\delta) := (v_w, v_m, v_t)$ and $\bar{v}_\delta := \max\{v_w, v_m, v_t\}$. Because the solution (u_δ, v_δ) is feasible, we have $v_\delta \ge \bar{v}_\delta$.

We want to prove that $v_\delta = \bar{v}_\delta$. To do so, we can prove that the function $\phi = v_\delta - \bar{v}_\delta$ is zero on its domain.

Consider the perturbation

$$v^\lambda := (1-\lambda)v_\delta + \lambda\bar{v}_\delta,$$

which is still convex and non-decreasing in both variables, and supermodular for $\lambda \in (0, 1)$. We will prove that for λ small enough, the pair (u_δ, v^λ) also respects the other constraints, i.e. (4.15a) and (4.15b), unless the continuous function $\phi \cong 0$ in K. If this is the case, this pair improves the objective, which is a contradiction and we conclude that $v_\delta = \bar{v}_\delta$.

Noting that

$$\begin{aligned}v^\lambda &= v_\delta - \lambda\phi \\ &= \bar{v}_\delta + (1-\lambda)\phi\end{aligned}$$

we have

$$
\begin{aligned}
& v^\lambda(\eta_w, w) + \frac{1}{\rho_L \eta_m} v^\lambda(\eta_m, m) - b_L(z_L(\eta_w, w; \eta_m, m)) - \delta \\
\geq & v_w(\eta_w, w) + \frac{1}{\rho_L \eta_m} v_\delta(\eta_m, m) - b_L(z_L(\eta_w, w; \eta_m, m)) \\
& - \delta f + (1-\lambda)\phi(\eta_w, w) - \frac{\lambda}{\rho_L \eta_m}\phi(\eta_m, m) && (\bar{v}_\delta \geq v_w) \\
\geq & v_w(\eta_w, w) - \sup_{(\eta_m, m)\in \bar{K}} \left\{ b_L(z_L(\eta_w, w; \eta_m, m)) + \delta f - \frac{1}{\rho_L \eta_m} v_\delta(\eta_m, m) \right\} \\
& + (1-\lambda)\phi(\eta_w, w) - \frac{\lambda}{\rho_L \eta_m}\phi(\eta_m, m) \\
= & (1-\lambda)\phi(\eta_w, w) - \frac{\lambda}{\rho_L \eta_m}\phi(\eta_m, m) && (\text{def of } v_w \text{ (4.8)}) \\
= & \phi(\eta_w, w)\left(1 - \lambda\left(1 + \frac{1}{\rho_L \eta_m}\frac{\phi(\eta_m, m)}{\phi(\eta_w, w)}\right)\right) && (4.27)
\end{aligned}
$$

and

$$
\begin{aligned}
& v^\lambda(\eta_w, w) + \frac{1}{\rho_L \eta_m} v^\lambda(\eta_m, m) - b_L(z_L(\eta_w, w; \eta_m, m)) - \delta f \\
= & v_\delta(\eta_w, w) + \frac{1}{\rho_L \eta_m} \bar{v}_\delta(\eta_m, m) - b_L(z_L(\eta_w, w; \eta_m, m)) - \delta f - \lambda\phi(\eta_w, w) + \frac{1-\lambda}{\rho_L \eta_m}\phi(\eta_m, m) \\
\geq & \frac{\phi(\eta_m, m)}{\rho_L \eta_m}\left(1 - \lambda\left(1 + \rho_L \eta_m \frac{\phi(\eta_w, w)}{\phi(\eta_m, m)}\right)\right). && (4.28)
\end{aligned}
$$

We want to prove that

$$
v^\lambda(\eta_w, w) + \frac{1}{\rho_L \eta_m} v^\lambda(\eta_m, m) - b_L(z_L(\eta_w, w; \eta_m, m)) - \delta \geq 0 \qquad (4.15a)
$$

to show that the constraint is satisfied for v^λ. If one or both of $\phi(\eta_w, w)$ and $\phi(\eta_m, m)$ vanish, there is nothing to prove. If both $\phi(\eta_w, w)$ and $\phi(\eta_m, m)$ are positive, we have that one of (4.27) and (4.28) is positive taking $\lambda < 1/2$.

Now to prove that (4.15b) is satisfied, observe that by adding

$$
u_\delta(\iota, a) - c_\delta b_E(z_E(\iota, a; \eta_t, t))
$$

to

$$
\begin{aligned}
& \frac{1}{\rho_E \eta_t} v^\lambda(\eta_t, t) - v^\lambda(z_E(\iota, a; \eta_t, t)) \\
= & \frac{1}{\rho_E \eta_t} \bar{v}_\delta(\eta_t, t) - v_\delta(z_E(\iota, a; \eta_t, t)) + \frac{1-\lambda}{\rho_E \eta_t}\phi(\eta_t, t) + \lambda\phi(z_E(\iota, a; \eta_t, t)),
\end{aligned}
$$

we get

$$
\begin{aligned}
&\frac{1}{\rho_E \eta_t} v^\lambda(\eta_t, t) + u_\delta(\iota, a) - v^\lambda(z_E(\iota, a; \eta_t, t)) - c_\delta b_E(z_E(\iota, a; \eta_t, t)) - \delta \eta_t \\
\geq & \frac{1}{\rho_E \eta_t} v_t(\eta_t, t) + u_\delta(\iota, a) - v_\delta(z_E(\iota, a; \eta_t, t)) - c_\delta b_E(z_E(\iota, a; \eta_t, t)) - \delta \eta_t \\
& + \frac{1-\lambda}{\rho_E \eta_t} \phi(\eta_t, t) + \lambda \phi(z_E(\iota, a; \eta_t, t)) && \bar{v}_\delta \geq v_t \\
\geq & \frac{1}{\rho_E \eta_t} \Bigg(v_t(\eta_t, t) - \rho_E \eta_t \sup_{(\iota, a) \in \bar{A}} \Big\{ v_\delta(z_E(\iota, a; \eta_t, t)) + c_\delta b_E(z_E(\iota, a; \eta_t, t)) \\
& + \delta \eta_t - u_\delta(\iota, a) \Big\} \Bigg) + \frac{1-\lambda}{\rho_E \eta_t} \phi(\eta_t, t) + \lambda \phi(z_E(\iota, a; \eta_t, t)) \\
= & \frac{1-\lambda}{\rho_E \eta_t} \phi(\eta_t, t) + \lambda \phi(z_E(\iota, a; \eta_t, t)) && \text{by (4.10)} \\
\geq & 0.
\end{aligned}
$$

This proves that $v_\delta = \bar{v}_\delta$ on K.

• We also prove that the artificial constraints forcing v_δ to be non-decreasing and convex in both variables and supermodular are in fact respected in the original problem.

Similarly to Lemma 4.2.4, we can find bounds on derivatives of $\bar{v}_\delta$ as follows:

$$
\begin{aligned}
\frac{\partial \bar{v}_\delta(\eta, k)}{\partial k} &\geq \min \left\{ (1-\theta_E) \underline{b}'_L, \eta \rho_L \theta_L \underline{b}'_L, \rho_E \eta_t \theta_E \left(c_\delta \underline{b}'_E + \inf_{(\eta, k)} \frac{\partial}{\partial k} v_\delta(\eta, k) \right) \right\}; \\
\frac{\partial^2 \bar{v}_\delta(\eta, k)}{\partial k^2} &\geq \min \left\{ (1-\theta_E)^2 \underline{b}''_L, (\theta_L)^2 \eta \rho_L \underline{b}''_L, \rho_E \eta_t \theta_E^2 \left(c_\delta \underline{b}''_E + \inf_{(\eta, k)} \frac{\partial^2}{\partial k^2} v_\delta(\eta, k) \right) \right\}; \\
\frac{\partial^2 \bar{v}_\delta(\eta, k)}{\partial \eta \partial k} &\geq \min \left\{ \delta, \rho_L \theta_L \underline{b}'_L, \rho_E \theta_E \left(c_\delta \underline{b}'_E + \inf_{(\eta, k)} \frac{\partial}{\partial k} v_\delta(\eta, k) \right) \right\}; \\
\frac{\partial}{\partial \eta} \bar{v}_\delta &\geq \delta \eta; \text{ and} \\
\frac{\partial^2}{\partial \eta^2} \bar{v}_\delta &\geq \delta
\end{aligned}
$$

a.e.

So $\bar{v}_\delta = v_\delta$ is strictly non-decreasing, strictly convex in both variables, and strictly supermodular.

We now want to show that (u_δ, v_δ) minimizes (4.14) if we remove the artificial constraint. Suppose for contradiction that the objective function is lower at some other feasible pair $(u, v) \in F_\delta$.

If $u \in C^2(\bar{A})$, $v \in C^2(\bar{K})$, then

$$(u, v) = (1-s)(u_\delta, v_\delta) + s(u, v) \in F_\delta$$

also lowers the objective for $s > 0$. For a small enough s, v inherits the monotonicity and convexity properties of v_δ, contradicting the optimality of (u_δ, v_δ).

If $u \notin C^2(\bar{A})$, or $v \notin C^2(\bar{K})$, we want to get the same contradiction. First, we can assume u, v are continuous and bounded by the proof of Theorem 4.4.4. Then, we can apply the Stone-Weierstrass theorem to approximate u, v by uniformly smooth functions $(\tilde{u}_\sigma, \tilde{v}_\sigma)$ such that

$$u + \sigma \leq \tilde{u}_\sigma \leq u + 2\sigma$$
$$v \leq \tilde{v}_\sigma \leq v + \sigma.$$

With those bounds, we have that $(\tilde{u}_\sigma, \tilde{v}_\sigma)$ is in F_δ if (u, v) is.

Using a small enough σ, $(\tilde{u}_\sigma, \tilde{v}_\sigma)$ lowers the objective value, thus we can get the same contradiction.

We conclude that (u_δ, v_δ) minimizes (4.14) in F_δ.

Existence of optimizer for $\delta = 0$.

By Lemma 4.5.4, there exists ϵ_δ, λ_δ maximizing the dual problem (4.16). As $LP_*(\delta) = LP^*(\delta)$, we have

$$\alpha(u_\delta) + \delta \langle u_\delta + v_\delta \rangle_A = c_\delta \epsilon_\delta \left(b_E(b)\right) + \lambda_\delta \left(b_L(z_L)\right).$$

Lemma 4.5.1 allows us to find a subsequence $(u_{\delta_i}, v_{\delta_i})$ that admits a limit (u_0, v_0) pointwise on $\bar{A} \times \bar{K}$ and uniformly on compact subsets of

$$\left(\left[0, \inf_k \tilde{\eta}_k\right) \times \left[0, \inf_\eta \tilde{k}_\eta\right)\right)^2$$

where $\tilde{\eta}_k$, $\tilde{k}_\eta$ are defined in Lemma 4.5.1. We will write $\delta \to 0$ to refer to the converging subsequence.

We want to show that $k_\eta = \bar{k}$ for all η. Suppose for contradiction that $k_\eta < \bar{k}$. Then by Corollary 4.5.3, $u_\delta(\eta, a) \to \infty$ uniformly on $a \in [(k_\eta + \bar{k}), \bar{k}]$. As $v_{w/m/t}$ depend linearly on η, if $u_\delta(\eta, a) \to \infty$ for an $\eta \in [\underline{\eta}, \bar{\eta}]$ then it diverges for all $\eta \in [\underline{\eta}, \bar{\eta}]$. There exists an η' such that

$$\eta' \times [(k_\eta + \bar{k}), \bar{k}] \cap \bar{K} \neq \emptyset.$$

Thus, we have a contradiction using Fatou's Lemma as in Step 1 of the proof of this Theorem.

The fact that $\tilde{\eta}_k = \bar{\eta}$ for all k follows from the fact that $v_{w/m/t}$ depend linearly on η.

Properties of (u_0, v_0). All of those properties are proven as in [13].

- We show that equality still holds in (4.26) when $\delta = 0$. Let

$$\begin{aligned} f_\delta(\iota, a; \eta_t, t) :=& u_\delta(\iota, a) + \frac{1}{\rho_E \eta_t} v_\delta(\eta_t, t) - c_\delta b_E(z_E(\iota, a; \eta_t, t)) - \delta v_t \\ & - v_\delta(z_E(\iota, a; \eta_t, t)) \\ \geq& 0. \end{aligned}$$

As

$$\frac{\partial^2}{\partial \iota \partial \eta_t} f_\delta = 0 \qquad \text{and} \qquad \frac{\partial^2}{\partial a \partial t} f_\delta \leq 0,$$

the zero set of f_δ, Z_δ is non-decreasing in ι and η_t and in a and t.

The zero set Z_δ of f_δ is closed in $A \cap \bar{K}$ and

$$A \times \bar{K} \cap \operatorname{spt} \epsilon_\delta \subset Z_\delta$$

by Corollary 4.4.7.

Thus, for all $(\iota, a; \eta_t, t) \in Z_\delta$

$$\int_{(\iota,\bar{\eta}]\times(a,\bar{a}]\times\bar{K}} \epsilon_\delta(d\eta^s, da^s; d\eta^t, dk^t) \leq \int_{\bar{A}\times[\eta_t,\bar{\eta}]\times[t,\bar{k}]} \epsilon_\delta(d\eta^s, da^s; d\eta^t, dk^t).$$

Equations (4.17a) and (4.17b) give us the left and right marginals of ϵ. Expending the left marginal, we get

$$\int_{(\iota,\bar{\eta}]\times(a,\bar{a}]\times\bar{K}} \epsilon_\delta(d\eta^s, da^s; d\eta^t, dk^t) = \alpha\left((\iota,\bar{\eta}]\times(a,\bar{a}]\right) + \delta\Delta\iota\Delta a,$$

where $\Delta\iota = \bar{\eta} - \iota$ and $\Delta a = \bar{a} - a$.

Expending the right marginal gives:

$$\begin{aligned}
&\int_{\bar{A}\times[\eta_t,\bar{\eta}]\times[t,\bar{k}]} \epsilon_\delta(d\eta^s, da^s; d\eta^t, dk^t) \\
&\leq \rho_E\bar{\eta}\left(z_{E\#}\epsilon\left([\eta_t,\bar{\eta}]\times[t,\bar{k}]\right) + \delta\Delta\eta_t\Delta t \right.\\
&\qquad \left. -\lambda^1\left([\eta_t,\bar{\eta}]\times[t,\bar{k}]\right) - (\lambda/\rho_L\eta_2)^2\left([\eta_t,\bar{\eta}]\times[t,\bar{k}]\right)\right) \\
&\leq \rho_E\bar{\eta}\left(z_{E\#}\epsilon\left([\eta_t,\bar{\eta}]\times[t,\bar{k}]\right) + \delta\Delta\eta_t\delta t\right) \\
&\leq \rho_E\bar{\eta}\left(\alpha\left([\eta_t,\bar{\eta}]\times\left[\bar{a} - \frac{\Delta t}{1-\theta_E}, \bar{a}\right]\right) + \delta\Delta\eta_t\Delta t\right) \qquad \text{by Lemma 4.4.1.}
\end{aligned}$$

Putting those two facts together, we get that

$$\alpha\left((\iota,\bar{\eta}]\times(a,\bar{a}]\right) + \delta\Delta\iota\Delta a \leq \rho_E\bar{\eta}\left(\alpha\left([\eta_t,\bar{\eta}]\times\left[\bar{a} - \frac{\Delta t}{1-\theta_E}, \bar{a}\right]\right) + \delta\Delta\eta_t\Delta t\right).$$

When $\delta \to 0$, the left hand side stays away from zero. So the right hand side is strictly positive. We supposed that α is null on sets of Hausdorff dimension one, so.

$$\alpha\left(\bar{\eta}, [0,\bar{a}]\right) = \alpha\left([0,\bar{\eta}]\times\bar{a}\right) = 0$$

and conclude that $\eta_t < \bar{\eta}$ and $t < \bar{k}$. Thus, the supremum in (4.26) is attained inside K and thus the equality in (4.26) still holds.

- We show that $v_0 = \bar{v}_0$ on K. Taking $\delta \to 0$ in $v_\delta = \bar{v}_\delta$ gives the following:

$$v_0 = \lim_{\delta\to 0} v_\delta = \max\left\{\limsup_{\delta\to 0} v_\delta^w, \limsup_{\delta\to 0} v_\delta^m, \limsup_{\delta\to 0} v_\delta^t\right\}. \tag{4.29}$$

Suppose

$$\sup_{(\eta,k)\in\bar{K}} v_0(\eta,k) < \infty.$$

By definition (4.10) for (η, k), $k < \bar{k}$,

$$v_\delta^t(\eta,k) = \rho_E\eta\left(\sup_{(\iota,a)\in\bar{A}} c_\delta b_E(z_E(\iota,a;\eta,k)) + \delta\eta + v_\delta(z_E(\iota,a;\eta,k)) - u_\delta(\iota,a)\right). \tag{4.30}$$

The convergence of v_δ is uniform, because $z_E^1(\iota, a; \eta, k) < \bar{k}$. Moreover, by Corollary 4.5.3

$$u_0(\iota, a) \leq \liminf_{\delta \to 0} u_\delta(\iota, a).$$

Thus, we conclude that $\limsup_{\delta \to 0} v_\delta^t \leq v_0^t$.

Similarly, we can show that $\limsup_{\delta \to 0} v_\delta^w \leq v_0^w$ and $\limsup_{\delta \to 0} v_\delta^m \leq v_0^m$, so we conclude $v_0 \leq \max\{ v_0^w, v_0^m, v_0^t \}$. The opposite inequality follows from the constraints and we conclude that $v_0 = \bar{v}_0$.

Suppose $\lim_{\substack{k \to \bar{k} \\ k < \bar{k}}} v_0(\eta, k) = \infty$. For fixed $k \in [0, \bar{k})$, let

$$C_\delta = \sup_{(\iota, a) \in A} c_\delta b_E(z_E(\iota, a; \eta, k)) + v_\delta(z_E(\iota, a; \eta, k)).$$

Then $C_\delta \xrightarrow[\delta \to 0]{} C_0 < \infty$. Let δ_0 be small enough that $C_{\delta_0} < 2C_0$. As v_δ diverges, u_δ diverges too. So, we can (by making δ_0 smaller if necessary) have $u_\delta(\eta, \bar{a} - \delta_0) > 2C_0$ for all $\delta < \delta_0$. Thus, for $\delta < \delta_0$ the supremum (4.30) can be restricted to $a \in [0, \bar{a} - \delta_0]$. In this interval, the convergence $(u_\delta, v_\delta) \to (u_0, v_0)$ is uniform. So taking $\delta \to 0$ in (4.30) gives $\lim_{\delta \to 0} v_\delta^t = v_0^t$. Similarly, we can show $\lim_{\delta \to 0} v_\delta^{w/m} = v_0^{w/m}$. So we conclude $v_0 = \bar{v}_0$ by (4.29).

• Now, we can prove that u_0 diverges if and only if v_0 does. Let $\bar{k}^-$ denote the limit when $k \to \bar{k}$ and $\bar{\eta}^-$ when $\eta \to \bar{\eta}$.

First, suppose that $v_0(\eta, \bar{k}^-) = \infty$ or $v_0(\bar{\eta}^-, k) = \infty$. We can then show that $u_0 = \infty$ using a similar argument to the proof of Lemma 4.5.2.

Now if $v_0(\eta, \bar{k}^-) < \infty$, then

$$\begin{aligned} u_0(\eta, \bar{k}^-) &= \sup_{(\eta_t, t)} c_0 b_E\left(\eta, (1-\theta_E)\bar{k}^- + \theta_E t\right) + v_0\left(\eta, (1-\theta_E)\bar{k}^- + \theta_E t\right) - \frac{1}{\rho_E \eta_t} v_0(\eta_t, t) \\ &\leq \sup_{(\eta_t, t)} c_0 b_E\left(\eta, (1-\theta_E)\bar{k}^- + \theta_E t\right) + v_0\left(\eta, (1-\theta_E)\bar{k}^- + \theta_E t\right) \\ &< \infty. \end{aligned}$$

If $v_0(\bar{\eta}^-, k) < \infty$, then as $v_0 = \bar{v}_0$, it is linear in η. So $v_0(\eta, k) < \infty$ for all η. In that case, $u_0(\bar{\eta}^-, k)$ doesn't diverge because b_E is bounded from above.

• Finally, this can be extended to $\bar{K}$. If $v_0(\eta, \bar{k}^-) < v_0(\eta, \bar{k})$ we can replace $v_0(\eta, \bar{k}^-)$ by $v_0(\eta, \bar{k})$ without violating any constraints. It doesn't affect the value of $v_{w/m/t}$ or u_0 except possible by reducing $v_t(\eta, \bar{k})$ and $u_0(\eta, \bar{a})$. Thus, it only improves the objective value. The same is true for $(\bar{\eta}^-, k)$. We can therefore extend $v_0 = \bar{v}_0$ to $\bar{K}$.

□

4.6 Properties of optimal matchings

In this section, we discuss properties of optimizers for the primal (4.2) and dual problem (4.5).

First, we talk about positive assortativity of primal measures ϵ, λ.

Lemma 4.6.1 (Structure of wage functions)**.** *Let $b_L : [0, \bar{k}] \to \mathbf{R}$ be as introduced in section 4.1 then*

$$f_1(\eta_w, w; \eta_m, m) := b_L\left(z_L(\eta_w, w; \eta_m, m)\right) = b_L\left((1-\theta_L)w + \theta_L m\right)$$

is strictly supermodular in w, m.

Let $v : K \to \mathbf{R}$ be convex and non-decreasing in its second variable and supermodular. Then

$$f_2(\iota, a; \eta_t, t) = v(z_E(\iota, a; \eta_t, t))$$

is supermodular in a, t and supermodular in ι, t. The supermodularity is strict if v is strictly convex and strictly supermodular.

Proof. The proof is straightforward and mimics the proof of [13, Lemma 5, first claim]. For f_1, one simply has to take the right cross derivatives and see that they are (strictly) positive. For f_2, we approximate v by $v^i \in C^2(\bar{K})$ and follow the same argument. □

Lemma 4.6.2. *If $f(k_1, k_2) \geq 0$ is strictly submodular, then its zero set Z is strictly increasing.*

Proof. Suppose for contradiction that there exists $p = (k_1, k_2)$ and $p' = (k_1', k_2')$, both in Z such that $k_1 > k_1'$ and $k_2 < k_2'$. As f is submodular in k_1 and k_2, we have that

$$\begin{aligned} 0 &> f(k_1, k_2') + f(k_1', k_2) - f(k_1, k_2) - f(k_1', k_2') \\ &= f(k_1, k_2') + f(k_1', k_2) && p, p' \in Z \\ &\geq 0 && f \text{ is positive.} \end{aligned}$$

This is impossible and we conclude that Z is strictly increasing. □

The next lemma shows that, for a student, an increase in either the cognitive or the communication skill results in an increase in their teacher's cognitive skill. Moreover, an increase in workers' cognitive skill results in an increase in their manager's cognitive skill.

Lemma 4.6.3. *Let (u, v) be an optimizing pair for LP_*. Suppose v is strictly supermodular and strictly convex in its second variable. Let (ϵ, λ) be the measure that maximizes LP^*. Then $\epsilon(\iota, a; \eta_t, t)$ is positive assortative in ι, t and in a, t. Moreover, $\lambda(\eta_w, w; \eta_m, m)$ is always positive assortative in w, m.*

Proof. By Corollary 4.4.7, the support of ϵ is included in the zero set of

$$f(\iota, a; \eta_t, t) = u(\iota, a) + \frac{v(\eta_t, t)}{\rho_E \eta_t} - c_\delta b_E(z_E(\iota, a; \eta_t, t)) - v(z_E(\iota, a; \eta_t, t)) \geq 0.$$

By Lemma 4.6.1, f is strictly submodular in ι, t and in a, t. So by Lemma 4.6.3, its zero set is strictly increasing in ι, t and in a, t. Thus, the result follows for ϵ.

The proof for λ is similar, noting that by Corollary 4.4.7, the support of λ is included in the zero set of

$$g(\eta_w, w; \eta_m, m) = v(\eta_w, w) + \frac{v(\eta_m, m)}{\rho_L \eta_m} - b_L(\bar{z}_E(\eta_w, w; \eta_m, m)) \geq 0.$$

□

Lemma 4.6.4 (Endogenous distribution of adult skills)**.** *Let (ϵ, λ), be optimizers of LP^*.*

If α gives no mass to sets of Hausdorff dimension one, then there exists a function $\tau : A \to [0, \bar{k}]$ that associates a student's type (η, a) to the cognitive skill of teachers of students of this type such that

$$\pi_{\#}^{1,2,4}\epsilon = (\mathbb{1} \times \tau)_{\#}\alpha.$$

If ϵ is positive assortative in its first and last variables and in its second and last variables, the function τ is non-decreasing in both variables.

Moreover, if α is absolutely continuous, i.e. $\alpha(\iota, a) = \alpha^{ac}(\iota, a)d\iota da$ is given by a density $\alpha^{ac} \in L^1(A)$, then $\kappa(\eta, k) = \kappa^{ac}(\eta, k)d\eta dk$ is given by a related density $\kappa^{ac} \in L^1(K)$ satisfying

$$\alpha^{ac}(\eta, a) = \left((1-\theta_E) + \theta_E \frac{\partial}{\partial a}\tau(\eta, a)\right)\kappa^{ac}(\eta, (1-\theta_E)a + \theta_E \tau(\eta, a)$$

for a.e. $(\eta, a) \in A$.

Proof. The proof follows the proof of [13, Lemma 14, second and third claim]. As

$$f(\iota, a; \eta_t, t) = u(\iota, a) + \frac{v(\eta_t, t)}{\rho \eta_t} - cb_E(b(\iota, a; \eta_t, t)) - v(b(\iota, a; \eta_t, t))$$

is positive and is null on the support of ϵ, we have that

$$0 = f_a(\iota, a; \eta_t, t) = u_a(\iota, a_a) - (cb_E + v)(1-\theta_E)\big(\iota, (1-\theta_E)a + \theta_E t\big)$$

on the support of ϵ. As we assume that b_E is strictly convex and v is convex, we have that

$$\frac{\partial}{\partial k}(cb_E + v)\big(\eta, (1-\theta_E)a + \theta_E \tau(\eta, a)\big)$$

is invertible in its second variable if we fix its first variable, where $\tau(\eta, a)$ is the cognitive skill of a teacher of a student of type (η, a). Thus, τ is a well defined function except possibly on a $\mathcal{C}^1$-rectifiable set of Hausdorff dimension one, and if α is null on those sets,

$$\pi^{1,2,4}_{\#}\epsilon = (\mathbb{1} \times \tau)_{\#}\alpha$$

by Lemma 2.1.5. If ϵ is positive assortative between its first and last variables, then τ is increasing in its first variable and if ϵ is positive assortative between its second and last variables, then τ is increasing in its second variable.

Let $f(\eta, a) = (\eta, (1-\theta_E)a + \theta_E \tau(\eta, a))$ be the adult type of a student of type (η, a). This function is non-decreasing in both variables and pushes α forward to κ.

We have that

$$Df(\eta, a) = \begin{pmatrix} 1 & \theta_E \frac{\partial}{\partial \eta}\tau(\eta, a) \\ 0 & (1-\theta_E) + \theta_E \frac{\partial}{\partial a}\tau(\eta, a) \end{pmatrix}.$$

As τ is increasing in its second variable, $\det DF \geq (1-\theta_E)$. Thus, f is invertible, with inverse g having Lipschitz constant at most $\frac{1}{1-\theta_E}$.

We then have that for $K' \subset K$,

$$\kappa(K') = \alpha(g(K')).$$

Taking a $\mathcal{C}^1$-rectifiable set of Hausdorff dimension one shows that if α is null on those sets then so is κ. Moreover, taking K' to be an arbitrary set of measure zero shows that κ is absolutely continuous if α is.

As τ is defined on a compact set and is increasing, it is differentiable a.e. so the formula

$$\alpha^{ac}(\eta, a) = \left((1-\theta_E) + \theta_E \frac{\partial}{\partial a}\tau(\eta, a)\right) \kappa^{ac}(\eta, (1-\theta_E)a + \theta_E \tau(\eta, a)$$

follows from a simple change of variables. □

Theorem 4.6.5 (Positive assortative and unique optimizers). *Let (u, v) be an optimizing pair for LP_*. Suppose that v is strictly convex in k. The first derivative of v with respect to η is piecewise constant. The first derivative of v with respect to k is uniquely determined. The first order derivatives of u are uniquely determined. If α dominates an absolutely continuous measure whose support fills $\bar{A}$ then u is unique α-a.e.*

Proof. The proof follows the proof of [13, Theorem 15, fifth claim]. As v is strictly convex in k, it is continuous in k, has one-side derivatives in k that agrees except perhaps for countably many $k \in K$. Moreover, we proved that

$$v(\eta, k) = \bar{v}(\eta, k) = \max\{\tilde{v}_w(k), \eta\tilde{v}(k)\}$$

where $\tilde{v}(k) = \max\{\rho_L\tilde{v}_m(k), \rho_E\tilde{v}_t(k)\}$. So v is piecewise linear in η. Thus it is differentiable a.e.

Worker Let

$$g(\eta_w, w; \eta_w, w) = v(\eta_w, w) + \frac{v(\eta_m, m)}{\rho_L\eta_m} - b_L(z_L(\eta_w, w; \eta_m, m)) \geq 0.$$

The set of workers κ_w has full measure on $\pi^1(\text{spt }\lambda)$. As λ has full support on the zero set of g and $g \geq 0$, the first order necessary condition implies $g_{\eta_w} = g_w = 0$ i.e.

$$\frac{\partial}{\partial \eta_w} v(\eta_w, w) = 0$$
$$\frac{\partial}{\partial w} v(\eta_w, w) = (1-\theta_L)b_L'\left((1-\theta_L)w + \theta_L m\right).$$

Because b_L is strictly convex, there cannot be two such m unless v is not differentiable at (η_w, w). This proves that $\frac{\partial}{\partial w}v$ is uniquely determined on $\pi^1(\text{spt }\lambda) \cap \text{Dom }\frac{\partial}{\partial k}v$.

Manager The set of managers κ_m has full measure on $\pi^2(\text{spt }\lambda)$. As λ has full support on the zero set of g and $g \geq 0$ the first order necessary condition tells us that

$$\frac{\partial}{\partial \eta_m} g = \frac{\partial}{\partial m} g = 0$$

when g is differentiable. As $g_{\eta_m} = 0$, we have that $v_{\eta_m} = \frac{v}{\eta_m}$. Thus v_{η_m} is constant when it is differentiable.

As $g_m = 0$, we have

$$\frac{\partial}{\partial m} v = \rho_L\eta_m\theta_E b_L'((1-\theta_L)w + \theta_L m).$$

As in the worker's case, we can conclude that $\frac{\partial}{\partial m}v$ is uniquely determined on $\pi^1(\text{spt }\lambda) \cap \text{Dom }\frac{\partial}{\partial k}v$.

Teacher We can show that v_{η_t} is constant, in the same way we show that v_{η_m} is constant, by replacing g with

$$f(\iota, a; \eta_t, t) = u(\iota, a) + \frac{v(\eta_t, t)}{\rho_E\eta_t} - cb_E(z_E(\iota, a; \eta_t, t)) - v(z_E(\iota, a; \eta_t, t)).$$

To show that v_t is uniquely determined a.e., we use Proposition 4.2.5 (5). That proposition gives a recursive formula for $v_t(\eta_1, k_1)$ if $(\eta_1, k_1) \in \text{spt } \kappa_t \cap \text{Dom } v_k$ and $(\eta_2, k_2) \in \text{Dom } v_k$. The formula relates $v_t(\eta_1, k_1)$ and $v_t(\eta_2, k_2)$, where $(\eta_1, k_1) \in \pi^2(\text{spt } \epsilon)$ and (η_2, k_2) is the type of an adult whose teacher had type (η_1, k_1). The strict monotonicity we assumed for $v(\eta, \cdot)$ ensures that (η_2, k_2) is unique. Proposition 4.2.5 (5) then implies that the recursion ends after a finite number of students. Thus, v_t is uniquely determined.

Student By Lemma 4.2.3, u is convex and non-decreasing in k. This means that u is differentiable in k. The first order necessary condition for f gives us

$$u_a(\iota, a) = (1-\theta_E)\frac{\partial}{\partial k}\left(c_\delta b_E(\iota, k) + v(\iota, k)\right)|_{k=(1-\theta_E)a+\theta_E t}.$$

As $(c_\delta b_E + v)_k$ is strictly increasing in k, t is uniquely determined for κ-a.e. adult type. So u_a is uniquely determined for α-a.e. student type.

We don't know if u is differentiable in ι, so the first order condition reads $\pm f(\iota^\pm, a; \eta_t, t) \geq 0$, i.e.

$$(c_\delta b_E + v)_\iota(\iota^-, (1-\theta_E)a + \theta_E t) \geq u_\iota(\iota, a) \geq (c_\delta b_E + v)_\iota(\iota^+, (1-\theta_E)a + \theta_E t).$$

But as $c_\delta b_E + v$ is convex,

$$(c_\delta b_E + v)_\iota(\iota^-, (1-\theta_E)a + \theta_E t) \leq (c_\delta b_E + v)_\iota(\iota^+, (1-\theta_E)a + \theta_E t).$$

So,

$$u_\iota(\iota, a) = (c_\delta b_E + v)_\iota(\iota, (1-\theta_E)a + \theta_E t).$$

Next, we prove that u is unique. Suppose there exists another minimizer (v_0, u_0). As the derivative of u is unique for α-a.e. type of student, $u_0 = u + const$. The constant must vanish because the objective value is the same. □

When $\delta = 0$, we don't know if v is strictly convex in either variable. By Theorem 4.5.5, v is strictly convex in k if $c > 0$ or $\rho_E \eta \theta_E > 1$ for all η. If this is not the case, we can approximate v by forcing $c_\delta > 0$; for example, $c_\delta = \delta$. Thus, taking the limit when $\delta \to 0$, we know that there exists a minimizing v which is convex. We don't know if its derivative is uniquely determined anymore. We know that there exists an ϵ which is positively assortative between ι and t and between a and t, but there may be other optimal ϵ's.

Appendices

Appendix A

Technical Preleminary Results

In this appendix, we present proofs of prelemenary results that were too technical for inclusion in Chapter 2.

First, we present the proof of Lemma 2.1.6.

Proof of Lemma 2.1.6. As $\bar{K}$ is compact, the sup is always attained in the definition of g. Thus, it is possible to pick an arbitrary

$$k_a \in \arg\max_{k \in \bar{K}} f(a,k).$$[1]

First, we assume f is locally Lipschitz in a and uniformly Lipschitz in k and we prove that g is locally Lipschitz. Let $a_1 \in \bar{A}$. As f is locally Lipschitz in a_1, for every neighbourhood $a_1 \in U_{a_1} \subset \bar{A}$ f is Lipschitz in a within U_{a_1} with Lipschitz constant $C_{U_{a_1}}$. Thus, for all $a_2 \in U_{a_1}$, if $g(a_1) \geq g(a_2)$:

$$\begin{aligned}
|g(a_1) - g(a_2)| &= g(a_1) - g(a_2) \\
&= \sup_{k \in \bar{K}} f(a_1,k) - \sup_{k \in \bar{K}} f(a_2,k) \\
&\leq f(a_1,k_{a_1}) - f(a_2,k_{a_1}) \\
&\leq |f(a_1,k_{a_1}) - f(a_2,k_{a_1})| \\
&\leq C_{U_{a_1}} |a_1 - a_2| && \text{because } f \text{ is Lipschitz.}
\end{aligned}$$

By symmetry, we have the same bound when $g(a_1) < g(a_2)$. Thus g is locally Lipschitz, with same Lipschitz constant as f in its first variable.

Now, as for all $a_2 \in U_{a_1}$ we have

$$\begin{aligned}
\left| \frac{g(a_1) - g(a_2)}{a_1 - a_2} \right| &\leq C_{U_{a_1}} \\
&\leq \sup_{\substack{a \in U_{a_1} \\ k \in \bar{K}}} f_a(a,k)
\end{aligned}$$

we get the upper bound on $g'(a)$.

[1] This set might contain more then one point k_a.

For the lower bound on g' we will prove bound for one-sided derivatives:

$$\begin{aligned}\partial_{a^+} g(a) &\geq \inf_{a' \in U_a} \partial_{a^+} g(a') \\ &\geq \inf_{a' \in U_a} \lim_{\substack{a'' \to a' \\ a'' > a'}} \frac{g(a'') - g(a')}{a'' - a'} \\ &\geq \inf_{a' \in U_a} \lim_{\substack{a'' \to a' \\ a'' > a'}} \frac{f(a'', k_{a''}) - f(a', k_{a'})}{a'' - a'}.\end{aligned}$$

As $k_{a''} = \arg\max_k f(a'', k)$, we have

$$\begin{aligned}\partial_{a^+} g(a) &\geq \inf_{a' \in U_a} \lim_{\substack{a'' \to a' \\ a'' > a'}} \frac{f(a'', k_{a'}) - f(a', k_{a'})}{a'' - a'} \\ &\geq \inf_{\substack{a' \in U_a \\ k \in \bar{K}}} \lim_{\substack{a'' \to a' \\ a'' > a'}} \frac{f(a'', k) - f(a', k)}{a'' - a'} \\ &\geq \inf_{\substack{a' \in U_a \\ k \in \bar{K}}} \partial_{a^+} f(a', k).\end{aligned}$$

As f is locally Lipchitz in a, and g is locally Lipchitz they are differentiable a.e. and at their differentiability point $\partial_{a^+} f(a', k) = f_a(a', k)$ and $\partial_{a^+} g(a) = g'(a)$. We conclude that $g'(a) \geq \inf_{\substack{a' \in U_a \\ k \in \bar{K}}} f_a(a', k)$.

Finally, we show that if f is locally semi-convex in a and uniformly semi-convex in k, then $g(a)$ is locally semi-convex. First, note that as f is locally semi-convex in a, for all $a \in \bar{A}$ for every neighbourhood $a \in U_a$ there exists a constant

$$C_1 = - \inf_{k \in \bar{K},\ a' \in U_a} f_{aa}(a', k)$$

such that

$$f(a, k) + C_1 \frac{|a|^2}{2}$$

is convex with respect to a in U_a. Thus

$$g(a) + C_1 \frac{|a|^2}{2} = \sup_{k \in \bar{K}} \left\{ f(a, k) + C_1 \frac{|a|^2}{2} \right\}$$

is a supremum of convex functions with respect to a in U_a, thus it is convex with respect to a in U_a and we have the right bound on g''. □

Now, we present the proof of Lemma 2.2.1.

Proof of Lemma 2.2.1. (a) Let $\{ f_j \}$ be a sequence of real functions on K such that $f_j \to f$ in the sense of uniform convergence on compact subsets of K. Then for all i, $\|f_j - f\|_{L^\infty([0, \bar{k} - \frac{1}{i}])} \to 0$ as $j \to \infty$ and therefore $d_K(f_j, f) \to 0$.

Let $\{ f_j \}$ be a sequence of real functions on K such that $d_K(f_j, f) \to 0$. This means that for all i, $\|f_j - f\|_{L^\infty([0, \bar{k} - \frac{1}{i}])} \to 0$. Let $U \subset K$ be a compact subset. Then there exists an $i \in \mathbf{N}$ such that $U \subset \left[0, \bar{k} - \frac{1}{i}\right]$ and we conclude that $\|f_j - f\|_{L^\infty(U)} \to 0$.

(b) Suppose $\mu_j \to \mu$ in the weak-$*$ topology. Let $\epsilon > 0$. We will separate $K = K_\epsilon \cup K_{-\epsilon}$ where $K_\epsilon = [0, \bar{k} - \epsilon]$ and $K_{-\epsilon} = (\bar{k} - \epsilon, \bar{k})$.

On K_ϵ we have

$$\begin{aligned}\left| \int_{K_\epsilon} f_i d\mu_i - f d\mu \right| &= \left| \int_{K_\epsilon} f_i d\mu_i - f d\mu_i + f d\mu_i - f d\mu \right| \\ &\leq \int_{K_\epsilon} |f_i d\mu_i - f d\mu_i| + \left| \int_{K_\epsilon} f d\mu_i - f d\mu \right|.\end{aligned}$$

The second term goes to zero as μ_i weak-$*$ converges to μ and f is bounded and continuous. By (a), on K_ϵ, $\|f_i - f\|_{L^\infty(K_\epsilon)} \to 0$. Thus we have that

$$\begin{aligned}\int_{K_\epsilon} |f_i d\mu_i - f d\mu_i| &\leq \|f_i - f\|_{L^\infty(K_\epsilon)} \|\mu_i\|_{TV} \\ &\leq \|f_i - f\|_{L^\infty(K_\epsilon)} && \text{as } \mu_i \text{ is a probability measure} \\ &\to 0.\end{aligned}$$

On $K_{-\epsilon}$ we have

$$\left| \int_{K_{-\epsilon}} f_i d\mu_i - d\mu \right| \leq |f_i|_{L^\infty(K_{-\epsilon})} \mu_i(K_{-\epsilon}) + |f|_{L^\infty(K_{-\epsilon})} \mu(K_{-\epsilon}).$$

As f_i and f are uniformly bounded, the result follows from hypothesis (2.2).

(c) Any non-decreasing convex function of K is Lipschitz continuous on compact subsets of K.

Consider a compact subset $C \subset K = [0, \bar{k})$ such that $\bar{k} - \sup_{x \in C} x > \epsilon$. Consider a sequence $\{ f_i \}$ of non-decreasing, and convex functions $f_i : K \to \mathbf{R}$ bounded by b. Each f_i is Lipschitz on C with Lipschitz constant $\mathrm{Lip}_C f_i < \frac{b}{\epsilon}$. Then $\{ f_i \}$ is equi-Lipschitz and by the Arzelà-Ascoli theorem, there exists a subsequence that converges uniformly on C.

In order to conclude that $\{ f_i \}$ has a converging subsequence when metrized by d_K, we need to have a subsequence that converges independently of the compact subset $C \subset K$. Let $C_n = [0, \bar{k} - \frac{\bar{k}}{n}] \subset K$. For every compact subset $C \subset K$, there exists an n such that $C \subset C_n$.

First, note that as C_1 is compact, there exists a subsequence of $\{ f_i \}$ that converges uniformly on C_1. As C_2 is compact, there exists a subsequence of the subsequence on C_1 that converges uniformly on C_2. Similarly, as C_n is compact there exists a subsequence of the subsequence on C_{n-1} that converges uniformly on C_n.

We will use a diagonal argument to create a subsequence of $\{ f_i \}$ that converges on all C_n. Let f_{i_n} be the n-th element of the converging subsequence on C_n. Then for all n, $\{ f_{i_j} \}_{j=n}^\infty$ converges uniformly on C_n because it is a subsequence of the converging subsequence on C_n. Therefore $\{ f_{i_j} \}$ converges on C_n for all n.

□

Appendix B

Finite Time Horizon for Complete Information Model

One way to get a solution for the complete information model (3.9) is to first consider the same problem, for a finite time horizon. Then, (3.9) can be seen as the limit when this time horizon increases. Indeed, (3.9) is an infinite time horizon problem. In this appendix, we introduce the corresponding finite time horizon model.

For the finite time horizon model, we fix $T \in \mathbf{N}$ and solve until generation T. Let v_{T+1} be a positive, twice differentiable, non-decreasing, and convex function that represents the estimated wage function for the students of the last step. We assume v_{T+1} is bounded by a constant h. Let κ_1 be an absolutely continuous Borel probability measure that represents the initial distribution for adults. The goal is to optimize the society's total production over T generations through separating adults by profession, i.e. worker, manager and teacher, and matching workers to managers and students to teachers within a generation. That is, we are seeking measures

$$\{\,\epsilon_i\,\}_{i=1}^{T} \qquad \text{and} \qquad \{\,\lambda_i\,\}_{i=1}^{T}\,,$$

where ϵ_i represents the matching between students and teachers and λ_i represents the labour matching for generation i. These measures have to satisfy the following constraints:

$$\begin{aligned}
\pi_{1\#}\epsilon_i &= \alpha && i = 1, \ldots, T && \text{i.e. the distribution of students is known,}\\
\pi_{1\#}\lambda_1 + \frac{1}{N'}\pi_{2\#}\lambda_1 + \frac{1}{N}\pi_{2\#}\epsilon_1 &= \kappa_1 && && \text{i.e. the first distribution of adults is known and}\\
\pi_{1\#}\lambda_i + \frac{1}{N'}\pi_{2\#}\lambda_i + \frac{1}{N}\pi_{2\#}\epsilon_i &= z_{E\#}\epsilon_{i-1} && i = 2, \ldots, T && \text{i.e. distributions of adults are induced.}
\end{aligned}$$

The last $T-1$ constraints represent the fact that the distribution of adults for generation i is induced by the education matching of generation $i-1$.

The goal is to optimize the total productivity of the society plus the expected wage for students of the last generation. We include a discount factor $e^{-\beta}$ (take $\beta = 0$ to remove it) to represent the fact that immediate gain is more valuable than future gain. Thus, the proposed finite horizon complete information model is:

$$\begin{aligned}
C_T = \sup_{(\epsilon_i,\lambda_i)_{i=1}^T} \Bigg(& e^{-\beta(T+1)} \int_{A\times K} v_{T+1}\circ z_E d\epsilon_T + \sum_{i=1}^T e^{-\beta i} \int_{A\times K} cb_E \circ z_E d\epsilon_i \\
& + \sum_{i=1}^T e^{-\beta i} \int_{K\times K} b_L \circ z_L d\lambda_i \Bigg)
\end{aligned} \tag{B.1}$$

$$\begin{aligned}
\text{s.t.} \qquad\qquad & \pi_{1\#}\epsilon_i = \alpha && \text{for } i=1,\dots,T,\\
& \pi_{1\#}\lambda_i + \frac{1}{N'}\pi_{2\#}\lambda_i + \frac{1}{N}\pi_{2\#}\epsilon_i = z_{E\#}\epsilon_{i-1} && \text{for } i=1,\dots,T,
\end{aligned}$$

where, $\epsilon_0 = (\mathbb{1}\times\mathbb{1})_{\#}\kappa_1$.

Note that (B.1) depends on α, β, κ_1, v_{T+1}, $b_{E/L}$, $\theta_{E/L}$ and $z_{E/L}$.

The dual of this problem is given by:

$$C_T^* = \inf_{(u_i,v_i)_{i=1}^T} \int_K e^{-\beta} v_1(k) d\kappa_1(k) + \sum_{i=1}^T e^{-\beta i} \int_A u_i(s) d\alpha(s) \tag{B.2}$$

$$\begin{aligned}
\text{s.t.} \quad & u_i(s) + \frac{1}{N} v_i(t) \geq cb_E(z_E(s,t)) + e^{-\beta} v_{i+1}(z_E(s,t)) && i=1,\dots,T \quad \text{stability of education market}\\
& v_i(w) + \frac{1}{N'} v_i(m) \geq b_L(z_L(w,m)) && i=1,\dots,T \quad \text{stability of labour market.}
\end{aligned}$$

Strong duality, $C_T = C_T^*$, is proven in Section B.1. Section B.2 shows that the infinite horizon complete information model can be seen as the limit, as the number of generations increases, of the finite time horizon model C_T.

B.1 Proof of Duality, Complete Information, Finite Time Horizon

We'll prove that the optimal value for the primal is equal to the optimal value for the dual in the finite horizon complete information model. This proof is separated into two parts. First, we will show the standard duality inequality:

$$C_T \leq C_T^*. \tag{B.3}$$

Then, we will show that there is no duality gap. Both proofs are adaptations of the duality proofs of [13].

If we restrict the minimization (B.2) to continuous bounded wage functions (u_i, v_i) the inequality $C_T \leq C_T^*$ is straightforward. However, it is not at all clear how to show the infimum is attained in this class. Instead we need to establish duality in a larger class of functions to be sure the infimum is attained, and then appeal to the asymptotic analysis of Erlinger et al to conclude the minimizer likely belongs to the smaller class. Since the larger class includes unbounded functions, a careful argument is needed to ensure convergence of certain integrals in the proof. Fortunately, as shown in [13, Proposition 8] the doubling condition on α (3.7) guarantees that $v_i \in L^1(\bar{K}, z_{\#}\epsilon_{i-1})$, which is sufficient to establish the inequality desired.

Proposition B.1.1. *Let $A = K = [0, \bar{k})$, $\beta \geq 0$, $T \in \mathbf{N}$ and $v_{T+1} \in L^1\left(\bar{K}, \alpha\right)$. Let κ_1 be a Borel probability measure on $\bar{K}$ and let α be a Borel probability measure on A. Define $b_{E/L}$ and $z_{E/L}$ as in subsection 2.5. Suppose $\bigoplus_{i=1}^{T}\binom{\epsilon_i}{\lambda_i}$ is a feasible candidate for the primal problem C_T and $\bigoplus_{i=1}^{T}\binom{u_i}{v_i}$ is feasible for the dual problem C_T^* such that $u_i \in L^1\left(\bar{A}, \alpha\right)$ for all i, $v_1 \in L^1\left(\bar{K}, \kappa_1\right)$ and $v_i = v_i^c + v_i^g$ differs from a bounded continuous function v_i^c (c for continuous) by a non-decreasing function v_i^g (g for growth). If $v_i \in L^1\left(\bar{K}, z_{E\#}\epsilon_{i-1}\right)$ for $i = 2, \ldots, T+1$ we have*

$$
\begin{aligned}
& e^{-\beta T}\int_{A\times K} v_{T+1}\circ z_E d\epsilon_T + \sum_{i=1}^{T} e^{-\beta i}\int_{A\times K} cb_E\circ z_E d\epsilon_i \\
& + \sum_{i=1}^{T} e^{-\beta i}\int_{K\times K} b_L\circ z_L d\lambda_i \\
& \leq \int_K e^{-\beta} v_1 d\kappa_1 + \sum_{i=1}^{T} e^{-\beta i}\int_A u_i d\alpha.
\end{aligned}
$$

If α satisfies the doubling condition (3.7) then $v_i \in L^1\left(\bar{K}, z_{E\#}\epsilon_{i-1}\right)$ for $i = 2, \ldots, T$.

Proof. Suppose $\bigoplus_{i=1}^{T}\binom{\epsilon_i}{\lambda_i}$ is feasible for the primal problem C_T and $\bigoplus_{i=1}^{T}\binom{u_i}{v_i}$ is feasible for the dual problem C_T^*. By multiplying the stability constraints for education sector by $e^{-\beta i}$, integrating over ϵ_i and summing over i, we have:

$$
\begin{aligned}
\infty > & \sum_{i=1}^{T} e^{-\beta i}\int_A u_i d\alpha - \sum_{i=1}^{T} e^{-\beta i}\int_A cb_E\circ z_E d\epsilon_i \\
\geq & \sum_{i=1}^{T} e^{-\beta(i+1)}\int_K v_{i+1}\circ z_E d\epsilon_i - \sum_{i=1}^{T} e^{-\beta i}\int_K \frac{1}{N} v_i(t) d\epsilon_i(s,t). \qquad \text{(B.4)}
\end{aligned}
$$

Multiplying constraints that represent the prescribed distribution for adults' skills by $e^{-\beta i} v_i$ and summing them over i, we get:

$$
\begin{aligned}
& \int_{A\times K} e^{-\beta} v_1 d\kappa_1 + \sum_{i=2}^{T}\int_{A\times K} e^{-\beta i} v_i\circ z_E d\epsilon_{i-1} - \sum_{i=1}^{T} e^{-\beta i}\int_{A\times K}\frac{1}{N} v_i(t) d\epsilon_i(s,t) \qquad \text{(B.5)} \\
= & \sum_{i=1}^{T}\int_{K\times K} e^{-\beta i}\left(v_i(w) + \frac{1}{N'} v_i(m)\right) d\lambda_i(w,m) \\
\geq & \sum_{i=1}^{T}\int_{K\times K} e^{-\beta i} b_L\circ z_L d\lambda_i > 0,
\end{aligned}
$$

where the inequality follows from statibility of the labour market. Adding

$$
\int_{A\times K} e^{-\beta} v_1 d\kappa_1 - e^{-\beta(T+1)}\int_K v_{T+1}\circ z_E d\epsilon_T
$$

to (B.4) makes it equal to (B.5). This is where we need the integrability of v_i with respect to $z_{E\#}\epsilon_{i-1}$. Indeed, as (B.4) is bounded from above and (B.5) is bounded from below, if $v_i \in L^1\left(z_{E\#}\epsilon_{i-1}\right)$ then

$$
\sum_{i=1}^{T} e^{-\beta i}\frac{1}{N}\int_{A\times K} v_i(t) d\epsilon_i(s,t)
$$

also has to be finite. Thus, we have the desired inequality:

$$C_t \leq C_T^*.$$

To complete the proof, we need to show that the doubling condition (3.7) of α implies that $v_i \in L^1\left(\bar{A}, z_{E\#}\epsilon_{i-1}\right)$. Suppose each $v_i = v_i^c + v_i^g$ differs from a bounded continuous function v_i^c (c for continuous) by a non-decreasing function v_i^g (g for growth). As v_i^c is bounded, it is integrable.

The stability constraint for education markets implies that

$$\begin{aligned} u_i(s) &\geq e^{-\beta} v_{i+1}\left(z_E(s,t)\right) - \frac{1}{N} v_i(t) + c b_E(z_E(s,t)) \\ &\geq e^{-\beta} v_{i+1}\left(z_E(s,t)\right) - \frac{1}{N} v_i(t). \end{aligned}$$

Setting $s = t := k$ and summing over all i, we get:

$$\sum_{i=1}^{T} u_i(k) \geq e^{-\beta} v_{t+1}(k) + \sum_{i=2}^{T} \left(e^{-\beta} - \frac{1}{N}\right) v_i(k) - \frac{1}{N} v_1(k).$$

Thus as long as $e^{-\beta} - \frac{1}{N} \geq 0$, we get

$$\sum_{i=2}^{T} v_i(k) \leq \frac{1}{e^{-\beta} - \frac{1}{N}} \left(\frac{1}{N} v_1(k) - e^{-\beta} v_{T+1}(k) + \sum_{i=1}^{T} u_i(k)\right). \tag{B.6}$$

Now, using the layer-cake representation, we get:

$$\begin{aligned} \sum_{i=1}^{T} \int_K v_i^g(k) dz_{E\#}\epsilon_{i-1}(k) &= \sum_{i=1}^{T} \int_0^\infty \kappa_i \left(\left(v_i^g\right)^{-1} [y, \infty)\right) dy \qquad \text{layer-cake representation} \\ &= \sum_{i=1}^{T} \int_0^\infty \kappa_i \left[\bar{a} - \left(\bar{a} - \left(v_i^g\right)^{-1}(y)\right), \bar{a}\right] dy. \end{aligned}$$

Thus by [13, Lemma 14] and because α satisfies the doubling condition (3.7):

$$\begin{aligned} \sum_{i=1}^{T} \int_K v_i^g(k) dz_{E\#}\epsilon_{i-1}(k) &\leq \sum_{i=1}^{T} \int_0^\infty \alpha \left[\bar{k} - \frac{1}{1-\theta}\left(\bar{a} - \left(v_i^g\right)^{-1}(y)\right), \bar{a}\right] dy \\ &\leq \sum_{i=1}^{T} \int_0^\infty C^d \alpha \left[\bar{k} - \frac{1}{2^d} \frac{1}{1-\theta}\left(\bar{a} - \left(v_i^g\right)^{-1}(y)\right), \bar{a}\right] dy \qquad \text{by (3.7).} \end{aligned}$$

Letting $d = -\ln(1-\theta)/\ln(2)$ and using the layer-cake representation again, we have

$$\begin{aligned}\sum_{i=1}^{T}\int_K v_i^g(k)dz_{E\#\epsilon_{i-1}}(k) &\leq \sum_{i=1}^{T} C^{-\ln(1-\theta)/\ln(2)}\int_K v_i^g(y)d\alpha(y)\\ &\leq \sum_{i=1}^{T}\frac{C^{-\ln(1-\theta)/\ln(2)}}{e^{-\beta}-\frac{1}{N}}\int_K\left(\frac{1}{N}v_1\right.\\ &\qquad \left.-e^{-\beta}v_{T+1}+\sum_{i=1}^{T}u_i-\sum_{i=1}^{T}v_i^c\right)d\alpha \qquad \text{by (B.6)}\\ &< \infty.\end{aligned}$$

□

To prove the other inequality, we use Lemma 2.4.1.

Proposition B.1.2 (No duality gap)**.** *Let $\beta \geq 0$, $T \in \mathbf{N}$, $v_{T+1} : K \to \mathbf{R}^+$ and $A = K = [0, \bar{k}]$. Let κ_1 be a Borel probability measure on $\bar{K}$ and let α be a Borel probability measure on $\bar{A}$ that satisfies the doubling condition* (3.7)*. Define $b_{E/L}$ and $z_{E/L}$ as in subsection 2.5. The optimal value of the primal finite horizon complete information problem C_T is the same as the optimal value of the dual finite horizon complete information problem C_T^*.*

Proof. To prove that there is no duality gap, we will apply Lemma 2.4.1 with

$$\begin{aligned}\mathring{A} &= \bigoplus_{i=1}^{T}\left(C(\bar{A})\oplus C(\bar{K})\right),\\ \mathcal{B} &= \bigoplus_{i=1}^{T}\left(C(\bar{A}\times\bar{K})\oplus C(\bar{K}\times\bar{K})\right), \qquad \text{so}\\ \mathring{A}^* &= \bigoplus_{i=1}^{T}\Gamma(\bar{A})\oplus\Gamma(\bar{K}),\\ \mathcal{B}^* &= \bigoplus_{i=1}^{T}\Gamma(\bar{A}\times\bar{K})\oplus\Gamma(\bar{K}\times\bar{K}),\end{aligned}$$

where $\Gamma(X)$ is the set of Borel measures on X. Now, we define:

$$\begin{aligned}
\varphi : & \mathring{A} \to \mathbf{R} \cup \{\infty\} \\
& \bigoplus_{i=1}^{T} \begin{pmatrix} u_i \\ v_i \end{pmatrix} \mapsto \int_K e^{-\beta} v_1(k) d\kappa_1(k) + \sum_{i=1}^{T} e^{-\beta i} \int_A u_i(s) d\alpha(s); \\
\varphi^* : & \mathring{A}^* \to \mathbf{R} \cup \{\infty\} \\
& \bigoplus_{i=1}^{T} \begin{pmatrix} \mu_i \\ \nu_i \end{pmatrix} \mapsto \begin{cases} 0 & \text{if } \mu_i = \alpha \ \forall i = 1, \dots, T, \ \nu_1 = \kappa_1 \text{ and } \nu_i = 0 \ \forall i = 2, \dots, T \\ +\infty & \text{otherwise;} \end{cases} \\
\phi : & \mathcal{B} \to \mathbf{R} \cup \{\infty\} \\
& \bigoplus_{i=1}^{T} \begin{pmatrix} \tilde{u}_i \\ \tilde{v}_i \end{pmatrix} \mapsto \begin{cases} & \text{if } \tilde{u}_i(s,t) \geq e^{-\beta i)} c b_E(z_E(s,t)) \ \forall i = 1, \dots, T-1, \\ 0 & \tilde{u}_T(s,t) \geq e^{-\beta(T+1)} v_{T+1}(z_E(s,t)) + e^{-\beta i} c b_E(z_E(s,t)) \\ & \text{and } \tilde{v}_i(w,m) \geq e^{-\beta i} b_L(z_L(w,m)) \ \forall i = 1, \dots, T \\ +\infty & \text{otherwise;} \end{cases} \\
\phi^* : & \mathcal{B}^* \to \mathbf{R} \cup \{\infty\} \\
& \bigoplus_{i=1}^{T} \begin{pmatrix} \epsilon_i \\ \lambda_i \end{pmatrix} \mapsto \begin{cases} e^{-\beta(T+1)} \epsilon_T(v_{T+1}(z_E)) + & \\ \Sigma_{i=1}^{T} e^{-\beta i} \epsilon_i(c b_E(z_E)) + e^{-\beta i} \lambda_i(b_L(z_L)) & \text{if } \epsilon_i \leq 0 \text{ and } \lambda_i \leq 0 \ \forall i \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

It is easy to check that φ^* and ϕ^* are Legendre transforms of φ and ϕ respectively.

Let

$$H\left(\bigoplus_{i=1}^{T} \begin{pmatrix} u_i \\ v_i \end{pmatrix}\right)((s,t),(w,m)) = \bigoplus_{i=1}^{T-1} \begin{pmatrix} e^{-\beta i} u_i(s) + \frac{1}{N} e^{-\beta i} v_i(t) - e^{-\beta i} v_{i+1}(z_E(s,t)) \\ e^{-\beta i} v_i(w) + e^{-\beta i} \frac{1}{N'} v_i(m) \end{pmatrix} + \begin{pmatrix} e^{-\beta T} u_T(s) + \frac{1}{N} e^{-\beta T} v_T(t) \\ e^{-\beta T} v_T(w) + e^{-\beta T} \frac{1}{N'} v_T(m) \end{pmatrix}$$

and

$$H^*\left(\bigoplus_{i=1}^{T} \begin{pmatrix} \epsilon_i \\ \lambda_i \end{pmatrix}\right) = \begin{pmatrix} e^{-\beta} \pi_{1\#} \epsilon_1 \\ \frac{e^{-\beta}}{N} \pi_{2\#} \epsilon_1 + e^{-\beta} \pi_{1\#} \lambda_1 + \frac{e^{-\beta}}{N'} \pi_{2\#} \lambda_1 \end{pmatrix} + \bigoplus_{i=2}^{T} \begin{pmatrix} e^{-\beta i} \pi_{1\#} \epsilon_i \\ e^{-\beta i} \frac{1}{N} \pi_{2\#} \epsilon_i + e^{-\beta i} \pi_{1\#} \lambda_i + e^{-\beta i} \frac{1}{N'} \pi_{2\#} \lambda_i - e^{-\beta i} z_{E\#} \epsilon_{i-1} \end{pmatrix}.$$

The operator H^* is the adjoint of H. With this construction,

$$\inf_{\bigoplus_{i=1}^{T} \left(\begin{smallmatrix} u_i \\ v_i \end{smallmatrix}\right) \in \mathring{A}} \varphi\left(\bigoplus_{i=1}^{T} \left(\begin{smallmatrix} u_i \\ v_i \end{smallmatrix}\right)\right) + \phi\left(H \bigoplus_{i=1}^{T} \left(\begin{smallmatrix} u_i \\ v_i \end{smallmatrix}\right)\right)$$

is the restriction to C_T^* where the infimum is taken over continuous functions, thus

$$C_T^* \leq \inf_{\bigoplus_{i=1}^{T} \left(\begin{smallmatrix} u_i \\ v_i \end{smallmatrix}\right) \in \mathring{A}} \varphi\left(\bigoplus_{i=1}^{T} \left(\begin{smallmatrix} u_i \\ v_i \end{smallmatrix}\right)\right) + \phi\left(H \bigoplus_{i=1}^{T} \left(\begin{smallmatrix} u_i \\ v_i \end{smallmatrix}\right)\right).$$

Similarly,

$$C_T = \max_{\bigoplus_{i=1}^T \binom{\epsilon_i}{\lambda_i}} -\varphi^*\left(H^*\binom{\epsilon_i}{\lambda_i}\right) - \phi^*\left(-\binom{\epsilon_i}{\lambda_i}\right).$$

To apply Lemma 2.4.1, we need to show that ϕ is continuous and real-valued at some point in $H(\text{Dom }\varphi)$. Let $\bigoplus_{i=1}^T \binom{u}{v} \in \text{Dom }\varphi$. Then, $v_1 \in L^1(K, \kappa_1)$ and $u_i \in L^1(A, \alpha)$ which means that the objective function of C_T^* is finite at $\bigoplus_{i=1}^T \binom{u}{v} \in \text{Dom }\varphi$. Moreover, $\phi\left(H \bigoplus_{i=1}^T \binom{u}{v}\right) = 0$ if $\bigoplus_{i=1}^T \binom{u}{v}$ is feasible for C_T^*. Thus the conditions are achieved because the conditions of being feasible and finite for $LP_*(\delta)$ are open and there exist such functions, for example:

$$\bigoplus_{i=1}^T \begin{pmatrix} u \\ v \end{pmatrix} = \bigoplus_{i=1}^T \begin{pmatrix} b_L(\bar{k}) + b_E(\bar{k}) \\ b_L(\bar{k}) \end{pmatrix}.$$

Thus, by Lemma 2.4.1,

$$C_T^* \leq C_T.$$

□

Corollary B.1.3. *Fix $c, \theta, \theta', N, N'$. Let $\beta \geq 0$, $T \in \mathbf{N}$, $v_{T+1} : \bar{K} \to \mathbf{R}^+$ and $A = K = (0, \bar{k})$. Let κ_1 be a Borel probability measure on $\bar{K}$ and let α be a Borel probability measure on $\bar{A}$ that satisfies the doubling condition* (3.7)*. Define $b_{E/L}$ and $z_{E/L}$ as in subsection 2.5. A sequence of feasible measures $(\epsilon_i, \lambda_i)_{i=1}^T$ maximizes* (B.1) *if there exists $(u_i, v_i)_{i=1}^T$ that satisfy the constraints of* (B.2) *such that*

$$e^{-\beta(T+1)}\epsilon_T\left(v_{T+1} \circ z_E\right) + \sum_{i=1}^T e^{-\beta i}\left(c\epsilon_i(b_E \circ z_E) + \lambda_i(b_L \circ z_L)\right) = e^{-\beta}\kappa_1(v_1) + \sum_{i=1}^T e^{-\beta i}\alpha(u_i)$$

Of course, we can prove, as in Propositions 3.3.1 and 3.3.5, existence of primal and dual optimizers for the finite time horizon model.

B.2 From Finite to Infinite Time Horizon

To represent a countable number of generations, we consider what happens when $T \to \infty$. To ensure that C_∞ is well defined in that case, we verify that the limit $\lim_{T\to\infty} C_T$ exists and is independent of our choice of v_{T+1}.

Proposition B.2.1. *Fix $\beta \in (0, 1]$, $\theta, \theta' \in (0, 1)$, $N, N' > 1$ and a probability measure α on A. Define $z_{E/L}$ and $b_{E/L}$ as in subsection 2.5. For any absolutely continuous probability measure $\kappa_1 \in \mathcal{P}(A)$ and any sequence of non-decreasing, convex, bounded functions $\{v_{T+1}\}$, such that $v_{T+1} < h$ for all T we define:*

$$C_T(\kappa_1, v_{T+1}) = C_T.$$

The limit of $C_T(\kappa_1, v_{T+1})$ when $T \to \infty$ satisfies

$$0 \leq \lim_{T\to\infty} C_T(\kappa_1, v_{T+1}) = C(\kappa_1) < C < \infty$$

where $C(\kappa_1)$ is the solution to (3.9) *and is independent of $\{v_{T+1}\}$ and C is independent of κ_1 and $\{v_{T+1}\}$.*

Proof. Since ϵ_i and λ_i are supported on $A \times K = K \times K = [0, \bar{k}) \times [0, \bar{k})$,

$$\int_{A\times K} v_{T+1} \circ z_E d\epsilon_T \leq \max v_{T+1} \leq h \qquad \text{as } v_{T+1} \text{ is bounded by } h, \tag{B.7}$$
$$\int_{A\times K} cb_E \circ z_E d\epsilon_i \leq c\bar{b}_E,$$
$$\int_{K\times K} b_L \circ z_L d\lambda_i \leq \bar{b}_L.$$

Thus, by the definition of C_T and using the formula for a geometric series, for $T \geq 1$,

$$\begin{aligned} C_T(\kappa_1, v_{T+1}) &< he^{-\beta(T+1)} + \frac{e^{-\beta}}{1-e^{-\beta}} \left(c\bar{b}_E + \bar{b}_L\right) \\ &\leq he^{-\beta} + \frac{e^{-\beta}}{1-e^{-\beta}} \left(c\bar{b}_E + \bar{b}_L\right) < C \end{aligned}$$

for C independent of κ_1 and v_{T+1}.

The only thing that is left to show is that the limit $\lim_{T\to\infty} C_T$ exists and is independent of v_{T+1}. To show this, we introduce a modified version of C_T:

$$\begin{aligned} \widetilde{C_T} = \sup_{(\epsilon_i,\lambda_i)_{i=1}^T} &\sum_{i=1}^T e^{-\beta i} \int_{A\times K} cb_E \circ z_E d\epsilon_i + \sum_{i=1}^T e^{-\beta i} \int_{K\times K} b_L \circ z_L d\lambda_i \\ \text{s.t.} \quad & \pi_{1\#}\epsilon_i = \alpha \qquad i = 1, \dots, T \\ & \pi_{1\#}\lambda_1 + \frac{1}{N'}\pi_{2\#}\lambda_1 + \frac{1}{N}\pi_{2\#}\epsilon_1 = \kappa_1 \\ & \pi_{1\#}\lambda_i + \frac{1}{N'}\pi_{2\#}\lambda_i + \frac{1}{N}\pi_{2\#}\epsilon_i = z_{E\#}\epsilon_{i-1} \qquad i = 2, \dots, T. \end{aligned}$$

First, note that $\widetilde{C_T} \leq C_T$. Let $(\epsilon_i^n, \lambda_i^n)_{i=1}^T$ be an approximate optimizer for $\widetilde{C_T}$. In this case, $(\epsilon_i^n, \lambda_i^n)_{i=1}^T$ satisfies the constraints for C_T and:

$$\begin{aligned} \widetilde{C_T} &= \underline{\lim}_{n\to\infty} \sum_{i=1}^T e^{-\beta i} \int_{A\times K} cb_E \circ z_E d\epsilon_i^n + \sum_{i=1}^T e^{-\beta i} \int_{K\times K} b_L \circ z_L d\lambda_i^n \\ &\leq \underline{\lim}_{n\to\infty} e^{-\beta(T+1)} \int_{A\times K} v_{T+1} \circ z_E d\epsilon_T^n + \sum_{i=1}^T e^{-\beta i} \int_{A\times K} cb_E \circ z_E d\epsilon_i^n \\ &\qquad + \sum_{i=1}^T e^{-\beta i} \int_{K\times K} b_L \circ z_L d\lambda_i^n \\ &\leq \lim_{n\to\infty} C_T = C_T. \end{aligned}$$

Thus $\widetilde{C_T}$ is also bounded from above. It is also increasing in T, so $\lim_{T\to\infty} \widetilde{C_T}$ exists.

Secondly, note that $C_T - \widetilde{C_T} \leq e^{-\beta(T+1)}h$. Let $(\epsilon_i^n, \lambda_i^n)_{i=1}^T$ be an approximate optimizer for C_T.

Then,

$$\begin{aligned}C_T &= \lim_{n\to\infty} e^{-\beta(T+1)} \int_{A\times K} v_{T+1}\circ z_E d\epsilon_T^n + \sum_{i=1}^{T} e^{-\beta i}\int_{A\times K} cb_E\circ z_E d\epsilon_i^n \\ &\quad + \sum_{i=1}^{T} e^{-\beta i}\int_{K\times K} b_L\circ z_L d\lambda_i^n \\ &\le e^{-\beta(T+1)}h + \widetilde{C_T}.\end{aligned}$$

Thus, $\lim_{T\to\infty} C_T = \lim_{T\to\infty}\widetilde{C_T}$.

As $\lim_{T\to\infty}\widetilde{C_T}$ exists and is independent of v_{T+1}, the same is true of $\lim_{T\to\infty} C_T$. □

Now, we'll show that the optimizers $(u_i^T, v_i^T)_i$ of (B.2) converge to the optimizer of (3.10). To do so, we'll use Γ-convergence. First, we need to introduce some notation. We will denote the set of feasible wages

$$\mathcal{F} = \left\{ (u_i, v_i)_{i=1}^\infty \,\middle|\, \begin{array}{c} u_i{:}A\to\mathbf{R}, v_i{:}K\to\mathbf{R} \text{ bounded, non-decreasing, and convex} \\ u_i(s)+\frac{1}{N}v_i(t)\ge cb_E(z_E(s,t))+e^{-\beta}v_{i+1}(z_E(s,t)) \\ v_i(w)+\frac{1}{N'}v_i(m)\ge b_L(z_L(w,m)) \end{array} \right\}$$

with the following metric:

$$d_\infty\left((u_i,v_i)_{i=1}^\infty, (\hat{u}_i,\hat{v}_i)_{i=1}^\infty\right) = \sum_{i=1}^\infty e^{-\beta i}\left(d_K(u_i,\hat{u}_i) + d_K(v_i,\hat{v}_i)\right).$$

Let W and W_T map the feasible wage functions to the real numbers:

$$\begin{aligned} &W : \mathcal{F}\to\mathbf{R} : \\ &(u_i,v_i)_{i=1}^\infty \mapsto \int_K e^{-\beta} v_1(k)d\kappa_1(k) + \sum_{i=1}^\infty e^{-\beta i}\int_A u_i(s)d\alpha(s); \\ &W_T : \mathcal{F}\to\mathbf{R} : \\ &(u_i,v_i)_{i=1}^\infty \mapsto \int_K e^{-\beta} v_1(k)d\kappa_1(k) + \sum_{i=1}^T e^{-\beta i}\int_A u_i(s)d\alpha(s). \end{aligned}$$

In the next proposition, we show that $W_T \xrightarrow{\Gamma} W$.

Proposition B.2.2. *Let $b_{E/L}$ be as defined in subsection 2.5. Fix $N_{E/L}\ge 1$ and $\kappa_1\in P^{ac}(K)$ and $\alpha\in P^{ac}(A)$. Then W_T Γ-converges to W.*

Proof. Let $(u_i,v_i)_{i=1}^\infty\in\mathcal{F}$. Then for any $(u_i^T,v_i^T)_{i=1}^\infty$ such that $\lim_{T\to\infty}(u_i^T,v_i^T)_{i=1}^\infty = (u_i,v_i)_{i=1}^\infty$ with respect to d_∞ we have $u_i^T\to u_i$ and $v_i^T\to v_i$ with respect to d_k. Therefore, from

$$\left|W\left((u_i,v_i)_{i=1}^\infty\right) - W_T\left((u_i^T,v_i^T)_{i=1}^\infty\right)\right| \le e^{-\beta}\left|\kappa_1(v_1 - v_1^T)\right| + \sum_{i=1}^T e^{-\beta i}\left|\alpha(u_i-u_i^T)\right| + \sum_{i=T+1}^\infty e^{-\beta i}\alpha(u).$$

we conclude that $\lim_{T\to\infty} W_T\left((u_i^T,v_i^T)_{i=1}^\infty\right) = W\left((u_i,v_i)_{i=1}^\infty\right)$ with Lemma 2.2.1 (b). Any sequence for $(u_i,v_i)_{i=1}^\infty$ is its own recovery sequence. Therefore, both conditions from Definition 2.3.1 (Condition (1) (lim inf inequality) and Condition (2) (lim sup inequality)) are satisfy and $W_T\xrightarrow{\Gamma} W$. □

Corollary B.2.3. *Fix* $c > 0$, $\theta, \theta' \in [0,1]$, $N, N' > 1$, *and* $\kappa_1 \in L^\infty(\bar{K})$. *Fix a distribution for students' abilities* α *with* $\log \alpha \in L^\infty(\bar{A})$.

Let (u_i^T, v_i^T) *be an optimizer for* (B.2). *Any convergent subsequence* $\{(u_i^{T_l}, v_i^{T_l})\}_{l=1}^\infty$ *converge to an optimizer of* (3.10).

Proof. Fix i. For all T, u_i^T and v_i^T are bounded non-decreasing and convex so by Lemma 2.2.1 (c), there exists a converging subsequence in d_K $\{(u_i^{T_l}, v_i^{T_l})\}_{l=1}^\infty$. Using a diagonal argument, we can get a subsequence that is converging for all i.

By Theorem 2.3.3, $\lim_{l\to\infty}(u_i^{T_l}, v_i^{T_l}) = (u_i, v_i)_{i=1}^\infty$ is an optimizer of (3.10). □

Bibliography

[1] Najma Ahmad, Hwa Kil Kim, and Robert J. McCann. Optimal transportation, topology and uniqueness. *Bull. Math. Sci.*, 1:12–32, 2011.

[2] Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and Optimisation of Flows on Networks*. Springer, 2009.

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows*. Birkhäuser Basel, 2 edition, 2008.

[4] Paul Appell. *Mémoire sur les déblais et les remblais de systèmes continus ou discontinus*. Mémoire présentés pas divers savants à l'Académie royale des sciences de l'Institut de France. Sciences mathématiques et physiques, 1887.

[5] Lionel Artige. Endogenous growth and parental funding in education of an olg model with a fixed factor. *Centre de Recherche en Economie Publique et de la Population*, August 2010.

[6] Robert J. Barro and Xavier Sala i Martin. *Economic growth*. Cambridge, Mass. : MIT Press, 2nd edition, 2004.

[7] Robert J. Barro and Jong-Wha Lee. International comparisons of education attainment. *Journal of Monetary Economics*, 32:263–394, 1993.

[8] Gary S. Becker and Kevin M. Murphy. The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economomics*, 107(4):1137–1160, November 1992.

[9] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 2000.

[10] Jonathan M. Borwein and Qiji J. Zhu. *Techniques of Variational Analysis*. Springer, 2005.

[11] Andrea Braides. *Γ-convergence or beginners*. Oxford University Press, 2002.

[12] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris Sér. I Math.*, 1987.

[13] Alice Erlinger, Robert J. McCann, Xianwen Shi, Aloysius Siow, and Ronald Wolthoff. Academic wages and pyramid schemes: a mathematical model. *Journal of Functional Analysis*, 269:2709–2746, 2015.

[14] David Gale and Lloyd S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, January 1962.

[15] Alfred Galichon. *Optimal Transport Methods in Economics.* Princeton University Press, 2016.

[16] Wilfrid Gangbo. Xerox copy from the first pages of "rapport d'habilitation à diriger des recherches en mathematiques; quelques problemes d'analyse nonconvex". Soutenu le 19 Janvier 1995.

[17] Luis Garicano and Esteban Rossi-Hansberg. Organization and inequality in knowledge economy. *The Quarlerly Journal of Economics*, pages 1383–1435, November 2006.

[18] John Gill. The use of the sequence $f_n(z) = f_n \circ \cdots \circ f_1(z)$ in computing fixed points of continuous fractions products and series. *Applied Numerical Mathematics*, 8:469–476, 1991.

[19] Gerhard Glomm. Parental choice of human capital investment. *Journal of Development Economics*, 53:99–114, 1977.

[20] Gerhard Glomm and B. Ravikumar. Public versus private investment in human capital: Endogenous growth and income inequality. *Journal of Political Economy*, 100(4):828–834, August 1992.

[21] Li Hao and Wing Suen. Technological changes and labor market segregation. http://faculty.arts.ubc.ca/lhao/research/Kremer.pdf, 2001.

[22] Leonid Kantorovich. On the translocation of masses. *Management Science*, 5(1):1–4, Oct. 1958.

[23] Gégoire Loeper. On the regularity of the polar factorization for time dependent maps. *Calc. Var. Partial Differential Equations*, 2004.

[24] Robert E. Lucas, Jr. On the size distribution of business firms. *The Bell Journal of Economics*, 9(2):508–523, Autumn 1978.

[25] Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke mathematical journal*, 80(2):309–323, November 1995.

[26] Robert J. McCann. Academic wages, singularities, phase transitions and pyramid schemes. *Proceedings of the International Congress of Mathematicians (Seoul 2014)*, Vol. III Invited Lectures:835–849, 2014.

[27] Robert J. McCann, Xianwen Shi, Aloysius Siow, and Ronald Wolthoff. Becker meets ricardo: Multisector matching with communication and cognitive skills. *Journal of Law, Economics and Organization*, 31(4):690–720, 2013.

[28] Robert J. McCann and Maxim Trokhimtchouk. Optimal partition of a large labor force into working pairs. *Econom. Theory*, 2010.

[29] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704, 1781.

[30] Yuri V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of probability and its applications*, I(2):157–214, 1956.

[31] Sherwin Rosen. The economics of superstars. *The American Economic Review*, 71(5):845–848, December 1981.

[32] Sherwin Rosen. Authority, control, and the distribution of earnings. *The Bell Journal of Economics*, 13(2):311–323, Autumn 1982.

[33] Alvin E. Roth. The origins, history, and design of the resident match. *(Reprinted) JAMA*, 289(7):909–912, Feburary 2003.

[34] Paul A. Samuelson. An exact consumption-loan model of interest with or without the social contrivance of money. *The Journal of Political Economy*, 66(6):467–482, December 1958.

[35] Lloyd S. Shapley and Martin Shubik. The assignment game i: The core. *International Journal of Game Theory*, 1(2):111–130, December 1971.

[36] Nancy L. Stokey and Robert E. Lucas. *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge, Massachusetts, and London, England, 1989.

[37] Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific J. Math.*, 5:285–309, 1955.

[38] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.

[39] Cédric Villani. *Optimal Transport: Old and New*. Springer, 2009.

[40] Alessio Figalli Young-Heon, Kim, and Robert J. McCann. Hölder continuity and injectivity of optimal maps. *Arch. Rational Mech. Anal.*, 209:747–795, 2013.

[41] Ayed Zambaa and Ben Hassen. Education financing and growth using overlapping generations model: a theoretical perspective. *Economic Annals*, 58(197), April - June 2013.