Distance to the Technology Frontier and the Allocation of Talent*

Tommaso Porzio†

Yale Department of Economics

JOB MARKET PAPER

Click here for most current version

January 24, 2016

Abstract

This paper develops a matching model - in which heterogeneous individuals form production teams and choose an appropriate technology - to study how the technological environment changes the pattern of matching and consequently the production structure of an economy. The main theoretical result is that, if technology improvements are expensive, teams bring high and low skilled individuals together and all teams choose similar technologies. In contrast, if improvements are inexpensive, talent concentrates in teams that choose the most advanced technologies, leading to larger dispersion of economic activity. Then, I apply the theory to study cross-country differences in the allocation of talent. Since relatively poor countries can make cheap and large improvements through technology adoption, the theory predicts that they should have stronger concentration of talent than relatively rich ones. I derive an empirical measure of the concentration of talent from the model to validate this prediction using micro data from several countries; in the cross-section, using a sample of 63 countries, and in the time-series, comparing the growth experiences of South Korea and United States. Finally, a dynamic extension of the theoretical framework demonstrates that endogenous allocation of talent coupled with localized technological progress may prevent cross-country convergence even in the absence of exogenous barriers.

*I am extremely grateful to my advisors, Mikhail Golosov, Giuseppe Moscarini, Nancy Qian, Aleh Tsyvinski, and Chris Udry for their invaluable guidance and support and to Benjamin Moll, Michael Peters and Larry Samuelson for helpful suggestions and stimulating discussions. I also thank for very helpful comments Joseph Altonji, Costas Arkolakis, David Atkin, Lorenzo Caliendo, Douglas Gollin, Pinelopi Goldberg, Samuel Kortum, David Lagakos, Ilse Lindenlaub, Costas Meghir, Nicholas Ryan, Alessandra Voena, Nicolas Werquin, and Noriko Amano, Gabriele Foà, Sebastian Heise, Ana Reynoso, Gabriella Santangelo, Jeff Weaver, Yu Jung Whang, and seminar participants at Yale.

†28 Hillhouse Avenue, New Haven, CT 06511, email: tommaso.porzio@yale.edu.

1 Introduction

More than a third of the world's population lives in countries whose GDP per capita is smaller than that of the United States in 1900. These individuals, however, have different opportunities available than Americans did at the end of the 19th century. In particular, they have access to the technologies invented during the past century. In light of such a different technological environment, it is natural to expect that developing countries organize production differently than developed countries have done in the past, and possibly even do today. Nonetheless, it is often implicitly assumed in the literature that developing countries today should follow in the footsteps of currently developed ones. This paper proposes a different view.

I develop a theoretical model of the assignment of talent to production teams in an environment of localized technological progress: that is, an environment where teams choose the most appropriate technology given the skills of their members.[1] The *main result* of the theory is that, allowing for an endogenous technology choice in an otherwise standard framework, changes the equilibrium pattern of matching and increases the similarity in skills of individuals working together. More specifically, I make standard assumptions on the production function to guarantee that, if technology is fixed, the equilibrium is symmetric: all teams are analogous and formed by high skilled individuals as managers and low skilled ones as their worker. If instead technology is endogenous and improvements are sufficiently cheap, the allocation of talent changes to an asymmetric equilibrium: some teams attract all high skilled individuals, while others are left only with low skilled ones. I apply this model to study cross-country differences in the allocation of talent. The theory predicts that in developing countries, since individuals have the possibility to adopt modern technology, the asymmetric equilibrium should prevail. I validate this prediction using micro data from several countries: the allocations of talent in developing countries today are different from the ones of developed countries both today and in the past.

I now provide a more detailed description of the model's environment. Each economy is inhabited by a continuum of individuals heterogeneous in their ability.[2] Production of output requires to combine labor input with a costly production technology. Labor input is produced in teams of two individuals. Technology is complementary with labor input and is chosen by each team subject to a convex cost of adoption. The production function of labor input has two key properties. First, there is a motive for *task-specialization*, since

[1] Along the lines of the *"New view of technological change"* of Atkinson and Stiglitz (1969).

[2] Talent is one dimensional: skilled individuals have an absolute advantage. This differs from Hsieh et al. (2013) and Lagakos and Waugh (2013) that study the allocation of talent with comparative advantages.

the two individuals decide who is the manager and who is the worker of the team and the marginal product of the manager's talent is higher than that of the worker's. Second, there is *complementarity*: the marginal product of each individual's talent is increasing in the talent of his partner. I solve for the competitive equilibrium; individuals pick an occupation and a production partner, taking into account the optimal technology choice of the manager-worker team that they form.

The equilibrium assignment is determined by the properties of the production function. Yet, unlike most matching models, complementarity is not sufficient to pin down the unique equilibrium. Complementarity in production dictates positive assortative matching between managers and workers, but in this setting the sets of managers and workers are themselves endogenous objects. As a result, the assignment depends on a complex trade-off between task-specialization and complementarity. Task-specialization alone would lead all the most skilled individuals to become managers and hire the low skilled ones as their workers, but complementarity pushes identical individuals to pair together. To overcome this difficulty, most previous work that combines occupational choice and matching uses functional form assumptions that guarantee that task-specialization is sufficiently strong with respect to complementarity, and thus a Spence-Mirrlees type condition holds and separates managers and workers into two connected sets.[3] Complementarity is then sufficient again to pin down the unique equilibrium, in which the most skilled individuals are managers and hire assortatively the less skilled ones. I make that same assumption for the production of *labor input* so that, if teams use *identical* technology, my setting replicates that same pattern of matching. Teams, however, in general do not choose identical technology.

Allowing for an endogenous choice of technology introduces an additional dimension to be taken into account to solve for the equilibrium. When an individual evaluates the production complementarity of his talent with that of potential partners, he must take into account both the complementarity in the production of labor input and the change in the optimal chosen technology, since technology and labor are complementary. As a result, the production complementarity - hence the allocation of talent - depends on the technological environment, as captured by the properties of the technology cost function. The main theoretical contribution of the paper is to formally study this relationship.

Endogenous choice of technology introduces a *new trade-off*. In equilibrium, the team's manager is more skilled than the worker, due to task-specialization. Additionally, labor-

[3]See for example Lucas (1978) and Garicano and Rossi-Hansberg (2006). An exception to this norm is the work of Kremer and Maskin (1996).

technology complementarity implies that more skilled teams choose a more advanced technology. Therefore, an individual faces a trade-off between being a manager in a team that uses a lower technology and being a worker in one that uses a higher technology. I show that this trade-off is modulated by the *convexity* of the cost of technology. In fact, the lower the convexity, the smaller is the cost difference between any two technologies, leading teams to have a larger gap in their optimal technology choices. As a result, the smaller the convexity, the more an individual can increase his used technology by becoming a worker and thus being matched with a more skilled partner. This argument implies that complementarity in production decreases in the convexity of the cost of technology.

More specifically, I show that when the convexity is sufficiently high, the same pattern of matching emphasized in the literature emerges: the most skilled individuals become managers and hire the low skilled ones, since task-specialization within teams *segments talent by occupation*. In contrast, when the convexity is sufficiently low, the opposite pattern appears: stronger complementarity *segregates talent by technology*. That is, the most skilled individuals concentrate in teams that use the most advanced technology, and some high skilled are therefore assigned to be workers. Intuitively, when convexity is high all teams use a similar technology and thus high skilled individuals wish to specialize in the most skill-intensive occupation, becoming managers. In contrast, when convexity is small, some teams use a much more advanced technology than others, and thus high-skilled individuals wish to work in these teams, even at the cost of being workers.

I then make simplifying functional form assumptions that allow to characterize analytically the unique equilibrium even between the two polar cases (talent segmentation and segregation), thus making the model amenable to an empirical analysis. I derive from this version of the model a measure of *concentration of talent,* which captures the expected similarity in skills of the members of a team.[4] I show that the less convex the cost of technology the stronger the concentration of talent.

Next, I apply the theory to study cross-country differences in organization of production. I provide an interpretation of the convexity of the cost of technology in terms of the distance of a country to the technology frontier, and show that the resulting empirical predictions are borne out in micro data from several countries.

It has been shown in the literature that cross-country variation in GDP per capita is, to a large extent, explained by differences in average technology.[5] Individuals in countries that are relatively poor with respect to the frontier have thus a unique opportunity: in

[4]The maximum concentration is when talent is segregated and the minimum when talent is segmented.
[5]See, among others, Klenow and Rodriguez-Clare (1997) and Comin and Ferrer (2013).

principle, they can make large improvements in their technology at a low cost by adopting the existing technology of rich countries. Motivated by this fact, I interpret relative GDP per capita across countries as a proxy for how expensive large technological improvements are, and hence for the convexity of the cost of technology. Under this interpretation, the theory gives sharp predictions for cross-country differences in the organizational structure of the economy. In particular, for countries farther from the frontier, the theory predictions are consistent with many familiar features of developing countries - such as larger dispersion of technology across teams (or more generally firms and sectors)[6] - and they highlight an aspect of development previously overlooked: a *stronger concentration of talent*.

In the main empirical exercise, I therefore use large sample labor force surveys for 63 countries, available from Integrated Public Use Microdata Series (IPUMS), to show that in the countries farther from the technology frontier - i.e. with lower relative GDP per capita - the concentration of talent is indeed higher. In the data, we do not directly observe either teams or individual skill. Nonetheless, I show that if education is positively correlated with skill and teams sort into sectors depending on the ability of their members,[7] then we can use available data on individual industry and education to compute the measure of concentration of talent as defined in the model. I construct this measure for each country-year pair and show how it varies as a function of distance to the frontier, both in the cross-section and in the time-series. I first compare countries in the cross-section. Concentration of talent is significantly negatively correlated with GDP per capita and the magnitude of cross-country differences is sizable. I then compare middle income countries today (such as Mexico and Brazil) with the United States in 1940, that was at the same level of development - and critically, it was closer to the technology frontier. Middle income countries today have significantly higher concentration of talent than the U.S. did in the past. This result alleviates the concern that cross-sectional differences are merely capturing differences in levels of development. Last, I compare the growth paths of South Korea and the United States in the past seven decades. In the United States, the concentration of talent has remained largely unchanged, consistent with the fact that they have been growing steadily as world leaders (i.e. on the frontier). In contrast, South Korea has seen a rapid decline in the concentration of talent as it has approached the technology frontier.

In the final part of the paper, I develop a dynamic extension of the model to study the

[6]See for example Caselli (2005), Hsieh and Klenow (2009), and Comin and Ferrer (2013).

[7]Sorting of teams to sectors can be justified through agglomeration economies at the industry level.

implications of the allocation of talent for economic growth, and show that countries that start to grow late may converge to a different growth path. I introduce four additional elements to the model. First, individuals are infinitely lived. Second, there is a technology frontier that grows exogenously. Third, there are two sectors, and technology is sector-specific. Fourth, individuals allocated to a sector decide only on marginal technological improvements towards the frontier. The state variables in the model are thus the distances to the frontier in each sector. I study the growth paths of countries that have identical fundamentals, but exogenously start to grow (or takeoff) at different points in time. The same mechanism of the static model is present. Therefore, when a country is far from the frontier, talent concentrates. There is an additional dynamic effect: the technology gap across sectors propagates over time depending on the allocation of talent. For example, if skilled individuals concentrate in one sector, they adopt more advanced technology, and in the next period there is a larger technology gap across sectors. I compute the model and show that the interaction of the described forces implies that, as long as individuals are sufficiently impatient, a country that starts to grow late - thus far from the technology frontier - may converge to a different balanced growth path, which has lower output, larger technology heterogeneity across sectors, and more concentration of talent.

Related Literature. The seminal work of Lucas (1978) highlighted the role of teams as means for high skilled individuals to specialize in managerial tasks. More recently, the literature on hierarchies of production, which builds upon Garicano (2000) and Garicano and Rossi-Hansberg (2006), emphasizes this same aspect. In these papers, the allocation of talent is fixed ex-ante by assuming a production function with sufficiently weak complementarities for a Spence-Mirrlees separability condition to hold: in any equilibrium the most skilled individuals are bound to be managers. Kremer (1993) instead proposes a production function in which each task is symmetric, and there is complementarity between team's members. As a result, teams put together individuals with identical skills. In my framework, depending on the properties of the cost of technology, the equilibrium might resemble either one of these two extremes or lie in an intermediate area. In the studying the trade-off between complementarity and task-specialization, my work is similar to a working paper by Kremer and Maskin (1996), that extends Kremer (1993) to allow for asymmetry across tasks and studies the effect of the increased segregation of skills in the U.S. labor market on wage inequality. My work diverges in studying how the production complementarity depends on the properties of the technological environment.

This paper shares with Kremer (1993) the goal of understanding cross-country differences in organizational structure. However, Kremer (1993) focuses on *average* differences across countries - e.g. in poor countries, firms are on average smaller - while I focus on cross-country differences in the within-country *distribution* of economic activity - e.g. in poor countries large and small firms coexist, while in rich ones all firms are similar in size.

The idea that distance to the frontier may impact organization of production is present in Acemoglu et al. (2006). They study selection into entrepreneurship with credit constraints. I thus see my work as complementary to theirs. Roys and Seshadri (2013) also studies differences across countries in the way in which production is organized. It uses a quantitative version of Garicano and Rossi-Hansberg (2006) in which human capital is endogenously accumulated, as in Ben-Porath (1967). Cross-country differences are therefore generated by changes in the distribution of talent, and not by changes in the pattern of matching, which is fixed ex-ante.

The application of the theory to developing countries fits into the debate on the causes of the existence of dual economies. Through the lens of my model, a dual economy emerges endogenously as a consequence of individuals in poor countries having the ability to adopt advanced technology from the frontier, and the resulting concentration of talent. This view is original, but resembles most closely the one in La Porta and Shleifer (2014), which emphasizes how duality is tightly linked to economic development.

The work Acemoglu (1999) and Caselli (1999) establishes, through mechanisms different from the one that I propose, a connection between the technological environment and the allocation of workers to jobs. Acemoglu (1999) focuses on the interaction between labor market frictions and the fact that firms have to commit ex-ante whether to create jobs for high or low skilled workers and shows conditions for a separating equilibrium to exist. Caselli (1999) shows that when new technologies are adopted, the most skilled are the most likely to start using them, thus separating themselves from the rest of the economy. Both these papers do not consider complementarity between individuals working together. My work instead focuses exactly on this latter channel and characterizes how the properties of the technological environment change the overall production complementarity, thus changing the assignment of workers to jobs.

A key feature of the model is that different teams choose to use a different technology, depending on the ability of their members. This idea has been explored by Basu and Weil (1998), who argue that each country might have a different appropriate technology, and by Acemoglu and Zilibotti (2001), that shows that part of the cross-country productivity gap might be explained by a mismatch between the low skill intensity of poor countries and the skill biased technologies invented in rich countries. I explore the same idea of appropriate technology, but focus on the implications for the pattern of matching.

The dynamic application of the theory, shares the emphasis on the role of talent allocation for growth of Murphy and Vishny (1991), that argues that the allocation of talent to rent seeking might lead to stagnation. In this paper, stagnation is instead determined by skilled individuals not being willing to reallocate where their talent is most needed to spur growth. Last, Parente and Prescott (1994) assumes the existence of barriers to technology adoption across countries and calibrate the barrier sizes that fit the data. In my work, barriers to technology adoption are instead endogenously determined. They are the consequence of a large technology gap across sectors, which depends on past allocation of talent and prevents skilled individuals from reallocating, thus leaving one part of the economy unable to adopt advanced technologies.

The structure of the paper is as follows. The theoretical results are provided in Section 2. In Section 3, I offer an interpretation of the theoretical results in terms of distance to the technology frontier and provide evidence in support of the resulting empirical predictions. In Section 4, I provide a dynamic extension to study the theory's implications for economic growth. Section 5 concludes. The proofs of all the results are gathered in the Appendix.

2 Theory: Choice of Technology and the Allocation of Talent

I develop a matching model in which heterogeneous individuals form production teams and choose an appropriate and costly technology. I use it to characterize how the allocation of talent depends on the properties of the cost of technology.[8]

2.1 Environment

I consider an economy inhabited by a mass one of heterogeneous individuals, indexed by a type $x \in \mathbb{X}$. The type set $\mathbb{X}$ is thought of as a set of types that indicate the *relative*[9] skill level, to the extent that if $x' > x$, then an individual of type x' is more able than x. In the main text, $\mathbb{X}$ is considered to be the interval $[0, 1]$. All results extend to the case of finite $\mathbb{X}$. This generality comes at the cost of additional notation; for this reason is left to appendix A.4. Individuals have an increasing and non-satiated utility function

[8] An example that highlights the main mechanism in a simplified economy is provided in Section A.3.

[9] Relative to the country specific skill distribution. x ranks individuals from the most to the least able.

of consumption and no disutility of labor.

Production Technology. Two agents must team up to produce output. Production requires them to perform two tasks, which are differentially sensitive to skill, and to choose a costly technology that is complementary with skills. The individual that performs the most skill intensive task is called the manager of the team, while the other member of the team is the worker. A manager of type x' and a worker of type x with technology a produce net output equal to

$$g(x', x, a) = af(x', x) - c(x', x, a)$$

where $f(x', x)$ is the production from the labor input of the team, and $c(x', x, a)$ is the cost of technology a. I interpret this cost as the amount necessary to purchase the vintage of capital associated with a given technology. Higher technologies are costlier, but yield higher returns by combining more capital to the same amount of labor input. This interpretation is modeled after the concept of *appropriate technology*, as in Basu and Weil (1998), with the difference that technology choice is not country specific - as in their setting - but each team within a country may have a different appropriate technology.

Few assumptions on the functions f and c guide the analysis and impose that, if technology is fixed, my setting replicates the same allocation of talent as seen in the literature. These assumptions thus allow to isolate the role of endogenous technology.

Assumption 1 (Production Function $\mathbf{f(x', x)}$). *(1.1) f is twice continuously differentiable. (1.2) f is increasing in both arguments: $f_1 > 0$ and $f_2 > 0$. (1.3) management is more skill-intensive: $f_1(x, x) > f_2(x, x)$. (1.4) skills of managers and workers are complementary: $f_{12} > 0$. (1.5) there is single-crossing:* $\min_{x' \in \mathbb{X}} f_1(x, x') > \max_{x' \in \mathbb{X}} f_2(x', x)$.

A few comments are in order. *(1.2)* captures the fact that more talented individuals are better at producing output. It is key for the results that also the skills of workers are useful for production. *(1.3)* captures the fact that, for given technology and partner, the individual's skill is more useful (has larger effect on output) if employed in a managerial position. This assumption is common in the literature. See, for example, the seminal paper by Lucas (1978), which assumes that only manager ability matters for production, and Garicano and Rossi-Hansberg (2006), which builds from primitives a production function that features this property.[10] *(1.4)* captures complementarity in production between tasks:

[10] Both these papers allow one manager to be matched with more than one worker. I assume one to one matching because the focus on the paper is not on firm size, but rather on skill composition in a firm. I

the skills of a manager are more useful if paired with a high skilled worker. This is also a natural assumption, pervasive in the literature. Lastly, *(1.5)* guarantees that the marginal effect of increasing the ability of a manager on the labor output of a team, f, is always larger than the effect of increasing the ability of a worker. This assumption separates, if technology is fixed, managers and workers and guarantees that the highest skilled should be managers, as in the previous literature.[11]

Before introducing the next assumption, it is useful to define $\alpha^*(x', x) \equiv c_3^{-1}(x', x, f(x', x))$, that is the solution of $\max_a af(x', x) - c(x', x, a)$.

Assumption 2 (Cost of Technology $\mathbf{c(x', x, a)}$). *(2.1) c is twice continuously differentiable, if not otherwise specified. (2.2) c is increasing and convex in a: $c_3 > 0$ and $c_{33} > 0$. (2.3) skills of managers are more valuable from a cost perspective $c_1 \leq c_2$ (2.4) c is additively separable in skills of managers and workers: $c_{12} = 0$; (2.5) technology and manager's skills are complementary $f_1(x', x) - c_{13}(x', x, a) > 0$; (2.6) technology and worker's skills are complementary: $f_2(x', x) - c_{23}(x', x, a) > 0$ $\forall a \in \Big[0, \alpha^*(x', x)\Big)$ and $f_1(x', x) - c_{13}(x', x, a) \geq 0$ if $a = \alpha^*(x', x)$; (2.7) the complementarity in technology-manager skill is stronger than the complementarity in technology-worker skill: $f_1(x', x) - f_2(x', x) > c_{13}(x', x, a) - c_{23}(x', x, a)$ $\forall a \in [0, \alpha^*(x', x)]$.*

Again, a few comments are in order. *(2.3)* implies that increasing the ability of the manager gives a larger decrease on the cost of technology (or a smaller increase). *(2.4)* is useful to isolate the effect of complementarity in production as coming either from $f_{12} > 0$ or from the choice of technology, and not directly through cost reduction. *(2.5)* and *(2.6)* are common, but non trivial assumptions, to the extent that they define the role of technology as an input in production that is complementary with skills.[12] This is consistent with most studies (see, for example, Goldin and Katz (1998), and in a developing country setting, Foster and Rosenzweig (1995)).[13] *(2.7)* more specifically implies that the ability of

build the simplest framework that isolates the effects of interest. Most results can be extended to a case in which one manager is matched with more than one worker, as long as there are no complementarities between workers themselves. This extension comes at the cost of stronger assumptions on complementarity. For example, supermodularity ($f_{12} > 0$) must be substituted with strict log supermodularity as in the setting of Eeckhout and Kircher (2012) and Grossman et al. (2013). Results are available upon request.

[11]Given the key role that this assumption plays, in A.2, I discuss it in details and show that (i) it corresponds to the Spence-Mirrlees condition in this environment, and (ii) it holds and plays a major role, in most settings in which heterogeneous workers must make an occupational choice; (iii) it can be rewritten as a condition on the difference between task-specialization and complementarity in f.

[12]Notice that *(2.5)* implies that $g_{13} > 0$ and *(2.6)* that $g_{23} > 0$. Also, *(2.6)* can be rewritten as $g_{13} > g_{23}$.

[13]This assumption may fail in some contexts, such as if computerization simplifies production tasks, requiring less able workers. However, it is suited to the empirical setting studied in this paper. In Section

the manager is more important in generating high returns from technology than the ability of the worker. This is consistent, and *(2.3)* as well, with recent studies emphasizing the role of managers in technology adoption and good practices (see Bloom and van Reenen (2007) and Gennaioli et al. (2013)).

An example that satisfies assumptions 1 and 2 is $g(x', x, a) = M(x') L(x) - \tilde{c}(a) L(x)$ with $M'(x) L(0) > M(1) L'(x)$; $M' > 0$; $L' > 0$; $\tilde{c}' > 0$; and $\tilde{c}'' > 0$. Simple functional forms for $M(\cdot)$, $L(\cdot)$, and $c(\cdot)$ can be: $M(x) = x$; $L(x) = 1 + \gamma x$, $\gamma < 1$; and $\tilde{c}(a) = \frac{1}{\eta+1} a^{\eta+1}$.

Assignment of Talent to Teams. Production requires one manager and one worker. Individuals are not restricted ex-ante to belong to either group. We can interpret this setting as a standard one to one matching in which both sides of the market are themselves endogenous sets. Individuals in the set $\mathbb{X}$ decides which side of the matching to join, whether to belong to the set of workers, $\mathbb{W}$, or to the set of managers, $\mathbb{M}$. They then match with an individual from the other side. An allocation in this setting must in fact specify both the occupational choice of each individual and the production partner that he would get in either occupation.

Definition 1 (Allocation). *An allocation φ is a quintuple of functions $\{\omega, \mu, m, w, \alpha\}$, where (i) $\omega : \mathbb{X} \to [0, 1]$ assigns to each type the fraction of individuals that are workers; (ii) $\mu : \mathbb{X} \to [0, 1]$ assigns to each type the fraction of individuals that are managers; (iii) $m : \mathbb{X} \to \mathbb{X}$ assigns to each type the manager he would get if he is a worker; (iv) $w : \mathbb{X} \to \mathbb{X}$ assigns to each type the worker he would get if he is a manager; (v) $\alpha : \mathbb{X} \times \mathbb{X} \to \mathbb{A}$ assigns to each possible pair of types a production technology in $\mathbb{A}$.*

I call ω and μ the occupational choice functions, and m and w the matching functions.[14] The set $\mathbb{A}$ is the set of available technologies, which, if not otherwise specified, is equal to the set of positive real numbers.

Definition 2 (Feasible Allocation). *An allocation φ is feasible if (i) $\mu(x) + \omega(x) = 1$ $\forall x \in \mathbb{X}$; (ii) $x = w(m(x))$ if $\omega(x) > 0$ and $x = m(w(x))$ if $\mu(x) > 0$ $\forall x \in \mathbb{X}$; (iii)*

3.3, I introduce evidence from firm level data showing that the direct implication of this assumption, i.e. that high skilled individuals use more advanced technology, holds to a similar degree across countries of different income levels. Additionally, notice that the theoretical results hinge on the idea that different individuals have different appropriate technologies, and not necessarily on complementarity between skills and technology. The theory could be rewritten with strict substitutability between skills and technology, and still provide similar results.

[14] I define m and w as matching *functions*, rather than correspondences. This is without loss of generality due to the fact that $\mathbb{X} = [0, 1]$. I show this result in A.4 together with the case for $\mathbb{X}$ finite.

$\omega(x) = \mu(m(x))$ *if* $\omega(x) > 0$ *and* $\mu(x) = \omega(w(x))$ *if* $\mu(x) > 0$ $\forall x \in \mathbb{X}$. *I call* $\mathbb{F}$ *the set of all feasible allocations.*

Each feasibility constraint must hold for each type x. The first one guarantees that an individual cannot be simultaneously a manager and a worker, the second one that matches are reciprocal, and the third one that labor market clears. [15]

2.2 Planner Problem, Existence, and Decentralization

I consider the problem of a planner who, in order to maximize output, assigns individuals to occupations, forms production teams, and picks the technology used by each team. The planner therefore picks the feasible allocation that maximizes total output:

$$\max_{\varphi \in \mathbb{F}} \int_{\mathbb{X}} g\left(m\left(x\right), x, \alpha\left(m\left(x\right), x\right)\right) \omega\left(x\right) dx. \tag{1}$$

I focus on the planner problem because, as known[16] in the literature, if a competitive equilibrium exists, it decentralizes the optimal allocation. Moreover, the focus of this paper is to study the assignment of talent to teams, rather than the prices that support it. Nonetheless, I outline in Section A.5 the competitive equilibrium, and associated prices, that decentralize the planner's allocation. In the same appendix, I prove the equivalence of my setting to an associated bipartite matching problem.[17] I exploit this isomorphism to prove the next proposition.

Proposition 1 (Existence and Decentralization). *There exists an optimal allocation* φ^* *and* φ^* *can be decentralized in a competitive equilibrium.*

Uniqueness is not guaranteed. This does not represent a limitation, since I characterize the properties that *any* optimal allocation (and equilibrium) must satisfy.[18]

2.3 The General Case: Segmentation or Segregation of Talent

I next characterize how talent is assigned to teams in an optimal allocation.

[15] Notice that I have defined the matching function m and w on the whole support of x. The reason is that the allocation must define the partner also for eventual deviations. For example, if for a type x there are no workers (i.e. $\omega(x) = 0$), still we need to define what managers they would be matched with in case of a deviation.

[16] See for example Becker (1973) and Chiappori et al. (2014).

[17] This isomorphism follows from an insight in Chiappori et al. (2014).

[18] The comparative static is robust (Athey et al. (1998)): it does not depend on the equilibrium selection.

As a preliminary step, it is convenient to define the value of a team that takes into account the optimal technology choice. This is given by

$$v(x', x) \;=\; g(x', x, \alpha^*(x', x)), \tag{2}$$

and it is a function of only (x', x), since the optimal technology choice depends only on the team composition, and not on the optimal allocation φ^*. The assignment of individuals into teams depends on the properties of v, which are inherited from the properties of the production function of labor input, f, and the cost of technology, c.

The characterization of the optimal assignment proceeds in three steps. I first describe standard results that hold directly due to the properties of the value function v and that are common in settings with occupational choice and matching, such as Kremer and Maskin (1996) and Garicano and Rossi-Hansberg (2006). I then describe the role of endogenous technology choice and show that it introduces a trade-off between occupation and technology. Last, I show the main result that describes how the the properties of c affects the optimal assignment.

Standard Results. The ability of the manager and worker are complementary in the production of team value v (in equation 2).[19] As a result, the optimal allocation dictates positive assortative matching between managers and workers: skilled workers are assigned to skilled managers.

Lemma 1 **(Managers-Workers Assortative Matching).** *The matching function m^* is weakly increasing: if $x' > x$ then $m^*(x') \geq m^*(x)$.*

Additionally, the marginal effect on team output of increasing the ability of the manager is greater than the one of increasing the ability of the worker.[20] As a result, as long as two individuals use the same technology - as is the case when they work together -, the most skilled is assigned to the task that is more skill demanding, i.e. to be a manager.

Lemma 2 **(Workers Match Upwards).** *Managers are weakly more skilled than workers: $m^*(x) \geq x$.*

This last lemma has an important implication: workers match *upwards*, that is, they work with a partner more skilled than themselves. This feature is at the core of the trade-off between occupation and technology that I describe next.

[19] This follows from assumptions *(1.4)*, *(2.4)*, *(2.5)* and *(2.6)* - as shown in the appendix.

[20] This follows from assumptions *(1.5)*, *(2.3)*, *(2.7)* - as shown in the appendix.

Occupation-Technology Tradeoff. Technology and individuals skills are complementary.[21] As a result, more skilled teams use a more advanced technology.

Lemma 3 (Optimal Technology). *The optimal technology of a team* (x', x) *is given by* $\alpha^*(x', x) = c_3^{-1}(x', x, f(x', x))$, *and satisfies* $\frac{\alpha_1^*}{\alpha^*} > \frac{\alpha_2^*}{\alpha^*} \geq 0$.

Lemma 3 and Lemma 2 together imply that any individual would use a more advanced technology if he is a worker rather than a manager.

Lemma 4 (Technology of Workers and Managers). *Any individual uses a higher technology as a worker than as a manager. I.e.* $\alpha^*(m^*(x), x) \geq \alpha^*(x, w^*(x))$.

Lemma 4 highlights the trade-off, between occupation and technology, that is generated by the endogenous choice of technology. An individual, by being a worker, hence taking a lower skill-intensive occupation, gets access to a more advanced technology. The key object in this trade-off is the gap between the technology that an individual x would use as a worker or as a manager. I define this η^*. It depends on the optimal matching functions:

$$\eta^*(x, m^*(x), w^*(x)) = \frac{\alpha^*(m^*(x), x)}{\alpha^*(x, w^*(x))}.$$

The convexity of the cost of technology, pinning down how different are the technologies appropriate to teams of different ability, determines the size of the technology gap.

Lemma 5 (Technology Gap). *For given* x, $m^*(x)$, *and* $w^*(x)$, *the technology gap* $\eta^*(x, m^*(x), w^*(x))$ *decreases in the convexity of* c *with respect to* a, $\frac{c_{33}}{c_3}$.

Main Result. The previous five lemmas completely characterize the optimal allocation φ^*, given the occupational choice functions ω^* and μ^*. However, the main difficulty in the problem is the characterization of these occupational choice functions, hence of the endogenous sets of workers and managers. Most of the matching literature focuses on two-sided matching, for example matching between husbands and wives as in Becker (1973). In those settings, positive assortative matching is sufficient to pin down the unique optimal allocation. In contrast, the two sides of the matching, the sets of managers and workers, are here themselves endogenous objects. As a result, positive assortative matching is not sufficient to pin down the optimal allocation.

I do not completely characterize ω^* and μ^*, but I describe two polar cases that might emerge in the optimal allocation, and show conditions for either one of them to be optimal.

[21] See assumptions *(2.5)*, *(2.6)*, *(2.7)*.

These two cases are polar with respect to the occupation-technology tradeoff, and they also correspond to the maximum and the minimum distance between the skill levels of individuals working together in a team.

Definition 5 (Segmentation by Occupation). *Talent is segmented by occupation if* $x' > x$ *and* $\mu^*(x) > 0$ *imply* $\mu^*(x') = 1$.

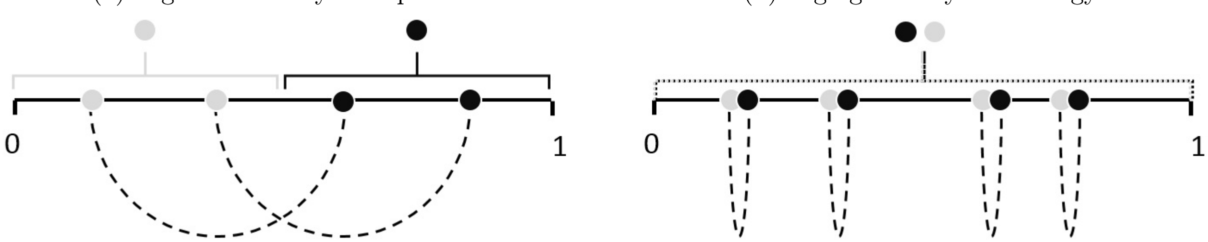
When talent is segmented by occupation, there will be a cutoff type such that all individuals with ability above this cutoff are managers, and all of those with ability below the cutoff are workers. This case is illustrated in Figure 1a below.

Definition 6 (Segregation by Technology). *Talent is segregated by technology if* $x' > x$ *implies* $\alpha^*(x, w^*(x)) \geq \alpha^*(m^*(x), x)$ *with probability* 1.

When talent is segregated by technology, more skilled individuals use a higher technology than lower skilled ones. Take any $x' > x$ such that x' is a manager and x is worker. For talent to be segregated by technology, it must be the case that $m^*(x) < x'$, since by Lemma 1, x would otherwise use a higher technology.[22] In fact, each worker's manager must be no more skilled than the next highest individual in the economy skill distribution. This second case is illustrated in Figures 1b.

Figure 1: Two Polar Cases of Allocation of Talent for $\mathbb{X} = [0, 1]$

(a) Segmentation by Occupation (b) Segregation by Technology



Notes: A grey dot represents a worker, while a black dot represents a manager. Dotted lines connect workers to their managers.

When talent is segmented, there is a larger distance between an individual's skill and his partner's. When talent is segregated, instead, individuals work with others of similar talent. Last, notice that even in the case in which talent is segmented by occupation, high skilled individuals use *on average* a higher technology. However, since workers match upwards, an individual x uses a better technology than a higher skilled one $x' > x$, if x

[22] Alternatively it could be that $\alpha^*(m^*(x), x) = \alpha^*(x', w^*(x'))$, but this equality may hold only if there are kinks in the cost of technology c, which we assumed away through assumption *(2.1)*.

is a worker, x' is a manager, and the manager of x, $m^*(x)$ is more skilled than x'. If talent is segmented, this happens for at least some worker. Talent, in fact, cannot be simultaneously segmented and segregated.

I next characterize which properties on the cost of technology are sufficient for the optimal allocation to satisfy either one of the two polar cases emphasized above. For this, it is useful to introduce a last piece of notation. Let the set $\mathbb{A}_{c,\mathbb{X}}$ be the set of all technologies that are optimal, given c, for at least one of the teams that can be formed given the type set $\mathbb{X}$. Therefore

$$\mathbb{A}_{c,\mathbb{X}} = \{\alpha^*(x', x) : (x', x) \in \mathbb{X} \times \mathbb{X}\}.$$

Proposition 2 (Segmentation or Segregation). *There exist $\kappa_1 \leq \kappa_2$ such that in any optimal allocation φ^*, if $\frac{c_{33}(x',x,a)}{c_3(x',x,a)} \geq \kappa_2$ for all $x', x \in \mathbb{X}$ and for all $a \in \mathbb{A}_{c,\mathbb{X}}$ then talent will be segmented by occupation. On the other hand, if $\frac{c_{33}(x',x,a)}{c_3(x',x,a)} \leq \kappa_1$ for all $x', x \in \mathbb{X}$ and for all $a \in \mathbb{A}_{c,\mathbb{X}}$, then talent will be segregated by technology.*

The convexity of the cost of technology modulates the role of occupation and technology in driving the optimal assignment - and more generally the complementarity of v - and thus changes the pattern of matching. If the convexity is high, the marginal cost increases steeply and thus every team has a similar optimal technology. The occupation drives the assignment and teams gather together high skilled managers and low skilled workers. If the convexity is low, the elasticity of optimal technology to ability is large. High skilled individuals are paired together to reap the benefits - through skill-technology complementarity - from using a high technology.

The formal proof of the proposition is left to the appendix. I nonetheless discuss here the main insights and difficulties. For the purpose of this discussion, I let the cost of technology not depend on the ability of either workers or managers, that is $c_1 = c_2 = 0$.

Recall that we need only to solve for the sets of managers and workers. Given those, Lemmas 1-5 characterize the optimal allocation. The planner, as usual, assigns individuals to the occupation where they have their comparative advantage. Hence, for example, most skilled individuals should be assigned to the occupation with a higher marginal value of their skills. The difficulty stems from the fact that the marginal value of an individual skills depend not only on his occupation, but also on the production partner and technology, due to complementarity. Hence the comparative advantage occupation of each individual depends on the choice of everyone else in the economy. More specifically, the marginal values of skills for an individual x in the two occupations, respectively manager and worker, are

$$
\begin{aligned}
v_1(x, w(x)) &= \alpha^*(x, w(x)) f_1(x, w(x)) \\
v_2(m(x), x) &= \alpha^*(m(x), x) f_2(m(x), x),
\end{aligned} \tag{3}
$$

where I used the envelope theorem and definition of the value, v, of a team (equation 2). Equation 3 shows that to compare the marginal value of skills for managers and workers it is necessary to know the matching functions, since they determine the production partner, and thus technology, in either occupation. The matching functions are themselves tied to the occupational choice through market clearing. We thus face a functional fixed point problem.[23] Nonetheless I can characterize the two polar cases even without solving it explicitly.

First, let's consider *talent segmentation*. Due to the previous argument, this requires that most skilled individuals have a comparative advantage in being managers, hence that $v_1(x, w(x)) > v_2(m(x), x)\ \forall x$. This inequality can be rewritten as $\forall x$

$$
\underbrace{\frac{f_1(x, w(x))}{f_2(m(x), x)}}_{\text{Task Specialization}} \geq \underbrace{\frac{\alpha^*(m(x), x)}{\alpha^*(x, w(x))}}_{\text{Technology Complementarity}}. \tag{4}
$$

The single-crossing assumption *(1.5)* introduces a motive for task-specialization in the production function, and guarantees that the left hand side is larger than 1 for any x, $m(x)$, and $w(x)$. Since workers match upwards, the right hand side must also be larger than 1 in general, since an individual would use a higher technology as a worker. However, if the convexity of the cost is large enough, it approaches 1: the gain in terms of complementarity with a better technology by being a worker disappears since all teams use a similar technology. As a result, this inequality is satisfied, and talent is segmented by occupation, since skills are more rewarded in managerial positions.

Next, let's consider *talent segregation*. When the convexity of the cost is low, differences across teams in optimal technology - hence the right hand side in equation 4 - are large. The increase in marginal value of skills due to the fact that workers use of a better technology, in fact, more than compensate the loss from being assigned to a less skill-

[23] This same fixed point problem is circumvented by most settings that combine an occupational choice and matching, such as Garicano and Rossi-Hansberg (2006). The reason being that in these settings functional form assumptions guarantee that $v_1 > v_2$; the matching pattern is fixed ex-ante and there is no need to tackle the fixed point problem.

sensitive occupation. The result is that the jobs that have higher marginal value of skills are those in which individual use a higher technology, irrespective of their occupation. Talent is then segregated by technology.

For intermediate values of c, it is necessary to solve directly the fixed point problem in order to characterize the properties of the assignment. I next undertake this task in a setting with simplified functional forms.

2.4 A Quasi-Linear Model: a Measure of Talent Concentration

Functional form restrictions allow to completely characterize the optimal allocation even outside of the two polar cases (talent segregation or segmentation). This setting generates the testable predictions that I use in the empirical application of Sections 3.2.

Assumption 3 (Quasi-Linear Model). For $\gamma \in (0,1)$ and $\zeta \in \left(\frac{1}{2}, 1\right)$

$$
\begin{aligned}
f\left(x', x\right) &= x'\left(1+\gamma x\right) \\
c\left(x', x, a; \underline{a}, \bar{a}\right) &= \tilde{c}\left(a; \underline{a}, \bar{a}\right)\left(1+\gamma x\right) \\
\tilde{c}\left(a; \underline{a}, \bar{a}\right) &= \begin{cases} 0 & \textit{if } a < \underline{a} \\ \zeta\left(a-\underline{a}\right) & \textit{if } a \in [\underline{a}, \bar{a}]. \\ \infty & \textit{if } a > \bar{a} \end{cases}
\end{aligned}
$$

These functional forms satisfy all previous assumptions, with the exception of twice differentiability of c (due to kinks in the cost of technology). Lemma 3 therefore does not longer apply: two differently skilled managers might have identical optimal technology. In fact, only two technologies, $\overline{a}$ and $\underline{a}$, are used due to the piece-wise linearity of the cost function and to the constant complementarity of f. I call them respectively the advanced and traditional technologies.

Lemma 6 (Optimal Technology in Quasi-Linear Model). *The optimal technology for a pair (x', x) is given by*

$$
\alpha^*\left(x', x\right) = \begin{cases} \bar{a} & \textit{if } x' \geq \zeta \\ \underline{a} & \textit{if } x' < \zeta \end{cases}.
$$

The key parameter in this simplified setting is the ratio between the advanced and the traditional technology

$$\eta \equiv \frac{\bar{a}}{\underline{a}}$$

and plays the same role of the convexity of c in the more general setting hence it modulates the trade-off between occupation and technology. Notice, in fact, that the technology gap takes either value 1 or η:

$$\eta^* \left(x, m^* \left(x\right), w^* \left(x\right)\right) \in \{1, \eta\}.$$

As a result a low η correspond to high convexity of c, according to Lemma 5. Additionally, only the ability of the manager matters for the optimal technology of a team. This gives tractability, while still satisfying one of the main properties of the solution, namely that any individual x would use a (weakly) higher technology if he is a worker rather than a manager.

Total output produced in an allocation φ is given - just as in the benchmark case - by

$$Y\left(\varphi; \underline{a}, \bar{a}\right) = \int_{\mathbb{X}} g\left(m\left(x\right), x, \alpha\left(m\left(x\right), x\right); \underline{a}, \bar{a}\right) \omega\left(x\right) dx,$$

where $g\left(x', x, a; \underline{a}, \bar{a}\right) = af\left(x, x'\right) - c\left(x, x', a; \underline{a}, \bar{a}\right)$.

Lemma 7 (Homogeneity). *The planner's problem is homogeneous in $\underline{a}$, that is*

$$\max_{\varphi \in \mathbb{F}} Y\left(\phi; \underline{a}, \bar{a}\right) = \underline{a}\left[\max_{\varphi \in \mathbb{F}} Y\left(\phi; 1, \eta\right)\right],$$

and

$$\arg\max_{\varphi \in \mathbb{F}} Y\left(\phi; \underline{a}, \bar{a}\right) = \arg\max_{\varphi \in \mathbb{F}} Y\left(\phi; 1, \eta\right)$$

Lemma 8 (Uniqueness). *Under assumption 3, the optimal allocation φ^* is unique.*

By these two lemmas, the unique optimal allocation depends only on the technology gap η. I next study the comparative statics with respect to it. To do so, I first define a measure of the concentration of talent that captures how close the allocation of talent in φ^* is to the previously described case of segregation of talent by technology, and thus ultimately how similar are the individuals working together in a team.

As preliminary steps for the definition, I first introduce the optimal technology used by an individual of type x. This is given by a random variable, call it $A^*\left(x\right)$, that takes into account the probability that he is employed as a manager or a worker

$$A^*(x) = \begin{cases} \alpha^*(m^*(x), x) & \text{w.p. } \omega^*(x) \\ \alpha^*(x, w^*(x)) & \text{w.p. } 1\text{-}\omega^*(x). \end{cases}$$

Next, I consider an hypothetical optimal technology, A_p, that is built to satisfy segregation and preserving the same mass of individuals using $\bar{a}$ and $\underline{a}$ as in A^*, formally

$$\begin{aligned} \forall x', x: \; x' > x \quad &\Rightarrow \quad A_p(x') \geq A_p(x) \\ \int \mathbb{1}(A_p(x) = \bar{a})\, dx \quad &= \quad \int \mathbb{1}(A^*(x) = \bar{a})\, dx. \end{aligned}$$

I can then introduce the definition of the concentration of talent, that exploits the difference between the average ability of individuals that use technologies $\bar{a}$ and $\underline{a}$.

Definition 8 (Concentration of Talent). *The concentration of talent in an allocation φ^* is*

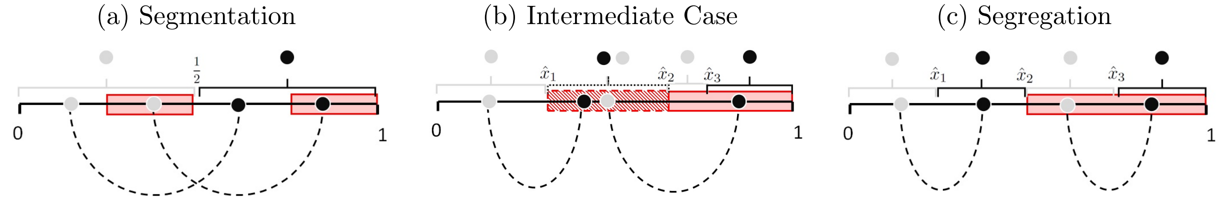
$$\pi = \frac{E(x|A^*(x) = \bar{a}) - E(x|A^*(x) = \underline{a})}{E(x|A_p(x) = \bar{a}) - E(x|A_p(x) = \underline{a})}. \tag{5}$$

I can then express π as a function of η, where the dependence comes from the fact that A^* and A_p depend on η, and describe the comparative statics of π with respect to η.

Proposition 3 (Concentration of Talent). *The concentration of talent $\pi(\eta)$ is non-decreasing in η. Additionally, there exist $\eta_1 < \eta_2$ such that if $\eta \leq \eta_1$ talent is segmented by occupation and $\pi(\eta) = \frac{1}{2}$; if $\eta \geq \eta_2$ talent is segregated by technology and $\pi(\eta) = 1$; and if $\eta \in (\eta_1, \eta_2)$ talent is nor segmented nor segregated and $\pi(\eta) \in \left(\frac{1}{2}, 1\right)$.*

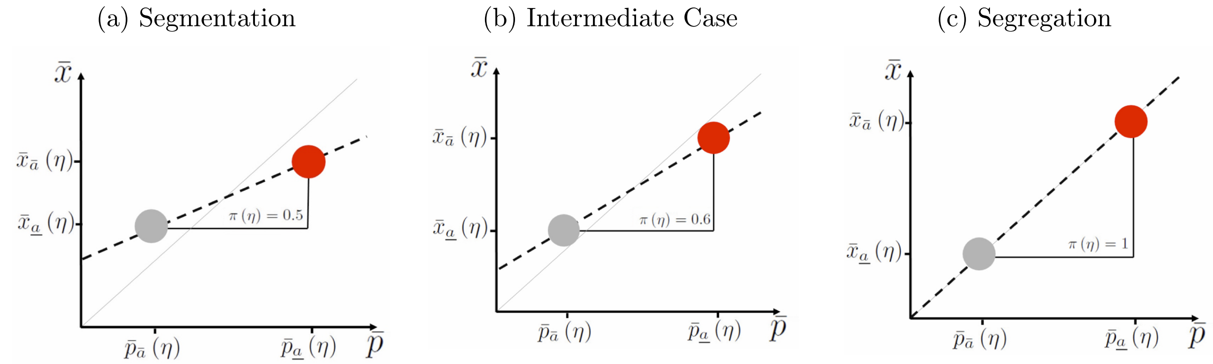
Figures 2 and 3 illustrate the optimal allocation and the measure π for three values of η. First consider 2c. It describes the occupational choice and technology used for a case in which $\eta \geq \eta_2$ and thus talent is segregated by technology: high skilled workers are allocated to the high technology $\bar{a}$. Among individuals that use the same technology, the most skilled are managers. The measure of concentration of talent, π, is constructed in Figure 3c. The figure shows $\bar{x}_a \equiv E(x|A^*(x) = a)$ plotted as a function of $\bar{p}_a \equiv E(x|A_p(x) = a)$ for $a \in \{\underline{a}, \bar{a}\}$. The measure of concentration of talent is the slope of the line. In this case, talent is segregated by technology, i.e. $\bar{x}_a = \bar{p}_a$ for all a, and as a result, $\pi = 1$. All the most skilled individuals are concentrated together in teams that use the advanced technology.

Figure 2: Allocation of Talent in the Quasi-Linear Model



Notes: the squared brackets put together individuals with the same occupation. Workers are highlighted with light grey square brackets, and managers with black ones. Dotted brackets signal mixing: some are workers and some managers. The red regions covers the set of individuals using the frontier technology $\bar{a}$. The red striped regions are present in mixing area on which the workers use $\bar{a}$, while the managers use $\underline{a}$. Dotted lines connect examples of workers and managers that are together in a team.

Figure 3: Measure of Concentration of Talent



Notes: in each figure I plot $\bar{x}_a$ as a function of $\bar{p}_a$. By definition, the measure of the concentration of talent is the slope of the line.

Now consider Figures 2a and 3a. They plot the allocation for $\eta \leq \eta_1$ and thus talent segmented by occupation: all the most skilled are managers and teams thus put together individuals of more different abilities. Only the most skilled managers, and the workers matched with them, use the advanced technology $\bar{a}$. Workers that use $\bar{a}$ are less skilled than some managers that use $\underline{a}$. As a result: $\bar{x}_{\bar{a}} < \bar{p}_{\bar{a}}$ and $\bar{x}_{\underline{a}} > \bar{p}_{\underline{a}}$ and the measure of concentration of talent is smaller than 1. In particular, it turns out that, since teams are formed by two individuals, $\pi = \frac{1}{2}$ when there is segmentation. This holds irrespective of the value of ζ (the cutoff that defines the manager that uses advanced technology).

Last, in Figures 2b and 3b, I show an intermediate case $\eta \in (\eta_1, \eta_2)$ in which there is neither segmentation nor segregation. Some fairly skilled individuals use the traditional technology, although less than before, and at the same time, some workers that use the advanced technology are more skilled than managers that use the traditional one. The result is that the concentration of talent has an intermediate value.

The ones above are just three special cases of the optimal allocation, that is completely characterized in Section A.8. As a result of the stricter functional form assumptions, I am able to directly solve for the fixed point in $\{\omega^*, \mu^*, m^*, w^*\}$. There are two key steps to do so. The *first* one is to notice that the optimal allocation can take only one of five possible shapes, where a shape is defined by the number of cutoffs that separate subsets of the type space where the occupational choice differ. For example, one shape is when talent is segmented by occupation. There is only one cutoff type and types below it are workers and types above it are managers. The *second* one is to note that the slope of the matching function $m(x)$ depends on the fraction of workers at x, $\omega(x)$, and fraction of managers at $m(x)$, $\mu(m(x))$. This allows to solve for ω and μ that guarantee that the marginal value of managers and workers is identical. A condition required over regions of the type space in which the optimal allocation dictates mixing, as in Figure 2b.[24]

Let me then discuss the transformation from one polar case to the other (details are in Section A.8). Consider the case with segmentation of talent in Figure 2a. When η is smaller than η_1, the gap between the advanced and traditional technology is small, thus every team uses a similar technology and high skilled individuals are assigned to be managers. As η increases it becomes more skill rewarding for an individual to become a worker and use $\bar{a}$, rather than being a manager and using $\underline{a}$. The reward from being a manager rather than a worker depends also on the production partner, and thus on the matching function m. For η slightly bigger than η_1, only the lowest skilled among the previous managers finds it optimal to become a worker. As η increases further, more and more individuals who, if they were managers would use $\underline{a}$, become workers in order to get access to $\bar{a}$. When $\eta \geq \eta_2$, access to the advanced technology drives the assignment, since the gap between technologies is so large that any individual would enjoy a higher marginal value of his skills if he is the worker using $\bar{a}$ rather than a manager using $\underline{a}$. The optimal allocation thus resembles a dual economy: within each technology, there is talent segmentation, but skills are segregated by technology.

The smooth transition from one polar case to the other as η increases is in contrast with most frictionless matching models, which feature a discrete jump between two polar cases. As an example, consider the case of a CES production function that leads to a perfectly positive or perfectly negative assortative matching depending on the value of the elasticity of substitution (see, for example, Grossman and Maggi (2000)).

[24] A similar insight is present in Low (2013) for a case with bipartite matching.

2.5 Model Extension: Self-Employment

Last, I allow individuals to choose whether to form a pair or to work alone. This extension provides an additional prediction - on the average skill level of self-employed - that can be validated in the data, and sheds light on the motives that drive self-employment in this environment. I start from the quasi-linear setting and model being self-employed, relative to being a manager, along the lines of Garicano and Rossi-Hansberg (2004). Under this setup, self-employment is a more skill intense occupation than being a worker, but less so than being a manager. Additional details and an analysis of the incidence of self-employment are left to Section A.9.

Assumption 4 (Self-Employment). *Output of a self-employed individual of ability x with technology a is equal to $af_s(x) - \tilde{c}(a)$, where f_s satisfies:* $\min_{x' \in \mathbb{X}} f_1(x, x') \geq f_s'(x) \geq \max_{x' \in \mathbb{X}} f_2(x', x)$.

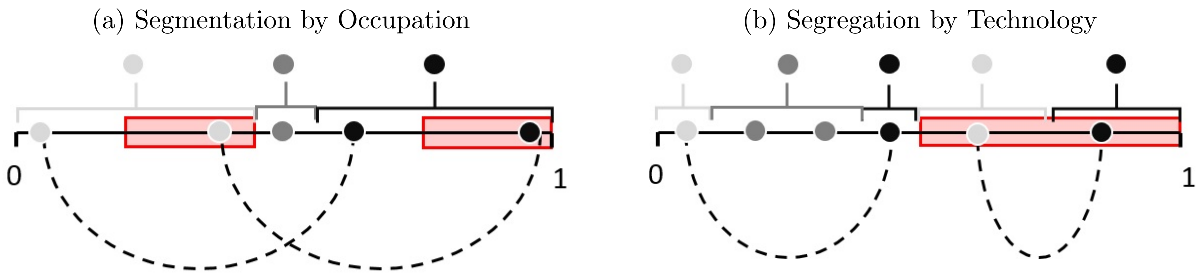
As a result of this assumption, in the optimal allocation, among individuals who use the same technology there is going to be talent segmentation by occupation: self-employed are more skilled than workers, but less so than managers. If everyone uses a similar technology, as when $\eta \sim 1$, the self-employed are thus going to be as skilled as the average. In Figure 4a, I depict this allocation. Instead, if the gap between the advanced and traditional technology, η, is large, then the complementarity between labor and technology implies that a worker using the advanced technology $\bar{a}$ has a higher marginal value of his talent than a self-employed using the traditional technology $\underline{a}$. Additionally, self-employed must be less skilled than managers, due to the fact that they have lower marginal value of skills. Since - due to the functional form assumptions - at least some manager chooses $\underline{a}$, this implies that no self-employed individual uses the advanced technology $\bar{a}$. This second case, which displays talent segregation by technology, is shown in Figure 4b. Self-employed are now less skilled than the average.

Definition 9 (Selection of Self-Employed). *The selection into self-employment under an allocation φ^* is given by $\vartheta = \int x\sigma^*(x)\,dx - \int x(1-\sigma^*(x))\,dx$, where $\sigma^* : \mathbb{X} \to [0,1]$ is the optimal occupational choice function that assigns to each type x the fraction of self-employed.*

This measure captures the difference between the average type of individuals who are self-employed or work in teams. As before, the model is homogeneous in $\underline{a}$, and I can thus express ϑ as a function of η and provide the main result of this section.

Proposition 4 (Selection of Self-Employed). *The selection of self-employed $\vartheta(\eta)$ is non-increasing in η.*

Figure 4: Allocation of Talent with Self-Employment



Notes: The squared brackets put together individuals with the same occupation. Workers are highlighted with light grey square brackets, self-employed with dark grey ones, and managers with black ones. The red regions covers the set of individuals using the frontier technology $\bar{a}$. Dotted lines connect workers and managers that are together in a team.

2.6 Taking Stock of the Theory

The theoretical analysis shows that the convexity of the cost of technology determines the shape of the economy's structure. Specifically, the convexity modulates the role of occupation and technology in determining the main purpose of team production and thus the optimal assignment. When the cost is sufficiently convex, so that most teams use the same technology, the allocation resembles the familiar structure from occupational choice problems: low skilled individuals are workers and high skilled ones are managers. The main purpose of team production is to put together differently skilled individuals to allow the most able ones to task-specialize.

The optimal allocation changes when the cost is less convex. The allocation in this case is asymmetric. Some teams attract skilled individuals (both managers and workers) and use advanced technologies. Some other teams instead are left with low skilled ones and use traditional technologies. Teams now concentrate skilled individuals to reap the benefits from the complementarity between advanced technologies and their skills. As a result, there is larger dispersion of talent and technology in the economy. Taking a first step towards the interpretation of the theoretical results, which will be provided in the next section, the possibility to adopt frontier technology leads to an endogenous formation of a dual economy. Teams that adopt advanced technology attract the most skilled individuals, leaving the rest of the economy with low talent and thus low technology.[25]

[25]This feature of the model resembles a mechanism roughly outlined by Acemoglu (2015) (page 454)

3 Evidence: Distance to the Frontier and Allocation of Talent

In this section, I apply the theory to a cross-country analysis and evaluate its empirical content, confronting its predictions with previous literature and with micro data from several countries.

First, I provide an interpretation of the convexity of the cost of technology in terms of the distance of a country to the technology frontier. This interpretation allows to generate empirical predictions on differences in the concentration of talent across countries.

It has been shown in the literature that cross-country variation in GDP per capita at one point in time is to a large extent explained by differences in average used technology (see Klenow and Rodriguez-Clare (1997) and Comin and Ferrer (2013)). As a result, individuals in countries relatively poor with respect to the frontier can, in principle, do large improvements in their technology at a low cost by adopting the existing technology of rich countries. They have in fact a unique opportunity: they can adopt technologies that are much more advanced than the ones they could otherwise produce, given their level of development. Instead, for individuals in countries already on the frontier, it is relatively more costly to do large technology improvements, since it would require to invent new technologies. Motivated by this, I interpret relative GDP per capita across countries as a proxy for how expensive large technological improvements are, hence for the convexity of the cost of technology.[26] Intuitively, differences in convexity capture the fact that innovation (on the frontier) is a gradual process, while adoption (far from frontier) can be rapid and direct.

Two comments are in order. *First*, I have not mentioned cross-country differences in the *level* of the cost of technology, but only in its convexity. The level of the cost of technology does not affect the allocation of talent, but does change average used technology and total output. Of course, therefore, countries farther from the frontier must also have a higher cost for *each* technology, in order to capture the fact that their technology level is lower. Similarly, the growth path of a country that remains steadily close to the technology frontier should be interpreted as a decrease in the cost of each technology, keeping constant

---

for the case of physical capital. He argues that it might be interesting to explore the possibility that "*if technologies imported from the world technology frontier have undergone much improvement only in high capital–labour ratios, then despite the relatively high price of capital, some firms in developing economies may end up choosing to operate at these high capital–labour ratios, leaving even lower capital– labour ratios for the rest of the economy.*"

[26] In the quasi-linear model in Section 2.4 the cost of technology has an even more direct interpretation. The frontier technology is $\bar{a}$ and it is identical for each country at a given point in time. The country specific traditional technology is instead $\underline{a}$. Countries farther from the frontier have lower $\underline{a}$, thus higher $\eta = \frac{\bar{a}}{\underline{a}}$.

the convexity. For example, this interpretation implies that, along the growth path of the United States, the convexity (hence the size of the support of used technology) remains fairly constant, but new technologies replace the old ones as they become obsolete.[27] *Second*, for the sake of clarity, the discussion so far has not emphasized cross-country differences in human capital. I load all cross-country differences on one primitive, the cost of technology. This should be interpreted as the net cost, which takes into account differences in its return. A high cost of technology is, in fact, isomorphic to a low return from it, which might be driven by the low level of human capital in countries far from the frontier. In Section A.14, I provide a microfoundation of the cost function to highlight how cross-country differences can be modeled as a race between the country's human capital and the technology frontier.

Next, I use this interpretation to confront the theory with data on cross-country differences in the organizational structure of their economies. The empirical analysis will comprehend three parts. *First*, in Section 3.1, I review the literature and show that the theory qualitatively replicates known empirical facts of developing countries. *Second*, in Section 3.2, I use household level data to construct an empirical measure of the concentration of talent as defined in the model, that I then compare across countries and time. I also use the same data to show, motivated by the model extension, cross-country differences in the selection of self-employed. *Third* and last, in Section 3.3, I provide additional evidence in support of the theory from firm level data. Firm level data have both advantages and disadvantages vis-à-vis the household level ones: they allow to have a more narrow notion of who works with whom, hence also to validate some of the model assumptions, but at the same time are plagued by cross-country comparability concerns.[28]

3.1 Confronting the Theory with Existing Empirical Evidence

The theory qualitatively replicates known features of developing countries. The model, in fact, predicts that as a result of a less convex cost of technology, developing - hence farther from the frontier - countries should be organized along an asymmetric equilibrium. Some teams are formed by high skilled individuals and use the advanced technology, while some other teams are formed by low skilled individuals and use the traditional one. This is in contrast with developed countries, where all teams should use a similarly advanced

[27] In Figures A.7 and A.8 I illustrate few examples of technology cost functions consistent with this interpretation.

[28] Specifically, I use the World Bank Enterprise Surveys. These data have been recently used by Asker et al. (2014), who also discuss some of the cross-country comparability concern.

technology and are formed by high skilled individuals in managerial positions and low skilled ones as their workers.

The model, therefore, predicts that in developing countries there is larger dispersion of productivity, both directly and through polarization of skill composition of teams. This fact is consistent with evidence documented among others by Caselli (2005), Hsieh and Klenow (2009), and Adamopoulos and Restuccia (2014). Moreover, while the overall level of technology is lower, the model predicts that, even in developing countries, within some niches of the economy, the technology used is the same as in developed ones. This is consistent with Rodrik (2012) that shows evidence on unconditional cross-country convergence within formal manufacturing, and with Comin and Ferrer (2013), that shows that modern technologies quickly diffuse to all countries, but the penetration rate is much lower in poor countries. Next, the theory predicts that in developing countries some very low skilled individuals are employed in managerial positions. This is consistent with evidence in Bloom and Van Reenen (2010), that documents a thick left tail of poorly managed firms and that firms with more educated workers have better management practices. More broadly, the asymmetric equilibrium resembles a dual economy, and duality is a feature often associated with developing countries (see for example La Porta and Shleifer (2014)). Last, the model extension predicts many low skilled self-employed in developing countries. In the model, these individuals become self-employed because those types that would be their managers in a developed country choose to be workers themselves to get access to the advanced technology. This type of self-employment resembles the notion of subsistence entrepreneurship in developing countries, as emphasized by Schoar (2010), Ardagna and Lusardi (2008) and Banerjee and Duflo (2011).

Most theories that provide an explanation for these empirical facts usually attributes them to larger market frictions in developing countries. This paper instead proposes a complementary explanation that differs in three aspects. *First*, these features of developing countries are generated in a first best world.[29] *Second*, they are associated with the distance to the frontier and the resulting option of technology adoption, rather than with a low level of development. *Third*, they are tightly connected, and in fact either caused or exacerbated, by a different assignment of individuals to teams.[30] Differences in the

[29]Other recent papers proposed competitive explanation for cross-country differences in productivity dispersion. For example, Lagakos (2013) argues too little car's ownership may lead to low retail productivity and Young (2013) argues that spacial sorting of workers can explain the rural-urban wage gap.

[30]One example makes more vivid differences in allocation of talent. Individuals working as cashiers in supermarket in India are on average more skilled than 80% of the population (data used are described in the next section). However, they do not perform a complex task. They would be defined "workers" in the model. We can also find in the same data supermarket managers. They are mildly more skilled than the

allocation of talent are a new feature of economic development, that has been previously overlooked. I next show that it has empirical content.

3.2 Evidence on Allocation of Talent from Household Level Data

Next, I use censuses and labor force surveys from 63 countries around the world and document cross-country differences in the allocation of talent.

3.2.1 Empirical Strategy

The quasi-linear model predicts (Propositions 3 and 4) that in countries with a less convex cost - hence, along the previous interpretation, farther from the frontier - talent is more concentrated and self-employed are more negatively selected.

The main empirical challenge is to construct, for each country, a scalar statistic that summarizes the information in the data on the concentration of talent. Labor force surveys in fact do not include information on production teams, i.e. on who works with whom. I build a measure that follows closely the definition of concentration of talent in the theory. Concentration of talent in the model (Equation 5) is given by the difference between the average skill of individuals using the advanced and the traditional technologies. This difference, in turn, is informative about team composition because teams sort into technology used depending on the ability of their members. In the data, I use a similar insight. Since I do not directly observe individual skills and technology choice, I use the concentration of educated individuals across sectors. The model predictions will hold for concentration of education across sectors if two assumptions are satisfied. *First*, individual ability and education years must be positively correlated. This assumption allows to use education as a proxy for ability. *Second*, teams must sort into sectors where similarly skilled teams are. This second assumption allows to use a sector as a proxy for the technology used by a team.

The *first* assumption allows to measure skill using an individual's years of education.[31] More precisely, for each country-year survey, I use the distribution of education years to compute a measure of skill, denoted as $\hat{x}$, between 0 and 1. For example, a value of $\hat{x} = 0.5$ means that this individual is more educated than 50% of the population in that country-

cashiers. What about a country close to frontier? In the United States cashiers are below the median of the population, while their managers are above the 70%. Thus, in U.S. simple occupations are taken by low skilled, which are working together with high skilled which take managerial roles.

[31] A prominent alternative measure of skills is individual income. I choose to use education years for three reasons: (i) in most countries, only wage income is available, and only a fraction of the developing country workforce receives a formal wage; (ii) even when non-wage income data are available, it is hard to compare with wage income, since it might capture non individual returns (e.g. family labor); (iii) income measures are available only for a subset of countries.

year survey. There are few concerns. First, education might measure skill with noise.[32] This is not a major problem, since due to the fact that I compare average skill by sector, measurement error will not bias the results. In particular, since there is a large number of individuals within each sector, measurement error should wash out. Second, although in the model skill is taken as an exogenous characteristic, education is the outcome of an endogenous choice. In order to address this concern, in Section A.10, I extend the model to allow for an endogenous education choice, given exogenous differences in ability. I show that under standard assumptions, years of education are increasing in ability, and the relationship between the convexity of the cost function and the optimal allocation is still present. Third, education might be a noisier measure of skills in poor countries, such as if some talented individuals are credit constrained and hence do not go to school even if they would have a high return from it. However, as I discuss in farther detail in Section 3.2.5, this would attenuate my results.

The *second* assumption allows to measure technology using the industry in which an individual works. In particular, I rank the implied technology used by each industry according to the average education of the individuals working in it. Therefore, according to my measurement, if an industry has a more educated workforce, it also has a more advanced technology. This is consistent with the model prediction that more skilled teams sort into higher production technologies.[33] It is also consistent with previous literature that argues that some sectors use technologies with higher degree of skill complementarity (see for example Buera et al. (2015)) and that documents large productivity gaps across sectors in developing countries.[34] Nonetheless a concern remains, that is, my result would be biased if industry is a worse proxy for technology used in countries *closer* to the frontier. I address this in Section 3.2.5.

The selection of self-employed can be directly measured in the data. I use again education as a measures of skill, and I observe whether individuals are self-employed. I can thus calculate the average skill of those who are self-employed and compare it with those who are not.

Last, equipped with the empirical measures of interest, I can compare them across countries. As already discussed, I use relative differences in GDP per capita as a proxy

[32] I use education to make an ordinal comparison across individuals within a country. As a result, any cross-country comparability concern is alleviated. For example, the concern of Hanushek and Woessmann (2008) - i.e. that cross-country education differences understate those in cognitive ability - does not apply.

[33] An extension of the model in which teams explicitly choose in which industry to locate, can replicate perfect sorting of teams to industries in the presence of economies of agglomeration at the industry level.

[34] See for example Caselli (2005) and Gollin et al. (2014) on agriculture productivity gaps, and Acemoglu and Zilibotti (2001) for a more disaggregated study.

for the convexity of the cost of technology - hence η in the quasi-linear model. More in details, it can be useful to recall that $\eta = \frac{\bar{a}}{\underline{a}}$. We can then - consistently with the previous discussion - interpret $\bar{a}$ as the frontier technology, and $\underline{a}$ as the country specific traditional one. In the model, for a given level of $\bar{a}$, total output is increasing in $\underline{a}$ (thus decreasing in η), since the higher the value of $\underline{a}$, the lower the cost of each technology. This formally motivates to use GDP per capita to measure proximity to the frontier and yields the *first* and main empirical comparison, that is, the comparison of rich and poor countries at one point in time. Similarly, the model predicts that when $\bar{a}$ is low - i.e. in the past when the frontier technology was less developed - $\underline{a}$ must be higher (thus η must be lower) in order to achieve a given level of GDP. This yields the *second* empirical comparison, that is, the comparison of poor countries today to countries that are currently rich at the time when the latter were at similar level of GDP per capita as the former today. Last, due to the homogeneity of output, if the GDP of a country grows faster than the frontier, then η should fall. This yields the *third* comparison: following a country over time as it approaches the frontier.

3.2.2 Data (Household Level)

I use large sample labor force surveys and censuses available from Integrated Public Use Microdata Series, International (IPUMS). The data cover 63 countries of differing income levels, from Rwanda and Tanzania to Switzerland and United States. Part of my analysis focuses on the South Korean growth experience, for which I use data from the Korean Longitudinal Study of Ageing (KLoSA) and, to perform robustness checks, the Korean Labor and Income Panel Study (KLIPS). All GDP per capita data are taken from the Penn World Table version 8.0.

In the IPUMS data, there are three main variable of interest: education, industry, and employment status. Completed years of education is coded from the educational attainment variable, and industries are standardized by IPUMS to be comparable across countries. Their industry definitions span 12 industries. Last, employment status records indicate whether an individual is a wage-worker, self-employed or an employer. In order to minimize comparability concerns, I restrict the sample to include only males, head of households and between 18 and 60 years old. For the baseline results, I also exclude self-employed, since they do not work in teams. All data are representative of the entire population from which they are drawn. Robustness checks and alternative sample selections are in Section 3.2.5. Data details are in Section A.11.

KLoSA is a survey gathered with the purpose of understanding the process of population aging in Korea. It has a sample size of approximately 10,000 individuals and it is representative of individuals older than 45. The survey has bi-annual frequency and started in 2006. Hence, it does not allow to directly trace the growth miracle in South Korea. However, there is a job supplement that asks the complete history of jobs for each individual. In particular, this contains information on the industry in which the respondent works, their employment status, and their education. Using these data, I retroactively construct cross-sections for each year from 1953 to 2006. There is one obvious concern with this procedure: average age of the individuals in my dataset changes over time by construction, and thus I may confound life-cycle and time-series trends. In Section A.12, I perform robustness checks to address this concern.

3.2.3 Concentration of Talent

I next use the labor force surveys to document the first empirical prediction.

Empirical Prediction 1 (Concentration of Talent). *The farther a country is from the technology frontier, the more educated individuals are concentrated into sectors.*

I build the empirical measure of concentration of talent in five steps. First, I compute a normalized measure of skill using the country-year specific cumulative density function of years of education, $\hat{x}_i = F(s_i)$, where s_i is the schooling year of individual i.[35] Second, I compute the average skill in each industry j: $\bar{\hat{x}}_j = E[\hat{x}_i | I_{ij} = 1]$, where I_{ij} is an indicator function equal to 1 if individual i works in industry j. The ranking of industries according to their average skill level provides the measure of the technology rank, that is, I rank industries with higher *average* education as having a higher technology. Third, I build a perfect sorting counterfactual in which I assign, keeping industry size constant, all the most skilled individuals to the industry with the highest average education (hence measured technology). All the highest skilled ones among the remaining workforce are then assigned into the second highest and so on.[36] Fourth, I compute the average skill in each industry under the perfect sorting counterfactual: $\bar{\hat{p}}_j = E\left[\hat{x}_i | I^C_{ij} = 1\right]$, where I^C_{ij} is the constructed indicator function. Fifth and last, I regress $\bar{\hat{x}}_j = B_0 + B_1 \bar{\hat{p}}_j + \varepsilon$.[37] The measure of

[35]The variable s takes only finite number of values. I therefore renormalize $\hat{x}$ in such a way that the lowest skilled individuals have ability $\hat{x} = 0$ and the highest skilled ones ability $\hat{x} = 1$. Results are robust to alternatives and available upon request.

[36]More formally, let $\hat{\mathbb{X}}_j$ be the observed set of individuals in industry j, of mass $v\left(\hat{\mathbb{X}}_j\right)$, then $\bar{\hat{x}}_j = E\left[\hat{x} | \hat{x} \in \hat{\mathbb{X}}_j\right]$. The counterfactual sets are given by $\hat{\mathbb{P}}_j \equiv \left\{\hat{x} : \hat{x} \in \left[\hat{P}_1(j), \hat{P}_2(j)\right]\right\}$ where $\hat{P}_1(j) \equiv \sum_{k:\, \bar{\hat{x}}_k < \bar{\hat{x}}_j} v\left(\hat{\mathbb{X}}_k\right)$ and $\hat{P}_2(j) \equiv \hat{P}_1(j) + v\left(\hat{\mathbb{X}}_j\right)$. Then $\bar{\hat{p}}_j = E\left[\hat{x} | \hat{x} \in \hat{\mathbb{P}}_j\right]$.

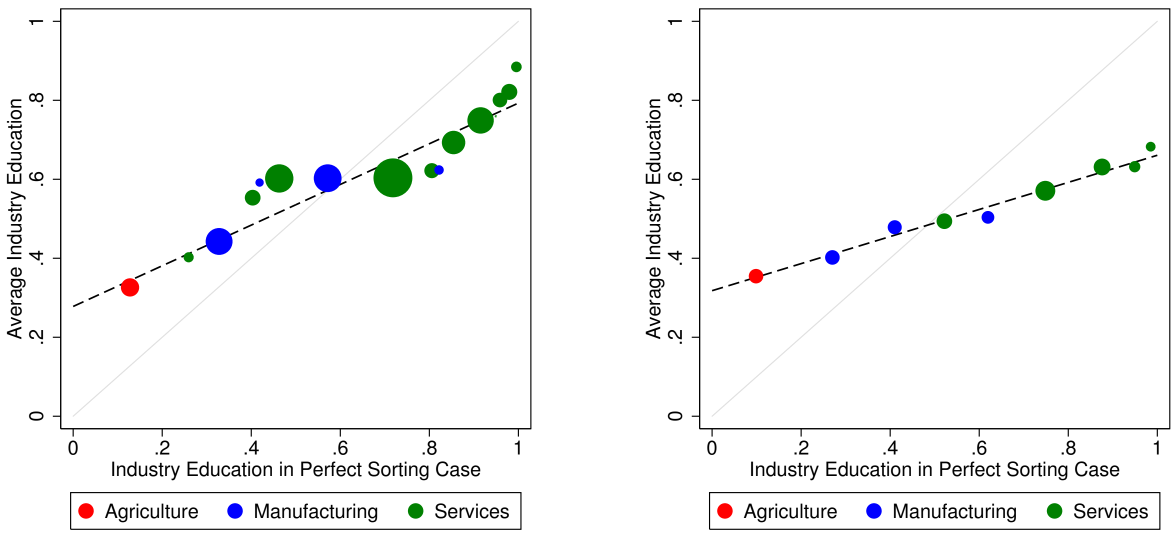
[37]I weight the regression by the number of individuals in each industry j. Unweighted results are similar and available upon request.

concentration of talent is $\hat{\pi} = \hat{B}_1$.

Note that by the definition of the least squares estimator, $\hat{\pi} = \frac{E\left[\bar{\hat{x}}_j - \bar{\hat{x}}_{j'}\right]}{\bar{\hat{p}}_j - \bar{\hat{p}}_{j'}}$. This measure is thus tightly linked to the definition of concentration of talent in the model, $\pi = \frac{\bar{x}_{\bar{a}} - \bar{x}_{\underline{a}}}{\bar{p}_{\bar{a}} - \bar{p}_{\underline{a}}}$. In fact, the only difference is that the empirical measure takes the average across different sectors, since there are more than two sectors in the data.

Figure 5: Construction of the Measure of Concentration of Talent

(a) Brazil, 2010 (b) United States, 1940



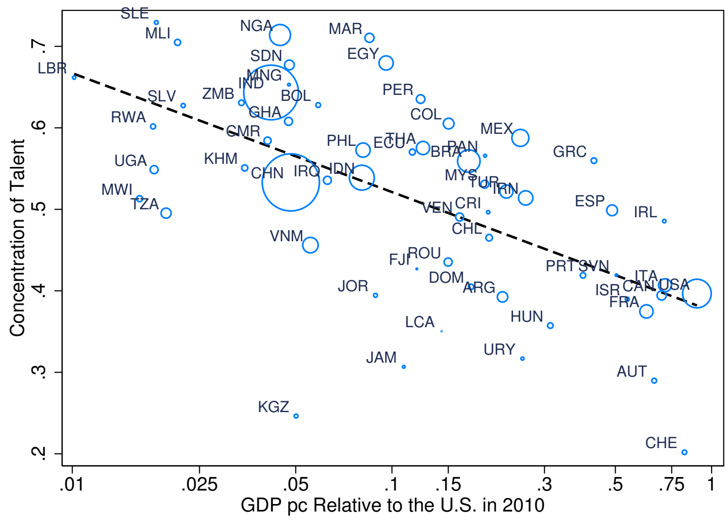
In Figure 5, I show two examples, Brazil in 2010, and United States in 1940, to illustrate how $\hat{\pi}$ is constructed. I plot the average skill in an industry, $\bar{\hat{x}}_j$, as a function of the skill in the perfect sorting counterfactual, $\bar{\hat{p}}_j$. Each dot in the figure corresponds to an industry and its size increases in the number of individuals there employed. A linear regression $\bar{\hat{x}}_j = \alpha + \pi \bar{\hat{p}}_j + \varepsilon$ fits the data well.[38] Last, notice that Brazil in 2010 has similar GDP per capita as U.S. had in 1940, however it has a higher concentration of talent (the regression line in the figure is steeper). This is consistent with the fact that Brazil in 2010 is farther from the technology frontier than the United States was in 1940.

I next build this measure of concentration of talent for each country-year pair in my sample and document systematic differences between countries depending on their distance from the frontier.[39] I do so along the three previously explained empirical comparisons.

[38] The average R^2 across countries from this regression is ~ 0.9 similarly in rich and poor countries.

[39] For brevity, I do not include Figure 5 for all the countries of my sample. They are however available on my website at https://sites.google.com/a/yale.edu/tommaso-porzio/home.

Figure 6: 1^{st} Comparison: Cross-county Differences in Concentration of Talent



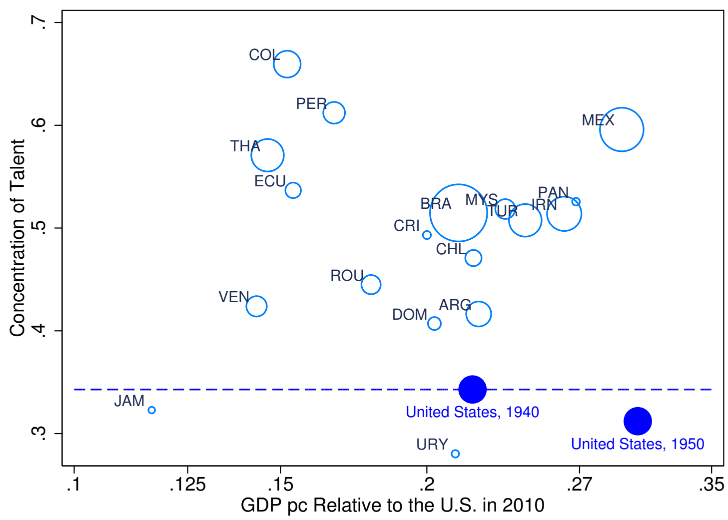
Notes: light blue circles show how populated each country is. The dotted line is the fit from a regression of concentration of talent on log of GDP per capita. The regression is not weighted by population, since I treat each country as one observation.

First, I compare countries with higher and lower level of GDP per capita by plotting the concentration of talent as a function of GDP per capita relative to one of the United States in 2010.[40] The result is shown in Figure 6: poor countries, i.e. those farther from the technology frontier, have larger concentration of talent.[41] In order to interpret the magnitude of cross-country differences, it is useful to conduct the following thought experiment. Consider a country with two industries and two types of workers, high and low-skilled. Each industry is of equal size, and half of the population is high-skilled and half is low-skilled. If $\hat{\pi}$ in this economy is equal to 0 it means that half of the high skilled individuals are in each industry. If $\hat{\pi} = 1$ it means that all high skilled individuals are in one industry, which I call the modern one. If $\hat{\pi} = \frac{1}{2}$, instead, 75% of the high skilled are in the modern industry, and hence a high-skilled individual is three times more likely to work in the modern industry. Using this thought experiment, the estimates imply that high-skilled individuals in poor countries would be approximately five times as likely to work in the modern industry as the traditional one, while in rich ones, they would be only twice as likely.

[40] For countries for which I have more than one cross-section, I compute the concentration of talent for each cross-section and take the average. Other alternatives yield similar results. Likewise, I have experimented with different measures of GDP per capita, which also does not affect the results.

[41] The regression line has a positive slope that is significant at 1% level. I am currently calculating standard errors that take into consideration that the measure of concentration of talent are themselves estimates. Estimates will be updated when available, but given the large sample sizes and the strong fit, significance is unlikely to be affected. This same argument holds - and will be omitted henceforth - for all cross-country regressions.

Figure 7: 2^{nd} Comparison: Developing Countries today and U.S. in the past



Notes: light blue circles show how populated each country is. The blue dotted line is at the level of concentration of talent of United States in 1940.

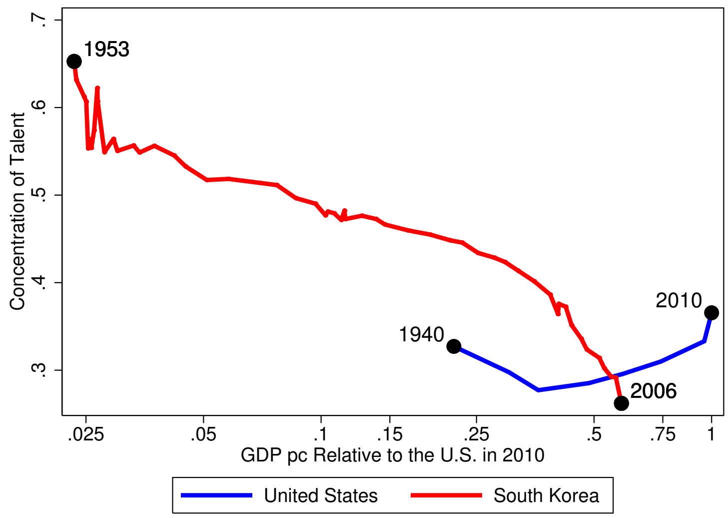
Second, I show that poor countries have higher concentration of talent than currently rich ones when they were at a comparable level of development. This alleviates the concern that the observed differences might be driven by differences in the level of development rather than the distance to the technology frontier. Specifically, I have comparable census data for the United States every ten years from 1940 to 2010.[42] GDP per capita in the United States in 1940 is comparable to the one of many middle income countries - such as Brazil, Mexico, Turkey, and Argentina - that I observed in my sample between 2000 and 2010. [43] I observe in fact 18 such countries, and among them, 16 have a higher concentration of talent than the U.S. used to have, as shown in Figure 7.[44]

[42]Before 1940 census data did not report education years.

[43]I have computed a similar comparison for France, for which I have data to calculate the measure of concentration of talent back to 1962. The results are very similar and available upon request.

[44]A permutation test of the null hypothesis that the U.S. is not different rejects the null hypothesis more than 99% of the time.

Figure 8: 3^{rd} Comparison: South Korea as it Approaches the Frontier



Third, I study the growth path of South Korea, a particularly interesting country due to fact that it converged to the frontier in the past 50 years. South Korea GDP per capita relative to the one of the United States increased in fact from only 7% to almost 60%.[45] In Figure 8, I plot the growth path for concentration of talent across sectors for both countries.[46] U.S. concentration of talent remained fairly constant along the growth path, consistent with fact that U.S. has been growing over this period constantly as a world leader, i.e. on the technology frontier. South Korea's concentration of talent instead decreased steeply. This last comparison alleviates the concern that cross-country differences might be driven by time invariant country characteristics that are correlated with GDP per capita.

3.2.4 Selection of Self-Employed

Next, I document the second empirical prediction of the model.

Empirical Prediction 2 (Selection of Self-Employed). *The farther a country is from the frontier, the lower the relative education level of its self-employed individuals.*

The IPUMS data report whether an individual is self-employed or a wage/salary worker. In order to build the empirical measure of self-employment, I use my measure of skill, namely the normalized number of years of education, and compute the difference

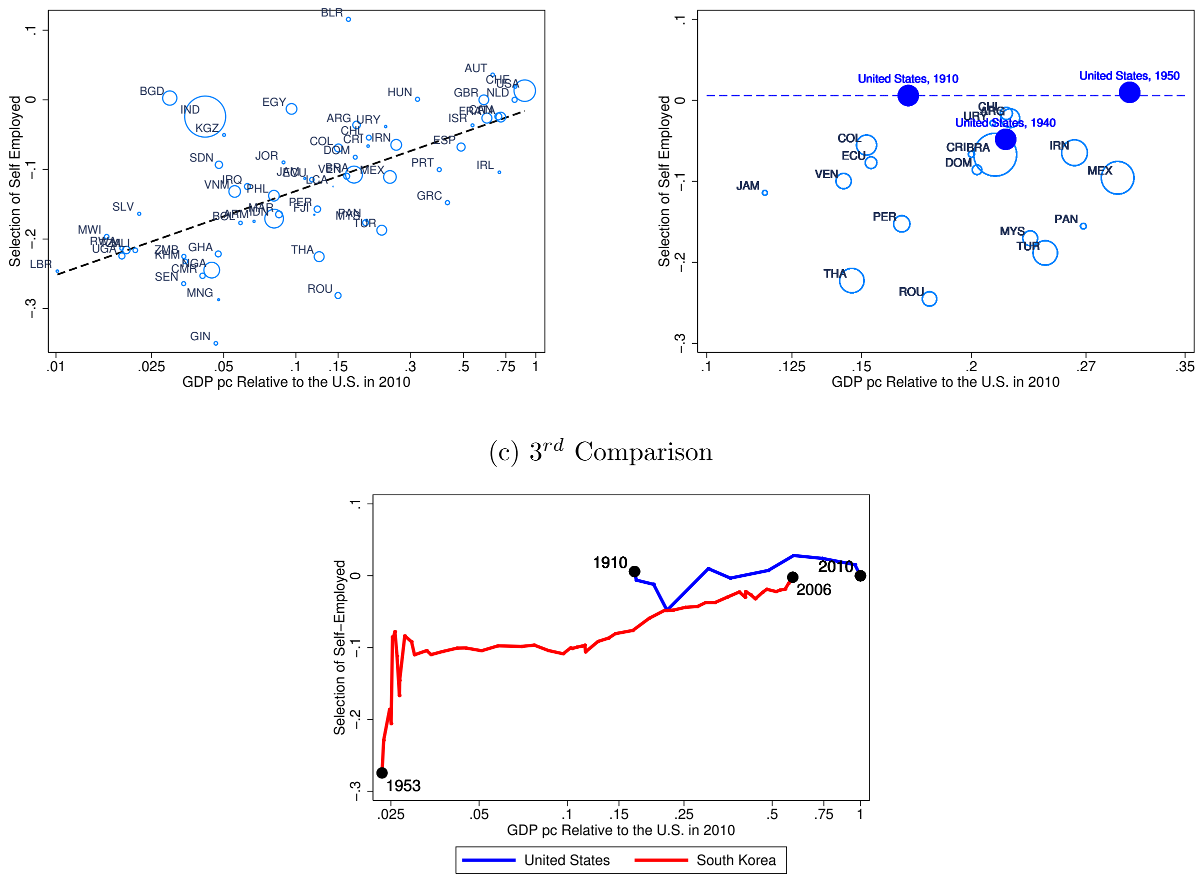
[45] These facts can be appreciated in Figure A.6.

[46] I use concentration of talent across sectors (hence aggregating industries to agriculture, manufacturing, or services) because industry measure is not comparable for United States and South Korea. Results with the concentration of talent across industries are nonetheless comparable and available upon request.

between the average skill of self-employed and the average skill of wage-workers. More details are in the Section A.11.

Figure 9: Selection of Self-Employed

(a) 1st Comparison (b) 2nd Comparison



Notes: light blue circles show how populated each country is. The blue dotted line in panel (b) is at the level of selection of self-employed of United States in 1910.

I follow the same three empirical comparisons as for the measure of concentration of talent. Results are shown in Figure 9. Poor countries - farther from the frontier - have more negatively selected self-employed, both than rich ones today and than United States when the latter were at similar level of development.[47,48] Last, as South Korea approached the frontier, the selection of its self-employed decreased.[49]

[47]A regression line from a cross-country regression has a negative slope that is significant at 1% and the permutation test of the null hypothesis that the U.S. is not different, rejects the hypothesis more than 99% of the times.

[48]I can compute the selection of self-employed for the United States as far back as 1910. However, from 1910 to 1940 information on number of years of education was not available. I use instead a measure of literacy rate.

[49]This result should be interpreted with particular caution. The reason being that the results for South

3.2.5 Robustness and Alternative Interpretations

I here explore the robustness of the main empirical result, that is the relationship between the concentration of talent and the distance to the technology frontier.

One main concern is that the underlying patterns of matching are identical across countries, and the observed differences are driven by mis-measurement resulting by the failure of either one of the two working assumptions. I argue, however, that failures of the assumptions would most likely attenuate my results. First, the documented cross-country patterns could be observed if individuals are perfectly matched on ability in all countries, and in poor countries, more able individuals are more schooled, while in rich countries the relationship between education and skills is non-monotonic.[50] This hypothesis however is at odds with the often made claim that in developing countries schooling choices are more constrained. (See for example Mestieri (2010)). Second, stronger sorting of *teams* into industries in developing countries could also explain the observed differences. For example, if in poor countries only the most skilled teams sort into high technology industries, while in rich countries both high and low skilled teams do so. A direct implication of this hypothesis, however, would be that within industries there should be very homogeneous teams, and thus little dispersion of used technology, in poor countries, and much greater dispersion in rich ones. This is at odds with empirical evidence that documents, even in narrowly defined industries, larger dispersion of productivity in poor countries.[51]

Next, I explore robustness to alternative sample selections or measures of concentration of talent. All results are reported in Table 1, and I refer below to its rows. For brevity, I focus on the cross-sectional comparison.

Korea are computed using the job history in one cross-section. As a result, the average age of individuals in the cross-section increases over time. Hence, these results may be capturing both a life-cycle and time-series component. This is a particularly relevant concern for self-employment, since it is known that entry into self-employment is correlated with age. (See for example Roys and Seshadri (2013)).

[50] Classic measurement error does not bias the results, as long as it averages to zero at the sectoral level.

[51] E.g., Hsieh and Klenow (2009) and Asker et al. (2014). Other reasons could generate cross-country differences in within industries dispersion other than teams compositions. Nonetheless, it is reassuring that failure of the second assumption would lead - through the lens of the model - to counterfactual empirical implications.

Table 1: Robustness Table

		Point Estimate	R^2
(1)	**Baseline**	-0.0708***	0.363
	Level of Industry Aggregation		
(2)	Sectors (Agr, Mfg, Ser)	-0.0861***	0.440
(3)	Unharmonized Industries	-0.0472***	0.173
	Sample Selection		
(4)	Include Non Household Heads	-0.0681***	0.355
(5)	Include Women	-0.0719***	0.379
(6)	Only Women	-0.0610***	0.150
(7)	Include Self-Employed	-0.0715***	0.364
	Role of Agriculture		
(8)	Drop Agriculture	-0.171***	0.553
(9)	Only Individuals non in Agriculture	-0.0266**	0.087
	Measure of Concentration of Talent		
(10)	Correlation	-0.0407***	0.249
(11)	Correlation using Normalized Skills	-0.0246**	0.103

Notes: *** means p-value<0.01; ** means p-value<0.05.

First, I explore alternative definition of industries. I aggregate industries at the sector level (agriculture, manufacturing, services) or I use, when available, finer definition of industries. This second alternative comes at the cost of lack of cross-country comparability, since for different countries I have different data at a different levels of aggregation. The results are robust to either industry definition (rows 2 and 3).[52] Second, I explore alternative sample selections. Results are robust to the inclusion of males non household head (row 4), or females (row 5). I then restrict the sample to *only* females (row 6). The fit is weaker, but the coefficient is still very similar. Last, I include individuals who report to be self-employed, and the result does not change (row 7). Third, given the large cross-country differences in the share of employment in agriculture, it may be useful to investigate whether the results are mostly driven by the prevalence of agriculture in developing countries.[53] I address this point through two exercises. I start by recomputing the measure of concentration of talent dropping agricultural industries in the previously described cross-industry regression used to compute $\hat{\pi}$, i.e. $\bar{\bar{x}}_j = B_0 + B_1\bar{\bar{p}}_j + \varepsilon$. Row 8 shows that the larger measure in poor countries does not come purely from the gap between agriculture and non-agriculture, but rather holds also within other sectors. I then

[52]Examples of Brazil 2010 and United States 1940 are in Figures A.2 and A.3 in the Section A.1. We can appreciate that the nice fit of the measure of concentration of talent is present at any level of aggregation.

[53]The comparison between currently poor countries and the U.S. in the past already hinted towards the fact that the result cannot uniquely be driven by differences in agricultural share, since - as shown in Herrendorf et al. (2014) - most countries follow the same structural transformation pattern as the one of the U.S. in the past.

consider only individuals who are not in agriculture, and recompute both the normalized measure of skill and the concentration of talent. This exercise calculates cross-country differences in concentration of talent if suddenly all individuals in agriculture dropped out of the labor force. Results (row 9) are weaker and smaller in magnitude but still show more concentration in poor countries. Fourth and last, I compute an alternative measure of concentration of talent, namely the correlation between individuals education and the average education in an industry , that is

$$\hat{\pi}_2 = \mathrm{Corr}\left(s_{ij}, \mathrm{E}\left(s_i | I_{ij} = 1\right)\right).$$

Under this alternative measure, which is equivalent to a variance decomposition exercise, poor countries have strong concentration of talent. Results hold both if computed with raw education (row 10) or with normalized skills $\hat{x}$ (row 11).

3.3 Evidence on Allocation of Talent from Firm Level Data

Last, I provide further evidence using firm level data, that allow to directly observe who works with whom, and thus to get closer to the model's notion of a team. In the model, a team is characterized by two key features: (i) it is made of individuals that interact with each other, to the extent that there is complementarity in production; and (ii) the members of a team use identical technology. A firm is the empirical object that better represents these two features. Individuals within a firm mostly share the same production technology, task-specialize based on their ability, and interact with each other in a manner that may generate complementarities.[54]

3.3.1 Data (Firm Level)

I use firm level data from the World Bank Enterprise Survey (WBES). Since the WBES questionnaire and methodology changed in 2006, there are two datasets available, both of which are used since they contain different variables. The first one, henceforth WBES 2006, covers the period 2002 to 2006. The second one, henceforth WBES 2014, covers the period 2007 to 2014.

As a measure of technology, in the WBES 2014 I use labor productivity, defined as sales

[54]Firms are usually made up of more than two individuals. However, we can interpret a team as made up of two groups of individuals. Under this interpretation, the only real disconnect between a team in the model and a firm in the data is the fact that I assume teams to have a fixed number of managers and workers. This should not be an empirically important departure, since - as already discussed - the main theoretical argument of the theory can be extended to a case in which the number of hired workers is a choice variable. Additionally, in the robustness section, I show that size differences do not seem to be driving the results.

divided by total number of employed workers; in the WBES 2006, I can use instead three different measures: (i) labor productivity; (ii) fraction of labor force that uses computers; (iii) self-assessed level of technology relative to main competitors.

As a measure of education, hence skill, in the WBES 2006, I use the answer to the question *"What percent of the workforce at your establishment have the following education levels?"*, which allows me to compute the whole distribution of education within a firm. In the WBES 2014, I use instead the answer to the question *"What is the average education of your production workers?"*. In the WBES 2006, I also observe the education of the top manager in the firm.

The sampling methodology targets formal (registered) companies with 5 or more employees. These data therefore are representative only of a selected fraction of the population, which is smaller for poor countries, where a large part of the population is employed in the informal sector. For this reason we should be cautious in interpreting the results. In Section 3.3.3, I farther discuss this concern. Additional details on the data are included in Section A.11.

3.3.2 Empirical Results

I use the data to provide support to three empirical predictions of the model. Results are summarized in Figures A.10, A.11, and A.12.

First. One fundamental assumption in the model is that there is complementarity between skills and technology and between skilled managers and skilled workers.

Empirical Prediction 3 **(Complementarity Skills-Technology).** *In both rich and poor countries, higher educated workers work in firms managed by higher educated managers and which use a more advanced technology.*

I use data from WBES 2006 and run, for each country across firms, a regression of each of my measures of technology on the average education of firm workforce. The results show that, similarly in rich and poor countries, high skilled individuals are more likely to work in high technology firms. I run the same regression for top manager education and show that high skilled individuals are also more likely to work in a firm whose top manager is highly educated.

Second. As a result of a less convex cost of technology, the model predicts a larger dispersion of used technology in countries farther from the frontier, both directly and indirectly through changes in the optimal allocation. In the model, far from frontier, some firms use a modern technology, while others find a backward one more convenient. High skilled are concentrated in the high technology firms, thus further increasing the gap in optimal technology. Close to the frontier, instead, most firms use the modern technology, and the gap in used technology should be smaller.

Empirical Prediction 4 (Dispersion of Technology). *Dispersion of used technology across firms is larger in poor countries.*

I show that the dispersion of technology is indeed larger in poor countries. I first use log labor productivity, compute the cross-sectional dispersion for each country, and show that it is significantly negatively correlated with GDP per capita. I then show that the same significant negative relationship with GDP per capita holds using more direct measures of technology, namely the fraction of workers that use a computer and the level of perceived technology.

Third. The model predicts that the farther a country is from the frontier, the more talented individuals are concentrated in firms (teams) that use relatively high technology. Thus the relative dispersion of skills *within* firms should decrease and the dispersion of skills *across* firms should increase.

Empirical Prediction 5 (Dispersion of Skills). *The ratio between the across firm variance and the overall variance of education is larger in poor countries.*

In the wave WBES 2006, I observe the distribution of individual education years for each firm. I use this to compute the overall variance of education in each country and decompose it between the within-firm and across-firm variance. The fraction of overall variance that is explained across firms decreases significantly in GDP per capita, that is, poor countries indeed have a larger concentration of talent across firms.[55]

In the WBES 2014 wave, I only observe for each firm its average education. Therefore, I cannot directly compute a variance decomposition exercise. Instead, I first show that cross-sectional dispersion in average firm education is significantly negatively correlated with GDP per capita, that is, in poor countries the across firms gap in average education is larger. Next, in order to account for the cross-country differences in distribution of education, I use the cross-sectional dispersion of education computed directly from the household data for the countries for which is available. Under the, arguably strong, assumption that the dispersion of education in the sample of individuals working in the

[55] I compute a similar variance decomposition using the normalized measure of education between 0 and 1, as in Section 3.2. Results are similar - although slightly less significant - and are available upon request.

WBES 2014 firms is the same as the one of the whole country, I can then compute the fraction of this dispersion that is explained across firms. When doing so, I show that the fraction of total dispersion of education explained by firms is also significantly negatively correlated with GDP per capita, as predicted by the model.

3.3.3 Robustness and Alternative Interpretations

The data cover only registered firms with more than five employees. They thus miss a large fraction of the population in poor countries, where average firm size is smaller and informality is widespread. We should thus interpret the evidence as suggestive. It can nonetheless be useful to discuss and address the most obvious concerns. Results for this robustness check are shown in Figure A.13.

The empirical results highlight that there is more dispersion of economic activity in poor countries. One concern would be if the set of firms included in the sample are, in poor countries, drawn more frequently from the tails of the overall distribution of firms. I show that in both rich and poor countries firms have similar size, with 100 employees on average. In developing countries, these firms are likely to be overly representative of the high productivity portion of the economy. Consistent with this observation, the workforce there employed is more educated than the country average, and especially so in the poorest countries. For this reason, we are likely observing a right truncation of the firm distribution in poor countries. This fact is suggestive that the results might even understate the amount of dispersion in poor countries, since we are not capturing the possibly large gap between firms included and those not included in the sample; that is between the formal and informal side of the economy.[56]

Another concern is that the larger dispersion across firms in poor countries may be driven by larger dispersion in size. However, the cross-sectional dispersion of firm size is, among the firms in the sample, smaller in poor countries, thus alleviating this concern.

Last, we might be concerned that the notion of technology in the data captures a different concept than the one in the model. The key assumption in the model is that technology can be improved subject to a cost. For example, this assumption might be violated if some firms have higher technology and labor productivity simply because their owners are more able. Firm owner ability is, in fact, an (approximately) fixed characteristics that arguably should respond little to the quality of the labor force. One question in the WBES 2006 helps to address this concern: "Over the last two years, what were the

[56]However, whether a right truncation would over or understate the observed dispersion in the data ultimately depends on the shape of the firm distribution. For this reason, this analysis should be interpreted as suggestive.

leading ways in which your establishment acquired technological innovations?". "Embodied in new machinery or equipment" is one of 13 possible choices. Across both rich and poor countries, approximately 60% of firms ranked this as their first choice.

4 Implications for Economic Growth: A Dynamic Environment

Until this point, the objective of the paper has been to propose a new theoretical mechanism highlighting differences in the organization of production across rich and poor countries. The empirical investigation of the previous section supports the main predictions of the theory, suggesting that it has empirical relevance. In this section, I take a first step towards understanding how (and whether) this mechanism plays a role in the process of economic growth. Towards this aim, I build a simple dynamic framework and show that endogenous talent allocation may lead to non-convergence across countries even in the absence of barriers to technology adoption. The conclusions are thus relevant for development policy.

Environment. I work with the simplest setting that allows to highlight the dynamic forces. Time is discrete. Each country is inhabited by four individuals that have linear utility over income, are infinitely lived and discount the future at rate β. They have types $x_1 < x_2 < x_3 < x_4$. In order to produce - just as in the static setting - individuals must form two teams, and each team picks a technology. In each country, there are two sectors, and one team must be assigned to one sector. All countries are identical, with the exception of an exogenous time at which they transition from stagnation to growth. For this reason, I omit a country subscript throughout. There are three main differences with respect to the static setting. *First*, the cost of technology is endogenous and depends on past investment in technology, that determines current level of technology over which individuals make marginal improvements. *Second*, the level of technology is sector specific, that implies that we must consider also the assignment of teams to sectors. *Third*, there is a world technology frontier that grows exogenously at rate λ, starting from an initial value $a_0 = 1$, such that the world frontier at time t will be $\bar{a}_t = \lambda \bar{a}_{t-1} = \lambda^t$. (I use $\bar{a}$ to distinguish the world technology frontier from other technologies).

Each country goes through two phases: stagnation and growth. The transition from stagnation to growth happens at a country specific and exogenous time t_0. During stagnation, only one technology is available, $a_0 = 1$, at zero cost, $c_t(a_0) = 0$. During growth, technology in each sector is chosen from the set of available technologies, that is the stock of all past technologies $\mathbb{A}_t = \{a_0, a_1, ..., \bar{a}_t\} = \{1, \lambda, ..., \lambda^t\}$. Cost of technology a at time t depends on frontier technology, $\bar{a}_t$, and on last period technology in the sector in which the team operates, a_{jt-1}.

Assumption 5 (Growth Model).

$$\begin{aligned} f(x, x') &= x(1 + \gamma x') \\ c(x, x', a; a_{jt-1}, \bar{a}_t) &= \left(\frac{a}{\gamma^{\frac{\log \bar{a}_t}{\log \lambda} - \frac{\log a}{\log \lambda}}} + a\kappa\gamma^{\frac{\log \bar{a}_t}{\log \lambda} - \frac{\log a_{jt-1}}{\log \lambda}} \right)(1 + \gamma x') \\ a &\in \mathbb{A}_t = \{a_0, a_1, ..., \bar{a}_t\}. \end{aligned}$$

A few comments are in order. First, these functional forms satisfy the assumptions of the static framework. Second, the complicated functional form for the cost of technology has both useful properties and an intuitive microfoundation. I leave their detailed description to Section A.14. For the purpose of this section, it is sufficient to highlight that for a given level of world frontier technology $\bar{a}_t$, if a given sector's technology, a_{jt-1}, is higher then the cost function will be lower for each technology and more convex. Third, the time of takeoff, t_0, determines the initial distance to the frontier in each sector: a country that starts to grow late faces an advanced frontier technology but has the same base level of technology a_0.

Planner's Problem. I consider the problem of a planner that takes as given the time of takeoff, t_0, and the evolution of the technology frontier and chooses a sequence of teams and technology in each sector to maximize the presented discount value of the sum of future income flows.[57,58] I show in A.13 that this problem admits a recursive formulation with states given by the distance to the frontier in the two sectors, defined as $d_1 \equiv \frac{\log \bar{a} - \log a_1}{\log \lambda}$ and $d_2 \equiv \frac{\log \bar{a} - \log a_1}{\log \lambda}$. Without loss of generality, I let $d_1 \leq d_2$.

Lemma 9 (Recursive Formulation). *If $\beta < \hat{\beta} \leq \frac{1}{\lambda}$ and $t > \hat{t}$, the planner's optimal allocation solves*

$$v(d_1, d_2) = \max_{d_1', d_2' \geq 0;\, y, z \in \{x_2, x_3\},\, y \neq z} g\left(x_4, y, d_1'; d_1\right) + g\left(z, x_1, d_2'; d_2\right) + \tilde{\beta} v\left(d_1', d_{2,}'\right)$$

[57] I am assuming that a country does not have a saving technology, but must consume all his production.

[58] The problem requires the planner not only to form two teams, but also to assign each team to a sector. The two sectors might in fact be ex-ante different due to different levels of technology.

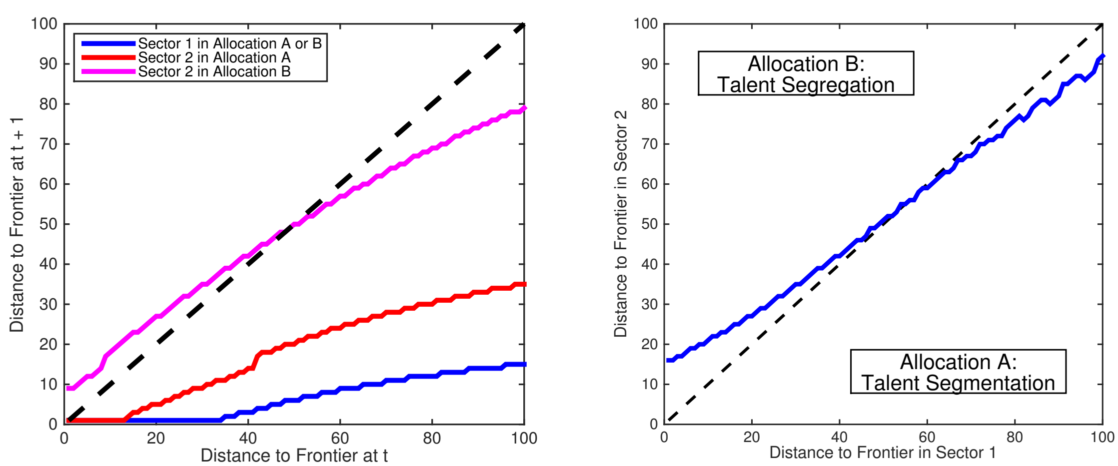
where $g\left(x', x, d^{'}; d\right)$ is the rescaled output flows net of technology cost for a team (x', x) that is in a sector at distance d from the frontier and picks technology at distance d'; $\tilde{\beta} = \beta\lambda$.

This formulation of the problem allows the planner to choose between two allocations, one with talent segmentation by occupation, which I call allocation A and is $\{(x_4, x_2), (x_3, x_1)\}$, and one with talent segregation by technology, allocation B, $\{(x_4, x_3), (x_2, x_1)\}$. Additionally, the most skilled team is assigned to the sector with higher technology. A planner that maximizes static flows would not choose any other allocation. A sufficiently low discount rate β thus guarantees that we are not restricting the choice of the planner.[59] Also, distance to the frontier must be, by definition, smaller than t. $t > \hat{t}$ guarantees that this constraint does not bind.

The *Dual Economy Trap.* I solve the model numerically, and compute the growth paths for different values of time of takeoff t_0. The interaction of allocation of talent and technology choice may generate path dependence: countries that take off late (high t_0) converge to a different balanced growth path, with large technological heterogeneity and depressed output, a situation that resembles a *dual economy trap.* I next illustrate the model solution for parameter values that lead to this sort of path dependence and highlight the main mechanism at play.

Figure 10: Solution of the Dynamic Model

(a) Optimal Technology (b) Optimal Allocation



In Figure 10a I plot, for both sectors and in either allocation, the optimal technology choice as a function of current distance to the frontier. The optimal technology depends

[59]Otherwise, it is possible to construct cases in which the planner might want, for example, to allocate the manager x_4 to the backward sector. This would lead to static losses, but the gains in terms of future technology possibility may outweigh them.

- due to the functional forms - only on the type of the manager, thus it varies across allocations only in sector 2. If a country is far from the frontier, the gap in optimal technology across sectors is larger, due to the lower convexity of the cost function. As a result, due to the stronger complementarity induced by technology choice - just as in the static theory - in countries far from the frontier is optimal to have talent segregation. This can be seen in Figure 10b, where I show in the space of distances to the frontier in sector 1 and 2 the regions in which it is optimal to have either allocation. A country at takeoff lies on the 45°. If takeoff is sufficiently late, thus far from the frontier, talent segregation is, at least initially, optimal. Additionally, back to Figure 10a we can notice that, when talent is segmented, both sectors converge to use the frontier technology ($d = 0$).[60] When talent is segregated, instead, sector 2 converges to a level of technology far from the frontier (in this numerical example, approximately 50 steps away from it).

After take-off, there are two additional forces to take into account. First, sector 1 converges faster towards the frontier due to the higher skill of the individuals allocated to it. This leads to a technology gap across sectors, that increases the return from allocating all the skilled individuals to sector 1 (talent segregation). Second, as a country approaches the frontier, the cost of technology becomes more convex. This makes instead talent segmentation more valuable. The growth path of a country is determined by the trade-off between these two forces.

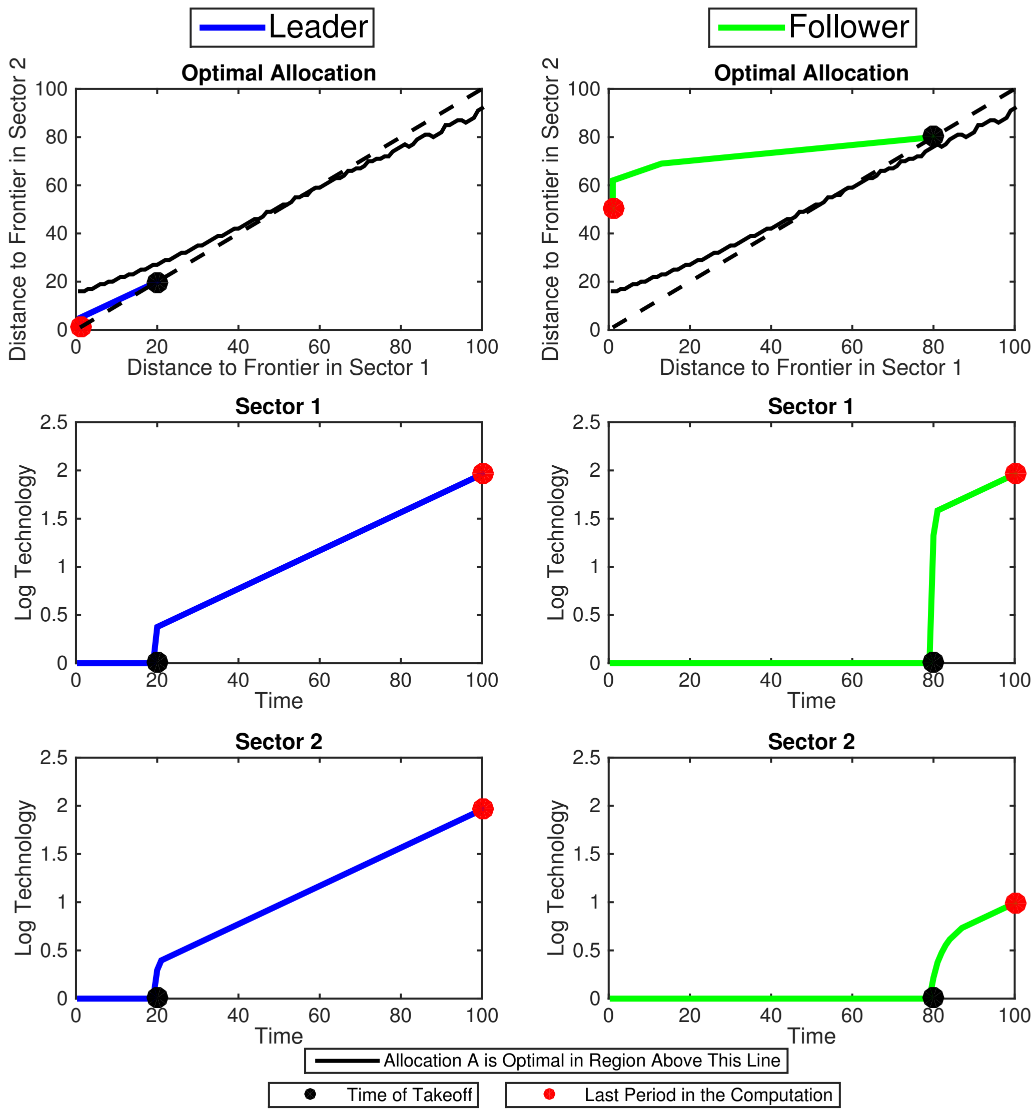
Consider now two countries, one that starts to grow early (leader), and one that starts to grow late (follower). Their paths are shown in Figure 11. The leader starts to grow close to the frontier, thus with talent segmentation. Some skilled individuals are allocated to each sector and, as a result, the technology gap across sectors remains low and both sectors converge to the frontier. The follower converges instead to a different balanced growth path. It starts to grow late, far from the frontier. This leads to segregation of talent. All skilled individuals are allocated to sector 1, that quickly catches up to the frontier. Even when sector 1 is on the frontier, the accumulated technology gap across sectors is large enough that is not optimal to revert to talent segmentation. As a result, the country converges to a dual economy trap, with depressed output and large cross-sectional technology dispersion. The initial time of take-off has persistent effects: there are two possible balanced growth paths, and a country reaches the one with large output and homogeneous technology only if it starts to grow early enough.

The two countries converge to different balanced growth paths despite the absence

[60]The convergence level of technology is given by the point in which the technology policy intersects with the 45 degrees line.

of any *exogenous* barriers to cross-country technology flows. The follower country builds *endogenous* barriers to adoption generated by the large technological gap between sectors, that prevents reallocation of talent where it is most needed to spur growth. The instantaneous loss from reallocating skilled individuals to the backward sector (sector 2) dominates in fact the future gains from reaching the balanced growth path with the higher output.

Figure 11: Growth Path for a Leader and a Follower



Lemma 10 **(Out of a Dual Economy Trap).** *For any time of take-off t_0, there exist $\delta(t_0)$ such that if the discount rate $\beta > \delta(t_0)$, the country does not fall into a dual economy trap, but rather converges to the same balanced growth as the one of a country with $t_0 = 0$.*

The dual economy trap is generated by a misalignment between short run and long run incentives, hence it is sustainable only as long as the discount rate is sufficiently low. A planner that gives enough weight to the long-run levels of output (high β) would like to homogeneously allocate talent to both sectors in such a way as to converge to the frontier. A short sighted planner in a country far from the frontier prefers instead to allocate all the most skilled individuals to one sector, and reap the benefits from quick convergence in that sector, even if this comes at the cost of future lack of overall (in both sectors) convergence.

Discussion. The dynamic framework highlights one new feature of the interaction between endogenous allocation of talent and technology choice. That is, the same possibility that has been emphasized in the literature as an *Advantage of Backwardness*[61], namely the fact that followers can directly adopt the frontier technology, may have unintended consequences that lead in fact to a *Disadvantage of Backwardness*. Countries, otherwise identical, that start to grow late (followers) fail to converge to the frontier despite free flows of technology. As emphasized, this is not due to inefficiency, but rather to misalignment of short and long run incentives. This prediction is consistent with the general lack of cross-country convergence. Growth miracles can also be rationalized within the model, and should be interpreted as the arrival of a more farsighted planner, that reallocates talent where most needed to spur growth.

It is important however to stress that this framework is purely illustrative of the proposed mechanism. Many features have been left out of the model, and this precludes from drawing sharp conclusions. For example, if skill could be endogenously accumulated at a sufficient low cost, countries may be able to grow out of the dual economy trap by investing in the human capital of low skilled individuals. Additionally, I have assumed that, within a country, technology is completely rival across sectors - e.g. embodied in capital. Relaxing this assumption, might also help a country to grow out of a dual economy trap. Last, the reasons that led some countries to start to grow late have been omitted in the discussion.

5 Conclusion

In this paper, I argue that the allocation of talent within countries is relevant for our understanding of economic development. More specifically, I highlight the role that tech-

[61] This term has been coined by Charles P. Kindleberger in a book review on the *Journal of Economic History* and refers to the work of Gerschenkron et al. (1962).

nology adoption plays in the internal organization of production and the growth prospects of developing countries today. To do so, I develop a new model of team formation in the presence of a localized technology choice. The model shows that the allocation of talent depends on the technological environment of a country, and in particular on how expensive large technological improvements are. The cheaper they are, the stronger the resulting production complementarity and thus the more assortative the equilibrium assignment, that is, the lower the skill gap between individuals working together. The theory thus predicts that poor countries - where individuals can increase their technology by a lot through adoption of the frontier one - have a different structure of the economy than rich ones, with larger dispersion of economic activity, and high skilled individuals clustered together. New evidence from micro data supports this prediction.

The way in which skilled individuals are allocated may matter for economic development and growth. An economy in which high and low skilled individuals interact will function differently from one in which individuals are segregated to work with others of similar skills. The paper explores one particular mechanism through which differences in the allocation of talent change the path of development. It shows that latecomers might adopt a growth strategy that allows them to reap the benefits from quickly catching up to the frontier in one sector, but this comes at the cost of falling towards a *dual economy trap* in which the backward sectors in the country remain stagnant.

Exploring additional mechanisms remains a promising area for future research. For example, economies in which individuals of different skills interact more frequently may be more (or less) efficient at diffusing ideas and good practices. A recent literature (e.g. Lucas and Moll (2014), Perla and Tonetti (2014), and Buera and Oberfield (2014)) develops theoretical models in which growth is generated through the diffusion of ideas. If embedded within these models, the different patterns of assignment in developing countries would affect the overall speed of learning and possibly generate a trade-off between the allocation of talent that maximizes output and the one that maximizes technology diffusion.[62]

[62]Similarly, Luttmer (2014) builds an assignment model in which student match with teachers for both production and learning. Evidence of slow accumulation of human capital over the life-cycle, that might be driven by slow learning due to lack of interaction of individuals of different skills, is documented in Lagakos et al. (2015).

References

Acemoglu, Daron, "Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence," *American Economic Review*, 1999, pp. 1259–1278.

_ , "Localised and Biased Technologies: Atkinson and Stiglitz's New View, Induced Innovations, and Directed Technological Change," *The Economic Journal*, 2015, *125* (583), 443–463.

_ **and Fabrizio Zilibotti**, "Productivity Differences," *Quarterly Journal of Economics*, May 2001, *116* (2), 563–606.

_ **, Philippe Aghion, and Fabrizio Zilibotti**, "Distance to frontier, selection, and economic growth," *Journal of the European Economic association*, 2006, *4* (1), 37–74.

Adamopoulos, Tasso and Diego Restuccia, "The size distribution of farms and international productivity differences," *The American Economic Review*, 2014, *104* (6), 1667–1697.

Ardagna, Silvia and Annamaria Lusardi, "Explaining international differences in entrepreneurship: The role of individual characteristics and regulatory constraints," Technical Report, National Bureau of Economic Research 2008.

Asker, John, Allan Collard-Wexler, and Jan De Loecker, "Dynamic inputs and resource (mis) allocation," *Journal of Political Economy*, 2014, *122* (5), 1013–1063.

Athey, Susan, Paul Milgrom, and John Roberts, "Robust comparative statics," *Manuscript, Department of Economics, Harvard University http://kuznets. fas. harvard. edu/˜ athey/draftmonograph98. pdf*, 1998.

Atkinson, Anthony B and Joseph E Stiglitz, "A new view of technological change," *The Economic Journal*, 1969, pp. 573–578.

Banerjee, Abhijit and Esther Duflo, *Poor economics: A radical rethinking of the way to fight global poverty*, PublicAffairs, 2011.

Basu, Susanto and David N Weil, "Appropriate Technology and Growth," *Quarterly Journal of Economics*, 1998, pp. 1025–1054.

Becker, Gary S, "A theory of marriage: Part I," *The Journal of Political Economy*, 1973, pp. 813–846.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 1967, *75*, 352.

Bloom, Nicholas and John Van Reenen, "Why do management practices differ across firms and countries?," *The Journal of Economic Perspectives*, 2010, pp. 203–224.

Bloom, Nick and John van Reenen, "Measuring and explaining management practices across firms and countries," *Quarterly Journal of Economics forthcoming*, 2007.

Buera, Francisco J and Ezra Oberfield, "The Global Diffusion of Ideas," in "2014 Meeting Papers" number 1099 Society for Economic Dynamics 2014.

— **, Joseph P Kaboski, and Richard Rogerson**, "Skill Biased Structural Change," Technical Report, National Bureau of Economic Research 2015.

Caselli, Francesco, "Technological revolutions," *American economic review*, 1999, pp. 78–102.

— , "Accounting for Cross-Country Income Differences," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 1, Elsevier, 2005, chapter 9, pp. 679–741.

Chiappori, Pierre-André, Alfred Galichon, and Bernard Salanié, "The roommate problem is more stable than you think," 2014.

— **, Robert J McCann, and Lars P Nesheim**, "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, 2010, *42* (2), 317–354.

Comin, Diego A and Martí Mestieri Ferrer, "If Technology has arrived everywhere, why has income diverged?," Technical Report, National Bureau of Economic Research 2013.

Eeckhout, Jan and Philipp Kircher, "Assortative matching with large firms: Span of control over more versus better workers," *Universitat Pompeu Fabra (Mimeo)*, 2012.

Feenstra, Robert C, Robert Inklaar, and Marcel P Timmer, "Penn world table version 8.0," *The Next Generation of the Penn World Table, available for download at www. ggdc. net/pwt*, 2013.

Foster, Andrew D and Mark R Rosenzweig, "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of political Economy*, 1995, pp. 1176–1209.

Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of political economy*, 2000, *108* (5), 874–904.

— **and Esteban Rossi-Hansberg**, "Inequality and the Organization of Knowledge," *American Economic Review*, 2004, pp. 197–202.

— **and** — , "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, 2006, *121* (4), 1383–1435.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "Human Capital and Regional Development*," *The Quarterly journal of economics*, 2013, *128* (1), 105–164.

Gerschenkron, Alexander et al., "Economic backwardness in historical perspective.," *Economic backwardness in historical perspective.*, 1962.

Goldin, Claudia and Lawrence F Katz, "The Origins of Technology-Skill Complementarity*," *The Quarterly journal of economics*, 1998, *113* (3), 693–732.

Gollin, Douglas, David Lagakos, and Michael E. Waugh, "The Agricultural Productivity Gap," *Quarterly Journal of Economics*, 2014, *129* (2), 939–993.

Grossman, Gene M and Giovanni Maggi, "Diversity and Trade," *American Economic Review*, 2000, pp. 1255–1275.

_ , Elhanan Helpman, and Philipp Kircher, "Matching and sorting in a global economy," Technical Report, National Bureau of Economic Research 2013.

Hanushek, Eric A and Ludger Woessmann, "The role of cognitive skills in economic development," *Journal of economic literature*, 2008, pp. 607–668.

Herrendorf, Berthold, Richard Rogerson, and Ákos Valentinyi, "Growth and Structural Transformation," *Handbook of Economic Growth*, 2014, *2*, 855–941.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, November 2009, *124* (4), 1403–1448.

_ , Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," NBER Working Papers 18693 January 2013.

Klenow, Pete and Andres Rodriguez-Clare, "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?," in "NBER Macroeconomics Annual 1997, Volume 12," National Bureau of Economic Research, 1997, pp. 73–114.

Kremer, Michael, "The O-ring theory of economic development," *The Quarterly Journal of Economics*, 1993, pp. 551–575.

_ and Eric Maskin, "Wage inequality and segregation by skill," Technical Report, National Bureau of Economic Research 1996.

Lagakos, David, "Explaining cross-country productivity differences in retail trade," *Available at SSRN 2214497*, 2013.

_ and Michael E Waugh, "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, 2013, *103* (2), 948–80.

_ , Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman, "Life-Cycle Wage Growth Across Countries," 2015.

Low, Corinne, "Pricing the Biological Clock: Reproductive Capital on the US Marriage Market," Technical Report, November 2013. Mimeo 2013.

Lucas, Robert E. and Benjamin Moll, "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 2014, *122* (1), 1 – 51.

Lucas, Robert E Jr., "On the Size Distribution of Business Firms," *Bell Journal*, 1978, *9*, 508–523.

Luttmer, Erzo GJ, "An Assignment Model of Knowledge Diffusion and Income Inequality," Technical Report 2014.

M., Andrei Shleifer Murphy Kevin and Robert W. Vishny, "The Allocation of Talent: Implications for Growth," *The Quarterly Journal of Economics*, May 1991, *106*, 503–530.

Maddison, Angus, *The world economy volume 1: A millennial perspective volume 2: Historical statistics*, Academic Foundation, 2007.

Mestieri, Martí, "Wealth distribution and human capital: How borrowing constraints shape educational systems," Technical Report, Working Paper 2010.

Minnesota Population Center, *Integrated Public Use Microdata Series, International: Version 6.1 [Machine-readable database]*, University of Minnesota, 2011.

Noldeke, Georg and Larry Samuelson, "Investment and Competitive Matching," 2014.

Parente, Stephen L. and Edward C. Prescott, "Barriers to Technology Adoption and Development," *Journal of Political Economy*, 1994, *102* (2), 298–321.

Perla, Jesse and Christopher Tonetti, "Equilibrium Imitation and Growth," *Journal of Political Economy*, 2014, *122* (1), 52 – 76.

Porta, Rafael La and Andrei Shleifer, "Informality and Development," *The Journal of Economic Perspectives*, 2014, pp. 109–126.

Rodrik, Dani, "Unconditional Convergence in Manufacturing*," *The Quarterly Journal of Economics*, 2012, p. qjs047.

Roys, Nicolas and Ananth Seshadri, "Economic development and the organization of production," Technical Report, mimeo 2013.

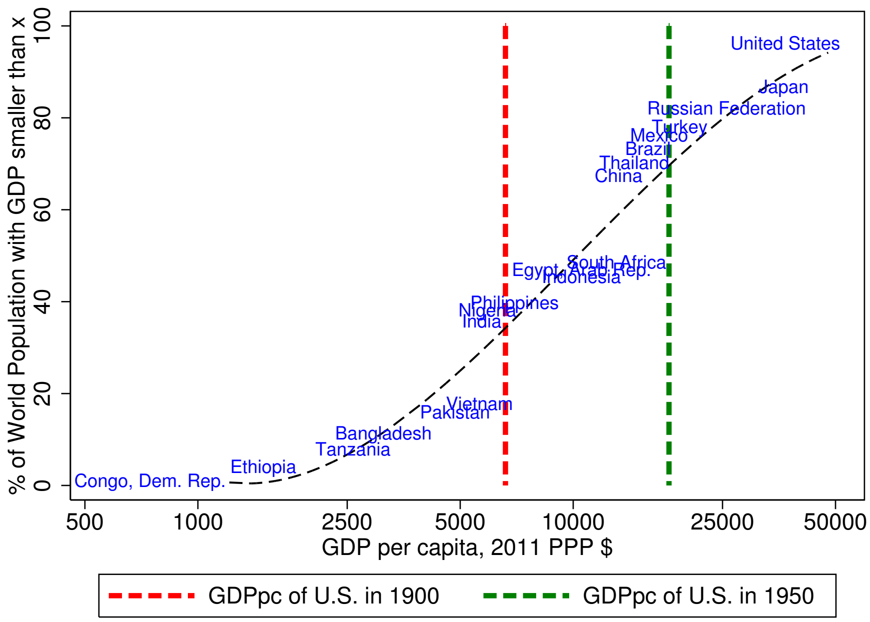
Schoar, Antoinette, "The divide between subsistence and transformational entrepreneurship," in "Innovation Policy and the Economy, Volume 10," University of Chicago Press, 2010, pp. 57–81.

Young, Alwyn, "Inequality, the Urban-Rural Gap and Migration*," *The Quarterly Journal of Economics*, 2013, p. qjt025.

A Appendix

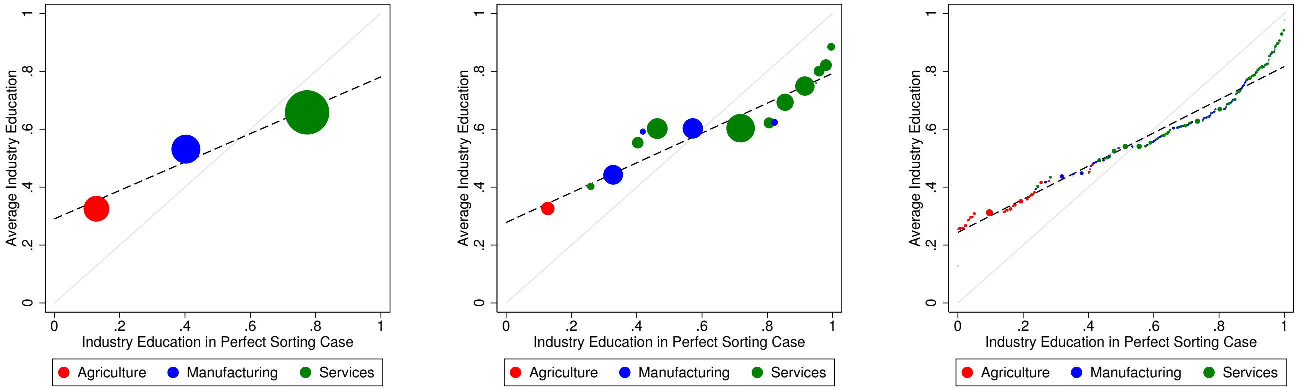
A.1 Additional Figures

Figure A.1: GDP per capita



Notes: GDP per capita and population is from PWT 8.0. Historical GDP per capita for the United States is from Maddison (2007)

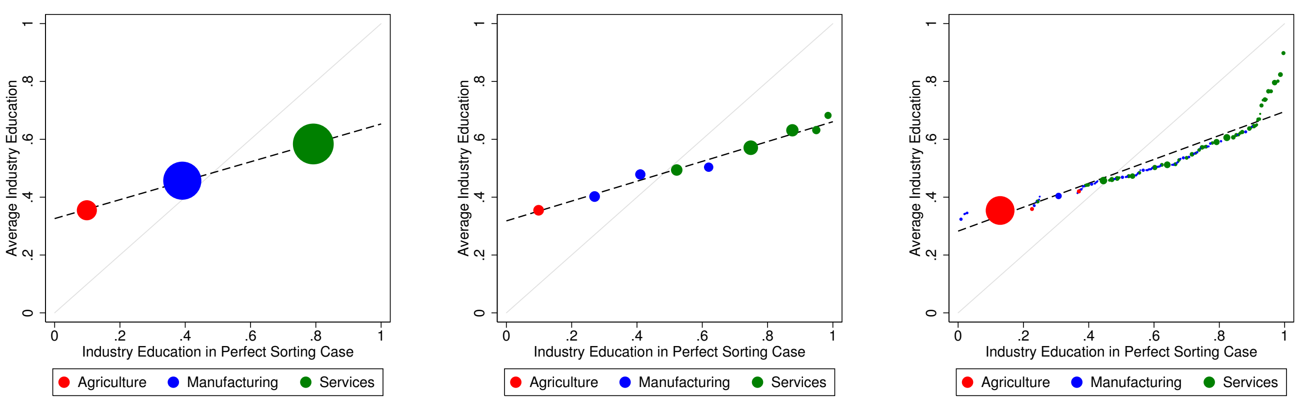
Figure A.2: Construction of Measure of Concentration of Talent, Brazil in 2010



(a) Across Sectors (b) Across Harmonized Industries (c) Across Unharmonized Industries

Notes: in each of three figures I follow the same procedure, with the difference that in the left one I refer to an industry as a sector, in the middle one as a 1-digit industry harmonized by IPUMS international, and in the right one as an unrecoded industry, which in the case of Brazil is at the 3-digits level. In each figure, I plot the average normalized education in an industry as a function of the average education in a counterfactual scenario in which there is perfect sorting of individuals across industries. Each dot correspond to an industry, as defined, and the size of the dot is increasing in the number of individuals employed in that industry. The dotted lines are the prediction from a linear regression weighted by the number of individuals in each industry. The slopes of the regression lines are the measures of the concentration of talent, which for Brazil in 2010 are, from left to right, 0.49, 0.51, 0.57.

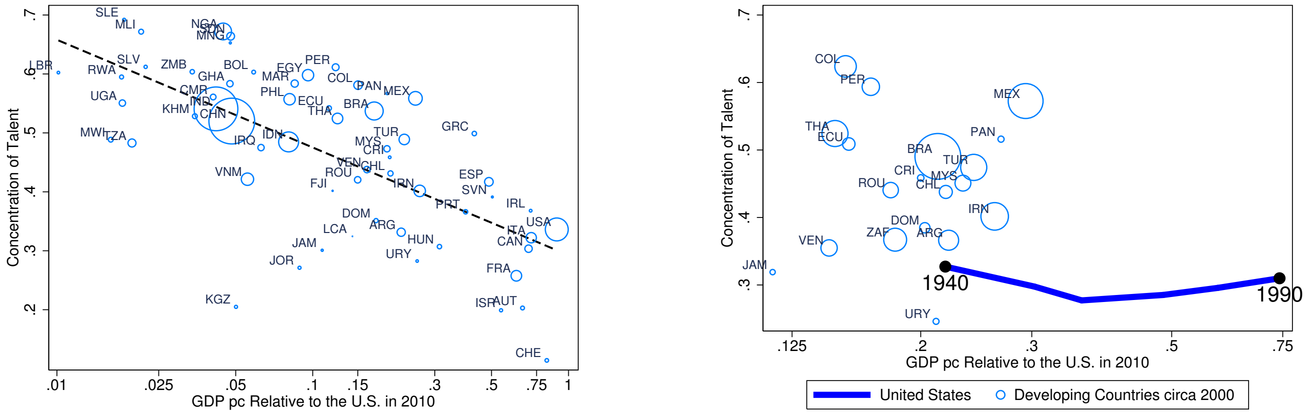
Figure A.3: Construction of Measure of Concentration of Talent, United States in 1940



(a) Across Sectors (b) Across Harmonized Industries (c) Across Unharmonized Industries

Notes: see Figure A.2. The slopes of the regression lines for United States in 1940 are, from left to right, 0.33; 0.34; 0.41.

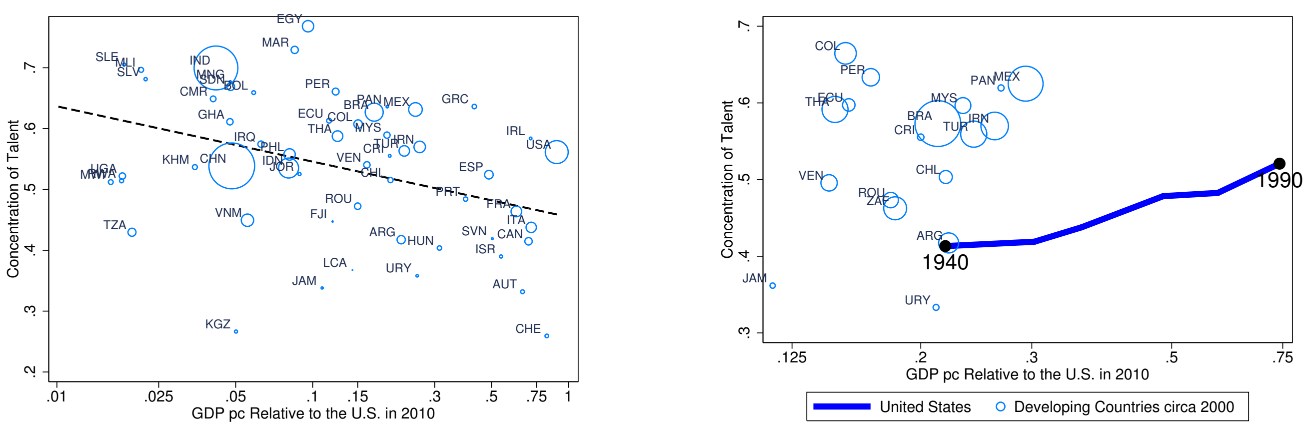
Figure A.4: Concentration of Talent Across Sectors



(a) Cross-sectional Differences (b) Comparison with U.S. in the past

Notes: see Figures 6 and 7. The difference with respect to those figures is that here I plot the concentration of talent across the three sectors - agriculture, manufacturing, services - rather than industries. Sectors are recoded from the harmonized variable industry.

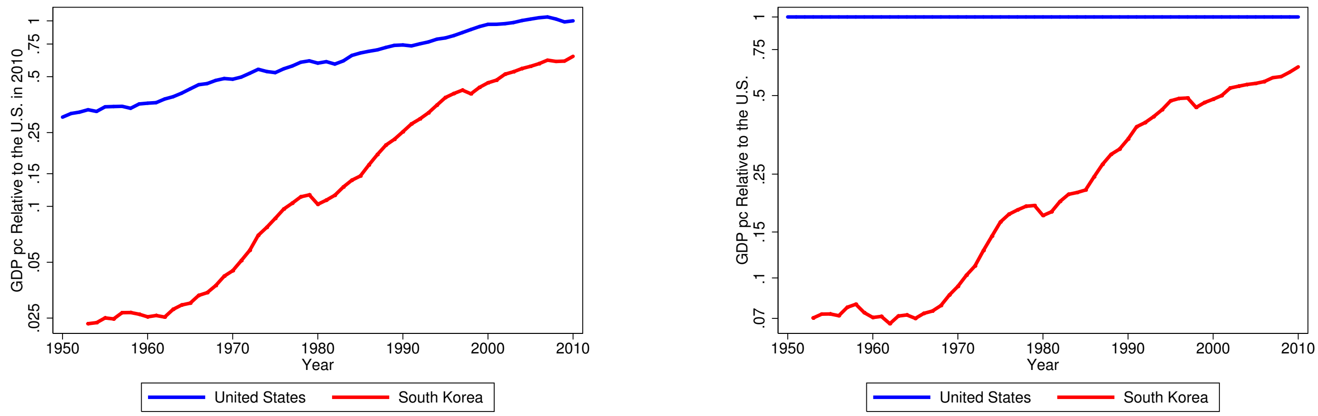
Figure A.5: Concentration of Talent Across Unrecoded Industries



(a) Cross-sectional Differences (b) Comparison with U.S. in the past

Notes: see Figures 6 and 7. The difference with respect to those figures is that here I plot the concentration of talent across unrecoded industries. Unrecoded industries are not harmonized nor across countries nor over time, but are more detailed than the harmonized one. In fact for most countries, the unrecoded industry variable - IND in the IPUMS dataset - provides information at the 3-digits level.

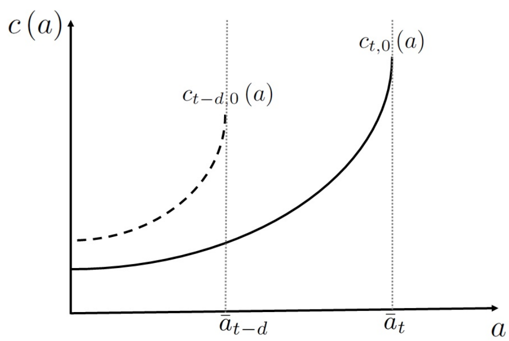
Figure A.6: Growth Paths of United States and South Korea



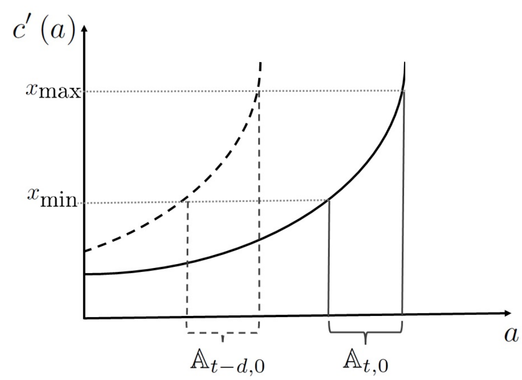
Notes: data are from the Penn World Table version 8.0. GDP per capita is computed as rgdpe/pop. Where rgdpe is expenditure side real GDP and pop is population size.

Figure A.7: Examples of Cost Function for a Country on the Frontier at Two Points in Time

(a) Cost of Technology

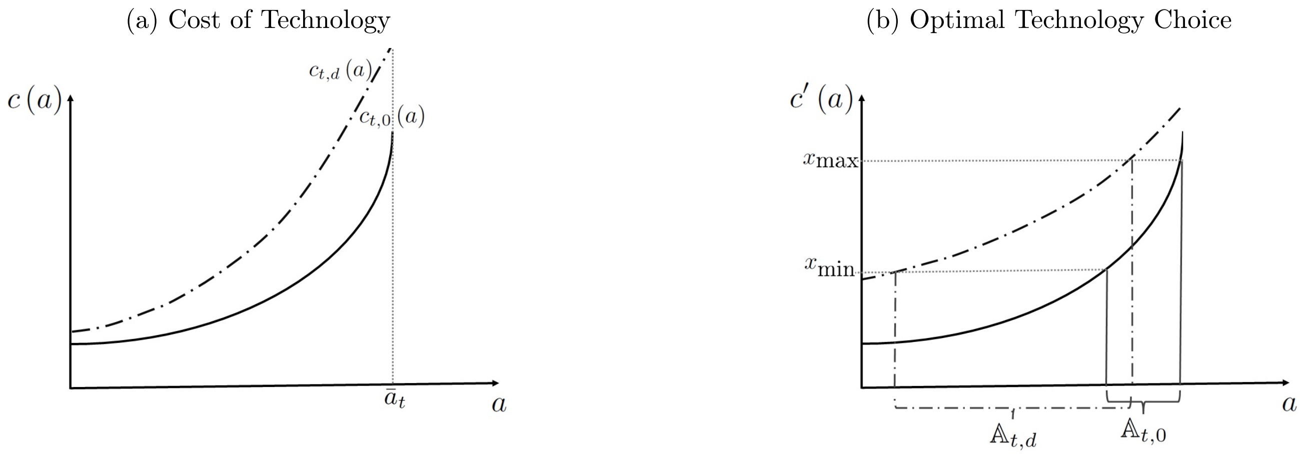


(b) Optimal Technology Choice



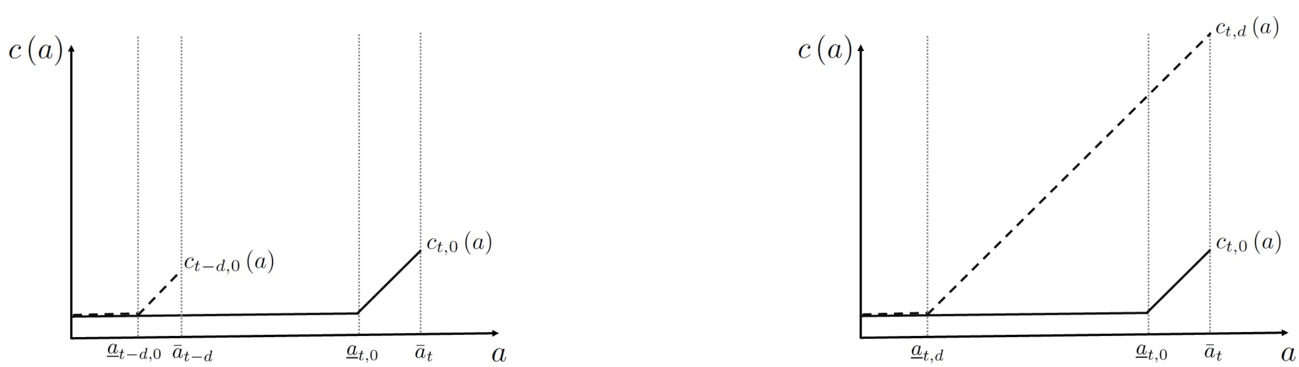
Notes: I plot in the left panel the cost functions for a country close to frontier at time $t-d$ and at time t. As the frontier grows, existing technologies become cheaper, and new ones are available. However, since the country is at a constant distance from the frontier, this does not change the curvature of the cost function. As a result, shown in the right panel where I plot the marginal cost of technology, the gap between the high and low technologies used is constant, and the set of used technologies improves over time. I compute the used technology set purely as an illustrative example. I am assuming that there are a high skilled and a low skilled individual, namely x_{max} and x_{min} which pick technology solving: $\max ax - c(a)$.

Figure A.8: Examples of Cost Function for a Country Close and One Far From Frontier



Notes: I the cost functions (left panel) and their marginals (right panel) for a country close to frontier at time t and one at d steps from it, still at time t. Both countries have access to the same set of technologies. However all technologies are more expensive for the country far from frontier. Additionally, the country far from the frontier has a less convex cost. The result is that far from the frontier the average level of technology is lower, but a large set of technologies are used, from backward to advanced.

Figure A.9: Examples of Cost of Technology Across Countries - Quasi Linear Case

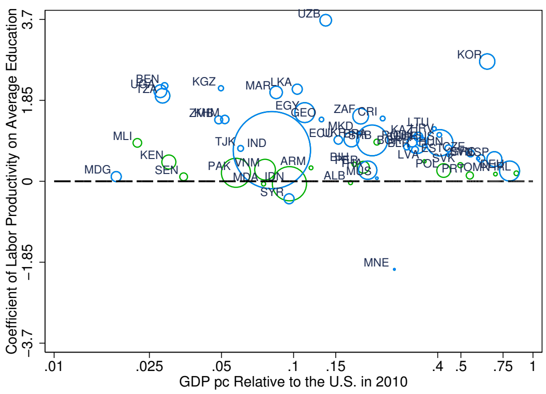


(a) Close to Frontier over Time (b) Close to Frontier and Far from it

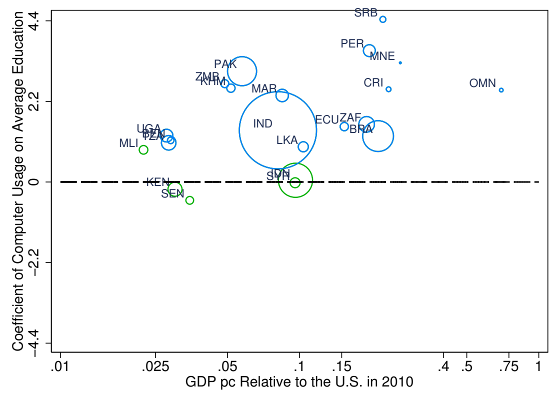
Notes: these two figures show the cost functions for a country close to the frontier at two different points in time and for one far from the frontier according to the cost function in the quasi-linear model.

Figure A.10: Complementarity Skill-Technology

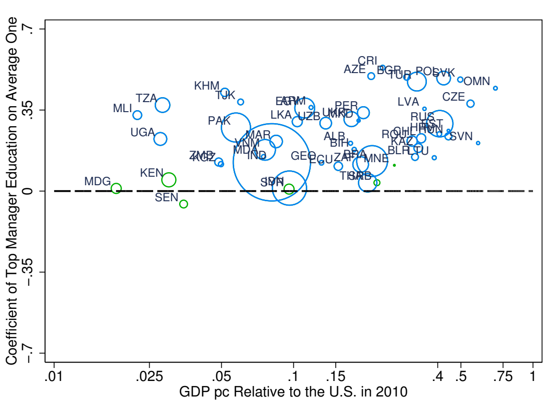
(a) Labor Productivity



(b) Computer Use



(c) Education of Top Manager

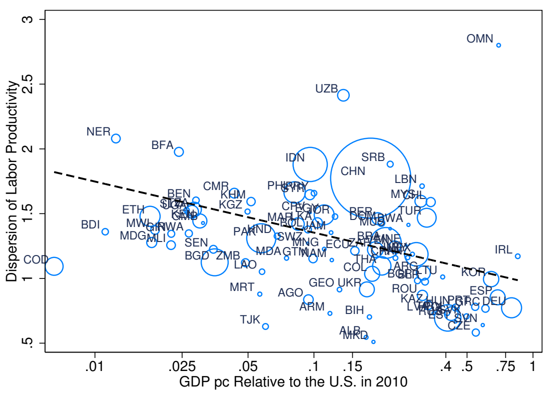


(d) Perceived Technology

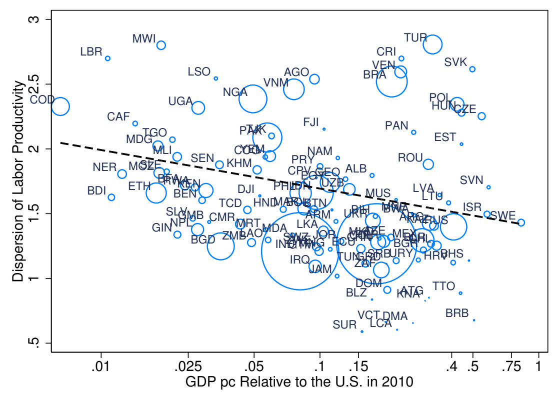
Notes: In each figure I plot the coefficients of a linear regression, computed across firms within a country, on the log of average education of the firm workforce. For example, let me describe the procedure to build the top left panel, since all other ones are identical. I first compute, for each country, a regression, across firms, of log labor productivity on the average education of the workforce of the firm. I store the coefficients of this regression and I plot them as a function of GDP per capita relative to the one of the U.S. in 2010. Each country is represented by its three digit country code and a circle, whose size depend on the population of the country, and whose color depend on whether the point estimates is significant (blue) or insignificant (green). The dotted line separates positive from negative coefficients. The data are from World Bank Enterprise Survey standardized waves 2002 to 2006.

Figure A.11: Dispersion of Used Technology Across Firms

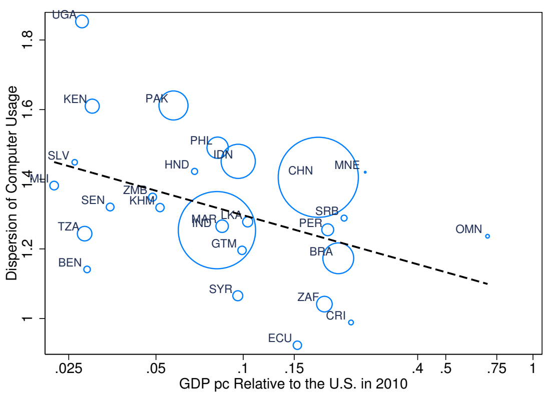
(a) Labor Productivity, WBES 2006



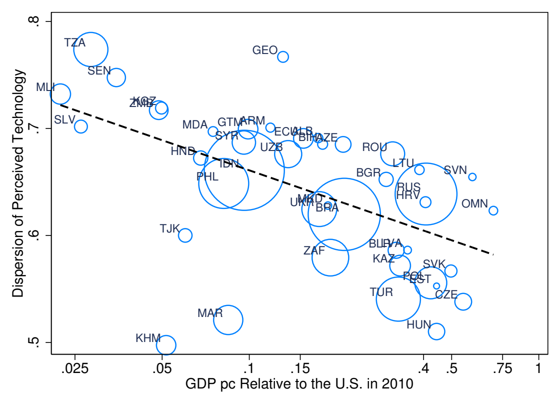
(b) Labor Productivity, WBES 2014



(c) Computer Use



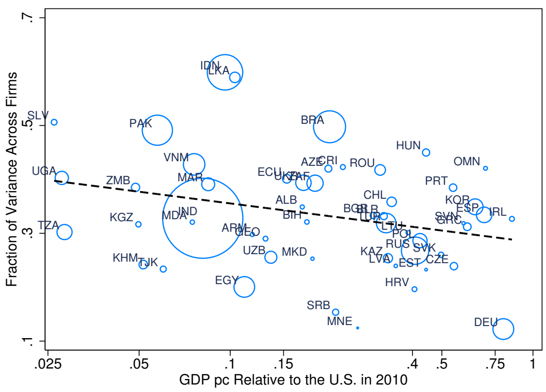
(d) Perceived Technology



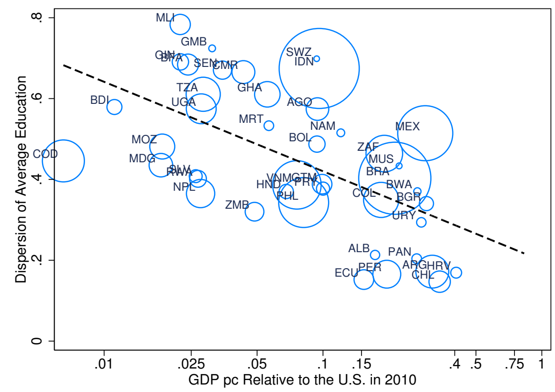
Notes: In each figure I plot standard deviation of log of a measure of the measure of education or technology shown in the figure title. Specifically, for each country I compute the cross-sectional, across firms, standard deviation of logs and then I plot the country estimates as a function of GDP per capita. Each country is represented by its three digit country code and a circle, whose size depend on the population of the country. The dotted black line is the regression line of a cross-country regression on log GDP per capita. All regressions are significant at 5% or below. Data are from World Bank Enterprise Survey standardized waves 2002 to 2006 (panels a and c and d) and 2007 to 2014 (panel b).

Figure A.12: Dispersion of Average Education Across Firms

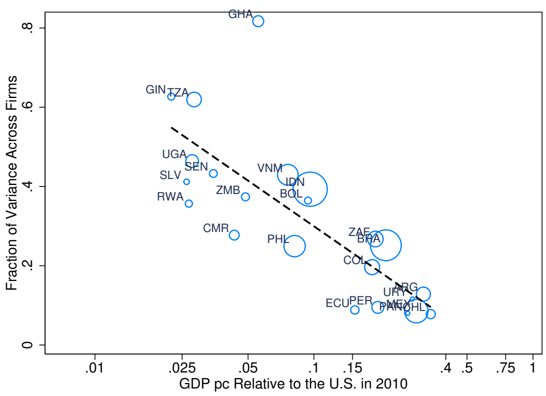
(a) Fraction of Total Variance Explained Across Firms, WBES 2006



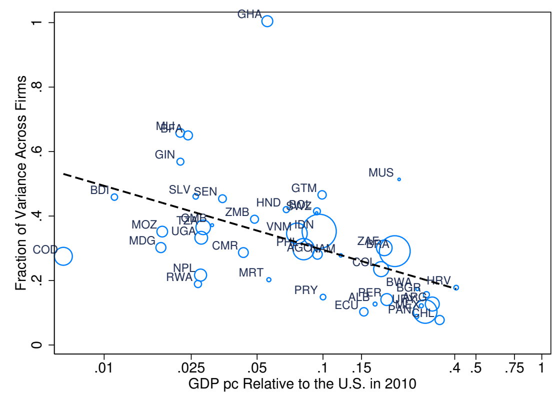
(b) Dispersion of Education, WBES 2014



(c) Fraction of Total Variance Explained Across Firms, WBES 2014 (a)

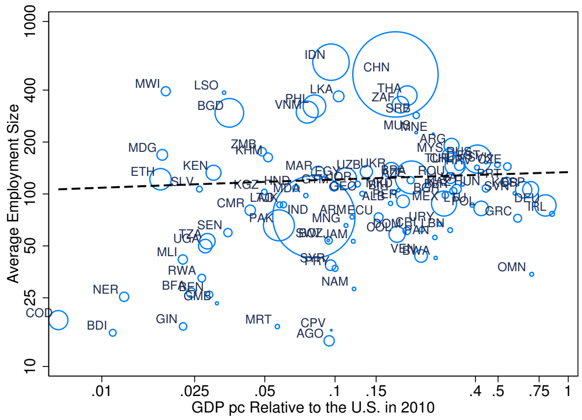


(d) Fraction of Total Variance Explained Across Firms, WBES 2014 (b)

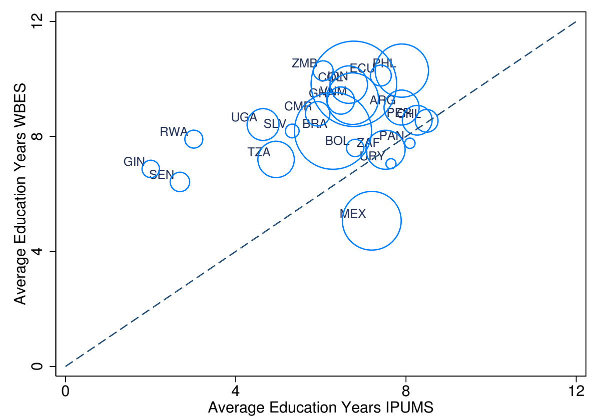


Notes: In the top left panel, I plot for each country as a function of GDP per capita the fraction of total variance of education which is "across firms". Specifically, for each country I compute the cross-sectional variance of education of the individuals hired at the firms in the sample. I can compute the country specific distribution of education using a variable that asks what is the fraction of the firm labor force with 6 years of education, between 6 and 9, between 9 and 12 and more than 12. I then decompose this total variance in the variance of education within firm and the variance of education across firms. For each country I compute the ratio between variance of education across firms and total variance and I plot it as a function of the country GDP per capita. In the top right panel, I compute for each country the cross-sectional variance of the log of average firm education and I plot is as a function of GDP per capita. In the bottom left panel, I plot the ratio between the variance of average education across firms and the overall variance of education, which is computed using micro data from IPUMS, the same used in 3.2. Notice that for some countries I do not have micro data, as such they do not appear in this figure. In order to overcome this limitation, in the right bottom panel I computed the predicted variance of education from a regression of variance of education on GDP per capita. In this way I can use all countries for which I have firm level data. Each country is represented by its three digit country code and a circle, whose size depend on the population of the country. The dotted black line is the regression line of a cross-country regression on log GDP per capita. All regressions are significant at 5% or below. Data are from World Bank Enterprise Survey standardized waves 2002 to 2006 (panels a and b) and 2007 to 2014 (panels c and d).

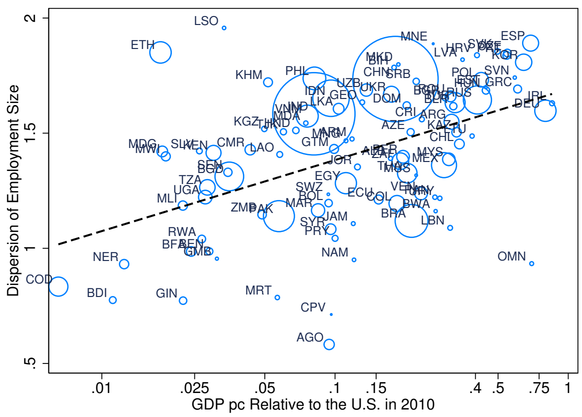
Figure A.13: Robustness for Firms Level Data



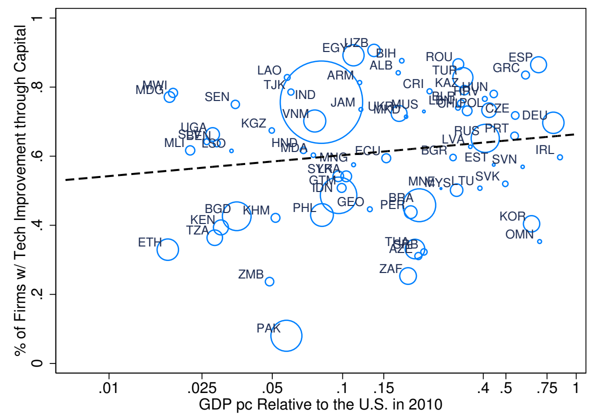
(a) Average Size



(b) Comparison Education in WBES and IPUMS



(c) Dispersion of Size



(d) Technology Embodied in Capital

Notes: In the top left panel I plot the average firm number of employees for each country as a function of the GDP per capita. In the the top right panel I plot the average education of workers in the firms in my data as a function of the average country education taken from IPUMS international micro data as described in Section 3.2. In the bottom left panel I plot the dispersion of log number of employees computed for each country as a function of GDP per capita. In the bottom right panel I plot for each country, still as a function of GDP per capita, the fraction of firms that to the question: "Over the last two years, what were the leading ways in which your establishment acquired technological innovations?: answer as leading way: "Embodied in new machinery or equipment". Each country is represented by its three digit country code and a circle, whose size depend on the population of the country. The dotted black line is the regression line of a cross-country regression on log GDP per capita. All regressions are significant at 5% or below. The data are from World Bank Enterprise Survey standardized waves 2002 to 2006.

A.2 Details on the Single-Crossing Assumption

I here discuss few properties of the single-crossing assumption *(1.5)*, that is $\min_{x'\in\mathbb{X}} f_1(x,x') > \max_{x'\in\mathbb{X}} f_2(x',x)$.

First, the single-crossing is tightly linked to the well-known Spence-Mirrlees condition in mechanism design. Consider a utility function $U = t - C(q,\vartheta)$, where t is a monetary transfer, q the action, and ϑ the unobserved type. The Spence-Mirrlees condition is $C_{q\vartheta} > 0$, and it allows to separate types. In my setting the action q is a binary choice, whether to be a manager or a worker (and the type ϑ is observed and given by the ability x). The Spence-Mirrlees thus reduces to a requirement that the marginal value of being manager (interpret q high) is higher than that of being worker (q low). This gives $f_1 > f_2$. However, in this setting we need to take into consideration that f_1 and f_2 depends themselves on the partner, through complementarity. Thus, in order for types to be separated into the two occupations, it must be that the marginal value of being a manager is higher than that of being a worker, regardless of the respective partners. This additional requirement gives the single-crossing (1.5).

Second, I show that the same single-crossing assumption holds, and plays a major role, in many settings in which heterogeneous workers must make an occupational choice. In order to see the role of this assumption, and the comparison with other settings, it is useful to briefly describe how it affects prices in a competitive equilibrium. Since other papers also use an intensive margin for number of workers, let the production function be $f(x',x,n(x))$ where $n(x)$ is the optimal number of workers given their ability x. The maximization problem of a manager x' that has to choose the type of his workers reads as

$$\pi(x') = \max_{x\in\mathbb{X}} f(x',x,n(x)) - w(x)\,n(x),$$

where $w(x)$ is the wage of a worker of type x. The first order and envelope conditions of this problem yields

$$\begin{aligned} w'(x) &= \frac{f_2(x',x,n(x)) + n'(x)\,f_3(x',x,n(x))}{n(x)} \\ \pi'(x) &= f_1(x',x,n(x)). \end{aligned}$$

It is then easy to see that, since individuals are not assigned ex-ante to be workers or managers, as long as $\pi'(x) > w'(x)\ \forall x$, high skilled are managers and low skilled workers. This is the essence of the single-crossing property, that is that the profit and wage schedule cross at most once. Next, I show that this property also holds in the cited settings. In Lucas (1978) the production function for a manager x' and n workers of type x is given by xn, where n does not depend on x. As a result $w'(x) = 0$, and single-crossing trivially holds. Next, consider the version of Garicano and Rossi-Hansberg (2006) with only two hierarchies, as in Garicano and Rossi-Hansberg (2004). The argument generalizes to the multi-hierarchy setting. The production function of a manager x' with $n(x) = \frac{1}{1-x}$ workers of type x is given by $x'n(x)$. Hence using the equations above, $\pi'(x') = n(x)$ and $w'(x) = xn(x)$. Since $x\in[0,1]$, single-crossing holds as well.

Third, in order to see how the single crossing assumptions relates to complementarity, we can rewrite the single-crossing, using the fact that $f_{12} > 0$ as: $f_1(x,x_{min}) > f_2(x_{max},x)$, where x_{min} and x_{max} are respectively the smallest and largest type in $\mathbb{X}$. Next, with a first order Taylor approximation around x on both sides we obtain

$$f_1(x,x) - f_{12}(x,x)(x - x_{min}) > f_2(x,x) + f_{12}(x,x)(x_{max} - x)$$

that rearranging becomes

$$f_1(x,x) - f_2(x,x) > f_{12}(x,x)(x_{max} - x_{min}). \tag{6}$$

This last inequality shows that in order for the single crossing to hold, the complementarity of the production function must be small with respect to the difference in skill-sensitivity of the complex and simple task.

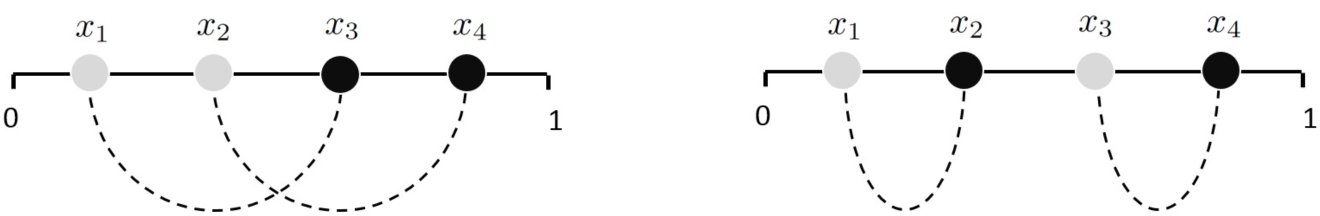
A.3 An Example in an Economy with Four Individuals

I here propose a stylized example that highlights the main mechanism at work. Consider four individuals of ability $x_1 > x_2 > x_3 > x_4$. Let them have to form two teams of a manager and a worker. I first assume that both teams must use the same production technology, which I call a, to produce their output. The net output of a team formed by a manager x' and a worker x is given by

$$a\left(\rho x' + x\right)$$

where $\rho > 1$ captures the fact that the skills of the manager are more important in determining the team's output. Under the optimal assignment of individuals to teams, the two most skilled individuals, x_3 and x_4, will be managers, and the less skilled, x_1 and x_2, will be line workers.[63] I refer to this assignment, illustrated in the left panel of Figure A.14, as *segmentation of talent* by occupation, to emphasize that high skilled individuals are in most task-intensive occupations. In this particular example, conditional on the vertical choice of occupation within a team, the horizontal assignment to teams is irrelevant due to the lack of complementarity in ability for this production function. Hence either manager can be paired with either worker. Intuitively, this is because the most skilled worker would be more productive, but the gain from this is the same regardless of the identity of the manager, since both managers use the same technology.

Figure A.14: Allocation of Talent for Four Individuals



(a) Segmentation by Occupation (b) Segregation by Technology

Notes: A grey dot represents a worker, while a black dot represents a manager. Dotted lines connect workers to their managers.

Next, I allow each team to pick its preferred technology. A team can use a simple technology, $\underline{a}$, at a cost of $\zeta\underline{a}$, or and advanced technology, $\bar{a}$, at a cost of $\zeta\bar{a}$. further, let $\zeta \in (\rho x_4 + x_1, \rho x_3 + x_2)$, so that regardless of whom he is paired with, only the most talented individual x_4 wishes to use the advanced technology. As a result, x_4 is the manager of the only team that uses $\bar{a}$. Next, consider the choice of x_3. He can be either the manager of a team that works with $\underline{a}$, or a worker for x_4, working with $\bar{a}$. His optimal role depends on the economic conditions, that in this example are summarized by the difference between the two available technologies. If $\frac{\bar{a}}{\underline{a}} > \rho$, the loss from using his talent towards a simple task is more than compensated for by the gain from using his skills to produce with the advanced technology.[64] If the gap between the available technologies is large enough, the assignment concentrates high skilled individuals together, in order for them to produce with the best available inputs. I refer to this assignment, illustrated in the right panel of Figure A.14, as *segregation of talent* by technology, to emphasize

[63] For example, consider the case in which x_2 is a manager and x_3 a worker. This would give total output equal to $a\left(\rho\left(x_4 + x_2\right) + \left(x_3 + x_1\right)\right)$ which is smaller than the alternative $a\left(\rho\left(x_4 + x_3\right) + \left(x_2 + x_1\right)\right)$ since $\rho > 1$.

[64] In order to see why this is the case, consider two alternative teams: $\{(x_4, x_2), (x_3, x_1)\}$ or $\{(x_4, x_3), (x_2, x_1)\}$. Any other alternative either produces less or equal output. The first assignment produces higher output if and only if $\bar{a}\left(\rho x_4 + x_2\right) - \zeta\bar{a} + \underline{a}\left(\rho x_3 + x_1\right) - \zeta\underline{a} > \bar{a}\left(\rho x_4 + x_3\right) - \zeta\bar{a} + \underline{a}\left(\rho x_2 + x_1\right) - \zeta\underline{a}$ which simplifies to $\frac{\bar{a}}{\underline{a}} < \rho$.

that high skilled individuals are assigned to the higher technology.

A.4 General Description of the Environment and Case with Finite $\mathbb{X}$

In this section I provide a more general description of the environment that allows to have $\mathbb{X}$ finite and that proves (Lemma 0) that the definition of the allocation and planner problem in the main text is without loss of generality. All the results in Section 2.3 in the main text hold and are proved for this general environment.

A.4.1 Environment

I consider an economy inhabited by heterogeneous individuals, which are indexed by a type $x \in \mathbb{X}$ and a name $i \in \mathbb{I}$. The type set $\mathbb{X}$ is thought of as a set of types that indicate the *relative*[65] skill level, to the extent that if $x' > x$, then all individuals of type x' are more able than x. Throughout the paper, $\mathbb{X}$ will be considered as either the interval $[0, 1]$ or a finite subset of it with an even number of elements.[66] The economy is inhabited by a continuum of individuals of each type x, hence $\mathbb{I} = [0, 1]$. Individuals have an increasing and non-satiated utility function of consumption and no disutility of labor.

Definition A.4.1 (Allocation). *An allocation φ is a quintuple of functions $\{\tilde{\omega}, \tilde{\mu}, \tilde{m}, \tilde{w}, \alpha\}$, where (i) $\tilde{\omega}:\ \mathbb{X}\times\mathbb{I} \to \{0,1\}$ is a function that takes value one if an individual is a worker; (i) $\tilde{\mu}:\ \mathbb{X}\times\mathbb{I} \to \{0,1\}$ is a function that takes value one if an individual is a manager; (iii) $\tilde{m}:\ \mathbb{W} \to \mathbb{M}$ assigns each individual in the set $\mathbb{W}$, that is closure of the set of workers $\{(x,i):\ \tilde{\omega}(x,i) = 1\}$, to a partner in the set $\mathbb{M}$, that is the closure of the set of managers $\{(x,i):\ \tilde{\mu}(x,i) = 1\}$; (iv) $\tilde{w}:\ \mathbb{M} \to \mathbb{W}$ assigns to each manager a corresponding worker; and (v) $\alpha:\ \mathbb{X}\times\mathbb{X} \to \mathbb{A}$ assigns to each possible pair of types a production technology in $\mathbb{A}$.*

The set $\mathbb{A}$ is the set of available technologies, which, if not otherwise specified, is equal to the set of positive real numbers.

Definition A.4.2 (Feasible Allocation). *An allocation φ is feasible if (i) $\tilde{\mu}(x,i) + \tilde{\omega}(x,i) = 1\ \forall (x,i) \in \mathbb{X}\times\mathbb{I}$; (ii) $\hat{m}$ and $\hat{w}$ are measure preserving bijection; and (iii) $m(x,i) = (x',i')$ implies $w(x',i') = (x,i)$ and $w(x,i) = (x',i')$ implies $m(x',i') = (x,i)\ \forall (x,i) \in \mathbb{X}\times\mathbb{I}$. I call $\mathbb{F}$ the set of all feasible allocations.*

The first feasibility constraint guarantees that an individual cannot be simultaneously a manager and a worker. The second one, that each worker is matched with exactly one manager, hence that labor market clears for each type.[67] The third one that matching is consistent, that is for example if an individual (x, i) is matched to a manager (x', i') than the manager (x', i') should be matched himself with worker (x, i). It is also useful to define few auxiliary functions which I use in the definition and characterization of the planner's problem.

Definition A.4.3 (Occupational Choice Functions). *Define (i) $\omega(x) = \int_{\mathbb{I}} \tilde{\omega}(x,i)\, di$ the function that describes the fractions of individuals of type x who are workers; and (ii) $\mu(x) = 1 - \omega(x)$ the function that describes the fractions of individuals of type x who are managers.*

I call ω and μ the occupational choice functions.

[65] Relative to the country specific skill distribution. x ranks individuals from the most to the least able.

[66] In particular, the set $\mathbb{X}$ is defined as $[0, 1]$ if the skill distribution is continuous, and as a finite and evened set if such distribution is discrete. The rational for this assumption is as follows. Consider an arbitrary distribution of (scalar) skills $s \sim F_s$. If F_s is a continuous distribution, then $X = F_s(s) \sim U[0, 1]$. Indeed, $X = \int_{-\infty}^{s} f_s(t)dt$, where f_s is the corresponding density of s. Therefore, for any $x \in [0, 1]$, $\mathbb{P}(X \leq x) = \int_{\{s:\int_{-\infty}^{s} f_s(t)dt \leq x\}} f_s(t)dt = x$. Then we can think of $\mathbb{X}$ as $Supp(F_X) = [0, 1]$ and names distributed uniformly in this set. If F_s is discrete, we consider the following transformation. Let N be an even integer such that $N \leq |Supp(s)|$, where the function $|\cdot|$ denotes cardinality. Define $\{y_n : n \leq N\}$ such that $\mathbb{P}[s \leq y_n] = \frac{n}{N}$. Consider the mapping $y_n \mapsto x_n$, where $x_n = \frac{y_n}{y_N - y_1}$. Then $\mathbb{X} = \{x_n : n \leq N\}$.

[67] Noldeke and Samuelson (2014) shows a case of a matching function that would violate the assumption of a measure preserving bijection.

Definition A.4.4 (Matching Functions). *I define (i) $m_x : \mathbb{W} \to \mathbb{X}$ to be the function that assigns to each worker the type of his manager, hence if $\tilde{m}(x,i) = (x', i')$ then $m_x(x,i) = x'$; (ii) $m : \mathbb{X} \to \mathbb{X}$ to be the function that assigns to each type of workers the average type of its managers, hence $m(x) = \int_{\mathbb{I}} m_x(x,i)\, di$ if $\omega(x) > 0$ and $m(x) = \int_{\mathbb{I}} m_x(\hat{x}, i)\, di$ where $\hat{x} = \max\{x' \in \mathbb{X} : x' < x \text{ and } \omega(x) > 0\}$ otherwise; (iii) $w_x : \mathbb{M} \to \mathbb{X}$ to be the function that assigns to each manager the type of his worker, hence if $\tilde{w}(x,i) = (x', i')$ then $w_x(x,i) = x'$; (ii) $w : \mathbb{X} \to \mathbb{X}$ to be the function that assigns to each type of manager the average type of its workers, hence $w(x) = \int_{\mathbb{I}} w_x(x,i)\, di$ if $\mu(x) > 0$ and $w(x) = \int_{\mathbb{I}} w_x(\hat{x}, i)\, di$ where $\hat{x} = \max\{x' \in \mathbb{X} : x' < x \text{ and } \mu(x) > 0\}$ otherwise.*[68]

With a slight abuse of notation, I call m and w the matching functions, and I focus on these functions for most of the characterization results.

A.4.2 Planner Problem, Existence, and Decentralization

Planner's Problem. The planner picks the feasible allocation that maximizes total output:

$$\max_{\varphi \in \mathbb{F}} Y(\varphi)$$

where output is given by[69]

$$Y(\varphi) = \int_{\mathbb{X}} \int_{\mathbb{I}} g\left(m_x(x,i), x, \alpha\left(m_x(x,i), x\right)\right) \tilde{\omega}(x,i)\, dx di.$$

Existence and Decentralization. I tackle the existence and uniqueness problem indirectly. I describe a competitive equilibrium in my setting, as shown in the appendix A.5, and then show that the first welfare theorem applies. I then recognize, using an insight from Chiappori et al. (2014), that my competitive equilibrium setting is isomorphic to an optimal transportation (Monge–Kantorovich) linear programming problem, for which Chiappori et al. (2010) shows existence results. As a result, a competitive equilibrium exists, and the competitive equilibrium allocation maximizes output. Thus it is the solution to the planner's problem. This gives the next proposition.

Proposition A.4.1 (Existence and Decentralization). *There exists an optimal allocation φ^*, which can be decentralized in a competitive equilibrium.*

This proposition does not guarantee uniqueness. In fact, the sufficient conditions for uniqueness shown in Chiappori et al. (2010) are not satisfied. Nonetheless, this does not represent a limitation of the setting, since I characterize the properties that *any* optimal allocation must satisfy. Hence, in the spirit of Athey et al. (1998), the comparative static exercise is robust, to the extent that it does not depend on the equilibrium selection.

For the characterization of the optimal allocation - shown in the main text - we do not need to keep track of names, and in fact I only work with the functions m, w, μ, and ω. Individual names i do not affect their productivity, given their type. Nonetheless, we had to keep track of them for the general definition of the problem in order to not constraint ex-ante the planner to assign all individuals of the given type to the same partner. For the case in which $\mathbb{X} = [0, 1]$, all workers of any type x are matched with an identical manager, and similarly all managers are matched with an identical worker.[70] For this reason, I skipped altogether the additional definitions

[68] Notice that I have defined m also outside of the optimal allocation, hence for those types for which there are no workers. This is useful when studying deviation to the optimal allocation.

[69] Note that the integral is taken summing only over individuals that are workers. I alternatively could take the integral over the fraction of managers, but cannot sum over both managers and workers, since that would double count the output of a team. Also, notice that for the case in which $\mathbb{X}$ is finite, the integral is a sum. The problem is otherwise identical.

[70] This intuition for this result is similar to the one for bipartite matching model. See for example Eeckhout and Kircher (2012).

of names in the main text. The next Lemma establishes formally this result.

Lemma 0 (Equivalence Result). *Let $\mathbb{X}=[0,1]$, and consider an optimal allocation $\varphi^*=\{\tilde{\mu}^*,\tilde{\omega}^*,\tilde{m}^*,\tilde{w}^*,\alpha^*\}$ that solves the planner problem described. First, $m_x^*(x,i)=m_x^*(x,i')$ and $w_x^*(x,i)=w_x^*(x,i')$ $\forall (x,i,i')\in\mathbb{X}\times\mathbb{I}\times\mathbb{I}$. Second, the set of functions $\{\omega^*,\mu^*,m^*,w^*,\alpha^*\}$, where $\omega^*(x)=\int_{\mathbb{I}}\tilde{\omega}(x,i)\,di$, $\mu^*(x)=1-\omega^*(x)$, $m^*(x)=\int\tilde{m}_x^*(x,i)\,di$, $w^*(x)=\int\tilde{w}_x^*(x,i)\,di$ maximize the problem*

$$\max\int_{\mathbb{X}} g\left(m\left(x\right),x,\alpha\left(m\left(x\right),x\right)\right)\omega\left(x\right)dx$$

$$s.t.$$

$$\begin{aligned}
\mu(x)+\omega(x) &= 1 \quad \forall x\\
\omega(x) &= \mu(m(x)) \quad \forall x \ s.t.\ \omega(x)>0\\
\mu(x) &= \omega(w(x)) \quad \forall x \ s.t.\ \mu(x)>0\\
x &= w(m(x)) \quad \forall x \ s.t.\ \omega(x)>0\\
x &= m(w(x)) \quad \forall x \ s.t.\ \mu(x)>0\\
\omega(x) &\leq 1 \quad \forall x\\
\mu(x) &\leq 1 \quad \forall x
\end{aligned}$$

This Lemma shows that the definition of the problem as included in the main text is equivalent, for the case with $\mathbb{X}=[0,1]$ to the more general one included in this section. Notice that the result shown in 2.3 hold - and are proved below - for this general setting, thus also for $\mathbb{X}$ finite, using in the characterization ω, μ, w and m as defined above.

A.5 Definition of Competitive Equilibrium and Proof of Existence (Proposition 1)

I here the describe a competitive environment that decentralizes the optimal allocation described in the paper. I first provide a broad description, then give a formal definition of the competitive equilibrium, followed by a result that shows the equivalence with bipartite matching and the consequent proof of existence.

A.5.1 Competitive Equilibrium

The environment is the same as described in A.4. There is a continuum of heterogeneous individuals with type $x\in\mathbb{X}$ and name $i\in\mathbb{I}$. An individual therefore is characterized by a pair (x,i). Production is in teams of two, according to the production function $g(x',x,a)$, where x' is the type of the manager, x is the type of the worker, and a is the technology.

I now describe individual optimization.

Problem of a Manager. I start from the problem of a manager of ability x'. He has to pick the optimal type of his worker x, and the technology a to maximize his profits, that is

$$y_\mu(x') = \max_{a\in\mathbb{R},x\in\mathbb{X}} g(x',x,a)-y_\omega(x)$$

taking as given $y_\omega(x)$, that is the equilibrium wage of a worker of type x. I define $\alpha^*(x', x)$ the optimal technology for a team of a manager x' and worker x

$$\alpha^*(x', x) \quad \in \quad \max_{a \in \mathbb{A}} g(x', x, a)$$

and I define $v(x', x)$ the output of a team taking into account optimal technology

$$v(x', x) \quad = \quad g(x', x, \alpha^*(x', x))$$

Notice that each name i of a given ability x solves the same problem, and has the same profit, hence we do not need to keep track of the names i for the definition of this problem. Names i are however relevant for market clearing, since different individuals of the same type might have a different partner (this turns out not to be true in equilibrium for the case with $\mathbb{X} = [0,1]$, but we do not want to impose it as an ex-ante restriction). In fact, I define $w_x(x', i)$ the function that assigns to a manager (x', i) the optimal type of his worker, that is

$$w_x(x', i) \in \max_{x \in \mathbb{X}} v(x', x) - y_\omega(x),$$

where I substituted the optimal technology choice that takes into account the type of the worker. Also, I define $m_x(x, i)$ to be the function that assigns the type of the manager of a worker (x, i). $m_x(x, i)$ is derived from $w_x(x', i)$ as described in market clearing.

Occupational Choice. Each individual (x, i) decides whether to be a manager or a worker. He picks the occupation that gives him the higher income taking as given the schedules $y_\pi(x)$ and $y_\omega(x)$ he would get in either occupation. Let $\tilde{\omega}(x, i)$ be the function that gives value 1 if individual (x, i) is a worker, and $\tilde{\mu}(x, i)$ be the same for managers. The occupational choice can written as

$$\begin{aligned}
\tilde{\omega}(x, i) &= 1 \text{ if } y_\omega(x) > y_\pi(x) \\
\tilde{\mu}(x, i) &= 1 \text{ if } y_\pi(x) > y_\omega(x) \\
\tilde{\omega}(x, i) &= 1 \text{ only if } y_\omega(x) \geq y_\pi(x) \\
\tilde{\mu}(x, i) &= 1 \text{ only if } y_\pi(x) \geq y_\omega(x).
\end{aligned}$$

Market Clearing. Market clearing requires that for each type x, the mass of managers of type x' equals the mass of workers of the types that have x' as a manager, that is

$$\int_{\mathbb{I}} \tilde{\mu}(x', i)\, di = \int_{\mathbb{X}} \int_{\mathbb{I}} 1(m_x(x, i) = x')\, \tilde{\omega}(x, i)\, dx di.$$

and that w_x and m_x are consistent, that is for each type x and x'

$$\int_{\mathbb{I}} 1(m_x(x, i) = x')\, \tilde{\omega}(x, i)\, di = \int_{\mathbb{I}} 1(w_x(x', i) = x)\, \tilde{\mu}(x', i)\, di.$$

Definition A.5.1 (Competitive Equilibrium). *A competitive equilibrium is given by wage and profit schedules* $y_\omega : \mathbb{X} \to \mathbb{R}$, $y_\mu : \mathbb{X} \to \mathbb{R}$; *occupational choice functions* $\tilde{\omega} : \mathbb{X} \times \mathbb{I} \to [0,1]$ *and* $\tilde{\mu} : \mathbb{X} \times \mathbb{I} \to [0,1]$; *matching functions* $m_x : \mathbb{X} \times \mathbb{I} \to \mathbb{X}$, $w_x : \mathbb{X} \times \mathbb{I} \to \mathbb{X}$; *and optimal technology choice* $\alpha^* : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, *such that*

1. *managers pick the optimal type of workers and technology;*

2. *individuals pick optimal occupation;*

3. *labor market clears.*

A.5.2 Equivalence Result

The competitive equilibrium described above is called in the matching literature a roommate problem (or one-sided matching), to emphasize the fact that individuals are allowed ex-ante to match with everyone (i.e. there are no males and females that must marry to each other). Extending an insight present in Chiappori et al. (2014) to the case with a continuum of types, I next show that this problem is isomorphic to a bipartite matching model (two-sided matching), for which we can then use known existence results from the optimal transportation literature. Since Chiappori et al. (2014) proves the equivalence for a finite number of types, I focus here on the continuous case, thus for $\mathbb{X} = [0, 1]$.

In order to define the associated bipartite model, I split for each type x the set of individuals (names) of that type into two groups of identical mass. More specifically, I let $\mathbb{I}_1 = \left[0, \frac{1}{2}\right]$ and $\mathbb{I}_2 = \left[\frac{1}{2}, 1\right]$, so that $\mathbb{I}_1 \cup \mathbb{I}_2 = \mathbb{I}$. I in fact separate the mass of individuals into two identical groups, and then I consider the problem of matching individuals in one group with individuals in the other group, that is a standard bipartite matching problem. For convention, I call an individual i of type x in the group 1, (x_1, i_1), and similarly for group 2, (x_2, i_2). I next define the associated bipartite matching problem.

Definition A..5.2 (Associated Bipartite Matching Problem). *The bipartite matching problem is given by income schedules $y_1 : \mathbb{X} \to \mathbb{R}$ and $y_2 : \mathbb{X} \to \mathbb{R}$; matching functions $m_{x_1} : \mathbb{X} \times \mathbb{I}_1 \to \mathbb{X}$ and $m_{x_2} : \mathbb{X} \times \mathbb{I}_2 \to \mathbb{X}$ that assigns to each individual in either group the type of their partner; and optimal technology $\alpha^* : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ such that*

1. *the income schedules solve*

$$
\begin{aligned}
y_1(x_1) &= \max_{x_2 \in \mathbb{X}} \hat{v}(x_1, x_2) - y_2(x_2) \\
y_2(x_2) &= \max_{x_1 \in \mathbb{X}} \hat{v}(x_1, x_2) - y_1(x_1)
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha^*(x', x) &\in \max_{a \in \mathbb{A}} g(x', x, a) \\
\hat{v}(x_1, x_2) &= \max\left\{g(x_1, x_2, \alpha^*(x_1, x_2)), g(x_2, x_1, \alpha^*(x_2, x_1))\right\}
\end{aligned}
$$

and the outside option of both individuals of not matching is identical for group 1 and 2 and equal to $\max\left\{y_\omega(0), y_\mu(0)\right\}$.

2. *the matching functions are such that for all $(x_1, i_1) \in \mathbb{X} \times \mathbb{I}_1$ and $(x_2, i_2) \in \mathbb{X} \times \mathbb{I}_2$*

$$
\begin{aligned}
m_{x_1}(x_1, i_1) &\in \max_{x_2 \in \mathbb{X}} \hat{v}(x_1, x_2) - y_2(x_2) \\
m_{x_2}(x_2, i_2) &\in \max_{x_1 \in \mathbb{X}} \hat{v}(x_1, x_2) - y_1(x_2)
\end{aligned}
$$

3. *market clear for each type x_1 and x_2, that is*

$$\begin{aligned} \frac{1}{2} &= \int_{\mathbb{X}}\int_{\mathbb{I}_1} 1\left(m_{x_1}(x,i)=x_2\right)dxdi \\ \frac{1}{2} &= \int_{\mathbb{X}}\int_{\mathbb{I}_2} 1\left(m_{x_2}(x,i)=x_1\right)dxdi \end{aligned}$$

I now show that the two problems are isomorphic, that is, given a solution to the associated bipartite matching problem (A.5.1) I can construct a solution to the competitive equilibrium in A.5.2 and vice versa. This is summarized in the next lemma.

Lemma A.5.1 (Isomorphism). *Consider a solution $\{y_1, y_2, m_{x_1}, m_{x_2}, \alpha\}$ of the problem in A.5.2, if it exists, then $\{y_\omega, y_\mu, \tilde{\omega}, \tilde{\mu}, m_x, \alpha\}$ solves the problem in A.5.1, where*

$$\begin{aligned} \tilde{\omega}_1(x_1,i_1) &= 1 \text{ if } m_{x_1}(x_1,i_1) \geq x_1 \\ \tilde{\omega}_2(x_2,i_2) &= 1 \text{ if } m_{x_2}(x_2,i_2) \geq x_2 \\ \tilde{\mu}_1(x_1,i_1) &= 1 \text{ if } m_{x_1}(x_1,i_1) < x_1 \\ \tilde{\mu}_2(x_2,i_2) &= 1 \text{ if } m_{x_2}(x_2,i_2) < x_2; \end{aligned}$$

$$\begin{aligned} y_\omega(x) &= \frac{1\left(\int_{\mathbb{I}_1}\tilde{\omega}_1(x,i_1)\,di_1>0\right)y_1(x)+1\left(\int_{\mathbb{I}_2}\tilde{\omega}_2(x,i_2)\,di_2>0\right)y_2(x)}{1\left(\int_{\mathbb{I}_1}\tilde{\omega}_1(x,i_1)\,di_1>0\right)+1\left(\int_{\mathbb{I}_2}\tilde{\omega}_2(x,i_2)\,di_2>0\right)} \\ y_\mu(x) &= \frac{1\left(\int_{\mathbb{I}_1}\tilde{\mu}_1(x,i_1)\,di_1>0\right)y_1(x)+1\left(\int_{\mathbb{I}_2}\tilde{\mu}_2(x,i_2)\,di_2>0\right)y_2(x)}{1\left(\int_{\mathbb{I}_1}\tilde{\mu}_1(x,i_1)\,di_1>0\right)+1\left(\int_{\mathbb{I}_2}\tilde{\mu}_2(x,i_2)\,di_2>0\right)}; \end{aligned}$$

$$\begin{aligned} \tilde{\omega}(x,i) &= 1(i=i_1)\,\tilde{\omega}_1(x,i_1)+1(i=i_2)\,\tilde{\omega}_2(x,i_2) \\ \tilde{\mu}(x,i) &= 1(i=i_1)\,\tilde{\mu}_1(x,i_1)+1(i=i_2)\,\tilde{\mu}_2(x,i_2); \end{aligned}$$

and

$$\begin{aligned} m_x(x,i) &= 1(i\in\mathbb{I}_1)\left(\tilde{\omega}_1(x,i)\,m_{x_1}(x,i)\right)+1(i\in\mathbb{I}_2)\left(\tilde{\omega}_2(x,i)\,m_{x_2}(x,i)\right) \\ w_x(x,i) &= 1(i\in\mathbb{I}_1)\left(\tilde{\mu}_1(x,i)\,m_{x_1}(x,i)\right)+1(i\in\mathbb{I}_2)\left(\tilde{\mu}_2(x,i)\,m_{x_2}(x,i)\right) \end{aligned}$$

Similarly, consider a solution $\{y_\omega, y_\mu, \tilde{\omega}, \tilde{\mu}, m_x, w_x, \alpha\}$ of the problem in A.5.1, if it exists, then $\{y_1, y_2, m_{x_1}, m_{x_2}, \alpha\}$ solves the problem A.5.2, where

$$\begin{aligned} m_{x_1}(x_1,i_1) &= \omega(x_1,i_1)\,m_x(x_1,i_1)+\mu(x_1,i_1)\,w_x(x_1,i_1) \\ m_{x_2}(x_2,i_2) &= \omega(x_2,i_2)\,m_x(x_2,i_2)+\mu(x_2,i_2)\,w_x(x_2,i_2) \end{aligned}$$

and

$$
\begin{aligned}
y_1(x_1) &= \frac{\int_{\mathbb{I}_1} \tilde{\omega}(x_1,i)\, y_\omega(x_1)\, di_1}{\int_{\mathbb{I}_1} \tilde{\omega}(x_1,i)\, di_1} + \frac{\int_{\mathbb{I}_1} \tilde{\mu}(x_1,i)\, y_\mu(x_1)\, di_1}{\int_{\mathbb{I}_1} \tilde{\mu}(x_1,i)\, di_1} \\
y_2(x_2) &= \frac{\int_{\mathbb{I}_2} \tilde{\omega}(x_2,i)\, y_\omega(x_1)\, di_2}{\int_{\mathbb{I}_2} \tilde{\omega}(x_2,i)\, di_2} + \frac{\int_{\mathbb{I}_2} \tilde{\mu}(x_2,i)\, y_\mu(x_2)\, di_2}{\int_{\mathbb{I}_2} \tilde{\mu}(x_2,i)\, di_2};
\end{aligned}
$$

and last the individuals into $\mathbb{I}$ *are rearranged such that market clearing holds with* $\mathbb{I}_1 = \left[0, \frac{1}{2}\right]$ *and* $\mathbb{I}_2 = \left[\frac{1}{2}, 1\right]$*, where this rearrangement is without loss of generality since all individuals of a given type* x *are identical.*

Proof. First notice that the optimal technology α^* is identical in the two settings by definition. Next, I focus on the income schedules. Let's notice that taking the envelope condition and the first order condition of the planner problem in A.5.1, we get that $\forall (x,i) \in \mathbb{X} \times \mathbb{I}$

$$
\begin{aligned}
y'_\mu(x) &= v_1(x, w_x(x,i)) \\
y'_\omega(x) &= v_2(m_x(x,i), x)
\end{aligned}
\tag{7}
$$

and that taking the first order condition in A.5.2 we get that

$$
\begin{aligned}
y'_1(x_1) &= \hat{v}_1(x_1, x_2) \\
y'_2(x_2) &= \hat{v}_2(x_1, x_2)
\end{aligned}
$$

which using the definition of $\hat{v}$ and the above defined functions can be rewritten as

$$
\begin{aligned}
y'_1(x_1) &= \tilde{\mu}_1(x_1, i_1)\, v_1(x_1, m_{x_1}(x_1, i_1)) + \tilde{\omega}_1(x_1, i_1)\, v_2(m_{x_1}(x_1, i_1), x_1) \\
y'_2(x_2) &= \tilde{\mu}_2(x_2, i_2)\, v_1(x_2, m_{x_2}(x_2, i_2)) + \tilde{\omega}_2(x_2, i_2)\, v_2(m_{x_2}(x_2, i_2), x_2)
\end{aligned}
\tag{8}
$$

which implies that, given the above construction, (7) holds if and only if (8) holds. Then, since $y_1(0) = y_2(0) = \max\{y_\omega(0), y_\mu(0)\}$ we have shown that the income schedules satisfy the isomorphism.

Last I need to show that market clearing holds. Market clearing in A.5.2 can be rewritten as

$$
\begin{aligned}
\int_{\mathbb{I}_1} \tilde{\mu}_1(x_1, i_i)\, di &= \int_{\mathbb{X}} \int_{\mathbb{I}_2} 1(m_{x_2}(x_2, i_2) = x_1)\, \tilde{\omega}_2(x_2, i_2)\, dx_2 di_2 \\
\int_{\mathbb{I}_2} \tilde{\mu}_2(x_2, i_2)\, di &= \int_{\mathbb{X}} \int_{\mathbb{I}_1} 1(m_{x_1}(x_1, i_1) = x_2)\, \tilde{\omega}_1(x_1, i_1)\, dx_1 di_1
\end{aligned}
$$

which, since $\mathbb{I} = \mathbb{I}_1 \cup \mathbb{I}_2$, implies that the market clearing in A.5.1 holds, that is

$$
\int_{\mathbb{I}} \tilde{\mu}(x', i)\, di = \int_{\mathbb{X}} \int_{\mathbb{I}} 1(m_x(x,i) = x')\, \tilde{\omega}(x,i)\, dx di
$$

Also, given market clearing in A.5.1, we can rearrange i in such a way that market clearing in A.5.2 holds as well. This rearrangement is without loss of generality, since names are irrelevant for output. This concludes the proof. □

A.5.3 Proof of the Existence

In order to prove existence, we can now apply known results for bipartite matching problems. Since the function $\hat{v}(x_1, x_2)$ as defined is, by the maximum theorem, continuous we can apply Theorem 1 of Chiappori et al. (2010), that proves existence of a stable matching that maximizes total output. Applying the isomorphism result, the stable matching corresponds to the solution of the competitive equilibrium defined in A.5.1. Hence we have shown that a solution that maximizes total output exists and provided a decentralization. □

It may be worth to remark that the condition (subtwist) that Theorem 3 of Chiappori et al. (2010) shows to be sufficient for uniqueness of an equilibrium in mixed strategy is not satisfied in this context.[71]

A.6 Proofs of Preliminary Results (Lemmas 0-5)

As a preliminary step, let's define the net output of team that pick the optimal technology as

$$v\left(x',x\right) = \max_a g\left(x',x,a\right),$$

where recall that $g\left(x',x,a\right) = af\left(x',x\right) - c\left(x',x,a\right)$. Taking the first order condition of the problem above we get

$$f\left(x',x\right) = c_3\left(x',x,\alpha^*\left(x',x\right)\right). \tag{9}$$

Using the envelope theorem, we can calculate the marginal products and the complementarity of v: [72]

$$\begin{aligned} v_1\left(x',x\right) &= \alpha^*\left(x',x\right)f_1\left(x',x\right) - c_1\left(x',x,\alpha^*\left(x',x\right)\right) \\ v_2\left(x',x\right) &= \alpha^*\left(x',x\right)f_2\left(x',x\right) - c_2\left(x',x,\alpha^*\left(x',x\right)\right) \\ v_{12}\left(x',x\right) &= \alpha^*\left(x',x\right)f_{12}\left(x',x\right) + c_{33}\left(x',x,\alpha^*\left(x',x\right)\right)\alpha_1^*\left(x',x\right)\alpha_2^*\left(x',x\right). \end{aligned} \tag{10}$$

where I used assumption *(2.4)*.

Proof of Lemma 0 I first show that $m_x^*\left(x,i\right) = m_x^*\left(x,i'\right)\ \forall\left(x,i,i'\right) \in \mathbb{X}\times\mathbb{I}\times\mathbb{I}$. I showed the equivalence between the planner's problem and the competitive equilibrium. Using the first order and envelope conditions in the competitive equilibrium I showed, in equation (7), that $\forall\left(x,i\right) \in \mathbb{X}\times\mathbb{I}$

$$\begin{aligned} y_\mu'\left(x\right) &= v_1\left(x,w_x^*\left(x,i\right)\right) \\ y_\omega'\left(x\right) &= v_2\left(m_x^*\left(x,i\right),x\right). \end{aligned}$$

Using (10) and assumption *(1.2)* this implies that $m_x^*\left(x,i\right) = m_x^*\left(x,i'\right)$ *and* $w_x^*\left(x,i\right) = w_x^*\left(x,i'\right)\ \forall\left(x,i,i'\right) \in \mathbb{X}\times\mathbb{I}\times\mathbb{I}$ must hold, All individuals of type x have the same marginal income and the production function is strictly increasing, in order for this to be satisfied - conditional on having the same occupation - they must be matched with identical partners.

Given this, and using the definition of ω^* and m^* we get

$$\int_{\mathbb{X}}\int_{\mathbb{I}} g\left(m_x^*\left(x,i\right),x,\alpha^*\left(m_x^*\left(x,i\right),x\right)\right)\tilde{\omega}^*\left(x,i\right)dxdi = \int_{\mathbb{X}} g\left(m^*\left(x\right),x,\alpha^*\left(m^*\left(x\right),x\right)\right)\omega^*\left(x\right)dx.$$

[71] This condition can be rewritten as $\hat{v}: \mathbb{X}\times\mathbb{X} \to [0,\infty)$, is subtwisted if there is a zero measure set $\mathbb{X}_L$ such that whenever $\hat{v}_1(x_1,x_2) = \hat{v}_1(x_1,x_2')$ for some $x_1 \in \mathbb{X}_1\backslash\mathbb{X}_L$ and $x_2 \neq x_2'$, then, a fortiori, x_1 is either the unique global maximum or the unique global minimum of the difference $\hat{v}(x_1,x_2) - \hat{v}(x_1,x_2')$ on $\mathbb{X}$.

[72] These results comes from the application of the envelope theorem. I omit argument of the functions for brevity. The first order conditions given $f = c_3$, which can be differentiated to get $f_1 = c_{33}\alpha_1^* + c_{13}$ and $f_2 = c_{33}\alpha_2^* + c_{23}$. Next, differentiating v with respect to its first argument we get $v_1 = \alpha^* f_1 + \alpha_1^* f - c_1 - c_3\alpha_1^*$ which simplifies to $v_1 = \alpha^* f_1 - c_1$. We can follow the similar steps for v_2. Next, differentiate v_1 with respect to its second argument to get $v_{12} = \alpha_{12}^* f + \alpha_1^* f_2 + \alpha_2^* f_1 + \alpha^* f_{12} - c_{12} - c_{13}\alpha_2^* - c_{23}\alpha_1^* - c_{33}\alpha_1^*\alpha_2^* - c_3\alpha_{12}^*$, which simplifies to $v_{12} = \alpha^* f_{12} - c_{12} + c_{33}\alpha_1^*\alpha_2^*$.

Last, the fact that $\varphi^* \in \mathbb{F}$ directly implies that the constraints of the problem are satisfied as well. This concludes the proof. $\square$

Proof of Lemma 1 By Lemma 1 and assumption *(1.4)* the function v has complementarity: $v_{12} > 0$. I prove the Lemma by contradiction. Let $x' > x$ and $m^*(x) > m^*(x')$, where recall that $m^*(x) = \int_{\mathbb{I}} m_x^*(x,i)\,di$. There must be $i, i' \in \mathbb{I}$ such that $m_x^*(x,i) > m_x^*(x',i')$, Call $x''' = m_x^*(x,i)$ and $x'' = m_x^*(x',i')$. In order for m^* to be induced by the optimal allocation it must be that

$$v(x''',x) + v(x'',x') \quad \geq \quad v(x'',x) + v(x''',x').$$

Rewrite the above, using the fundamental theorem of calculus (I use this theorem throughout the appendix, but acknowledge it only here) as

$$\int_{x''}^{x'''} v_1(z,x)\,dz \quad \geq \quad \int_{x''}^{x'''} v_1(z,x')\,dz$$

which can be farther rewritten as

$$0 \quad \geq \quad \int_{x''}^{x'''}\int_{x}^{x'} v_{12}(z,y)\,dydz$$

which contradicts $v_{12} > 0$. $\square$

Proof of Lemma 2 Consider two individuals of type x' and x. We need to show that

$$v(x',x) \quad \geq \quad v(x,x')$$

if and only if $x' \geq x$. Let $x' \geq x$ and rewrite the equation above as

$$\int_{x}^{x'} v_1(z,x)\,dz - \int_{x}^{x'} v_2(x,z)\,dz \quad \geq \quad 0.$$

Next, I show that $v_1(x',x) \geq v_2(x,x')\ \forall x' \geq x$. Use (10) and rewrite $v_1(x',x) \geq v_2(x,x')$ as

$$\alpha^*(x',x)\,f_1(x',x) - c_1(x',x,\alpha^*(x',x)) \quad \geq \quad \alpha^*(x,x')\,f_2(x,x') - c_2(x,x',\alpha^*(x',x))$$

which can be farther rewritten as

$$\alpha^*(x,x')\,f_1(x',x) - c_1(x',x,\alpha^*(x,x')) + \int_{\alpha^*(x,x')}^{\alpha^*(x',x)} (f_1(x',x) - c_{13}(x',x,a))\,da \quad \geq \quad \alpha^*(x,x')\,f_2(x,x') - c_2(x,x',\alpha^*(x,x')).$$

This last inequality holds since: (i) by assumption *(1.5)* $f_1(x',x) > f_2(x',x)$; (ii) by assumptions *(2.3)* and *(2.4)* $-c_1(x',x,\alpha^*(x,x')) > -c_2(x,x',\alpha^*(x,x'))$; (iii) by assumption *(2.5)* $f_1(x',x) - c_{13}(x',x,a) > 0$; and last (iv) by Lemma 1 $\alpha^*(x',x) > \alpha^*(x,x')$. Hence we have shown that if $x' \geq x$, then $v(x',x) \geq v(x,x')$. An identical argument for the case $x' < x$ completes the proof. $\square$

Proof of Lemma 3 Differentiate the first order condition 9 with respect to x' to get

$$\alpha_1^* (x', x) = \frac{f_1 (x', x) - c_{13} (x', x, \alpha^* (x', x))}{c_{33} (x', x, \alpha^* (x', x))}$$

and with respect to x to get

$$\alpha_2^* (x', x) = \frac{f_2 (x', x) - c_{23} (x', x, \alpha^* (x', x))}{c_{33} (x', x, \alpha^* (x', x))}$$

assumption *(2.5)* guarantees that $\alpha_1^* > 0$ and *(2.5)* that $\alpha_2^* \geq 0$. Assumption (2.7) guarantees that $\frac{a_1^*}{a^*} > \frac{a_2^*}{a^*}$. □

Proof of Lemma 4 By Lemma 3: $m^* (x) \geq x$ and thus using feasibility $w^* (x) \leq x$. By Lemma 1: $\alpha_1^* > 0$ and $\alpha_2^* \geq 0$. The result follows immediately. □

Proof of Lemma 5 The result follows immediately from the fact that both $\frac{\alpha_1^*}{\alpha^*}$ and $\frac{\alpha_2^*}{\alpha^*}$ are decreasing in $\frac{c_{33}}{c_3}$, as can be see from their explicit expression shown in the proof of Lemma 3 and from the fact that $\alpha^* = c_3^{-1} (x', x, f (x', x))$ and if c_3 is higher, than c_3^{-1} is smaller. □

A.7 Proof of Proposition 2

I provide a separate proof for the case in which $\mathbb{X}$ is finite and for the case in which $\mathbb{X} = [0, 1]$. I start from the former.

Proof of Proposition 2 for $\mathbb{X}$ **is a finite subset of** $[0, 1]$ I first prove the conditions for talent segmentation. The proof is by contradiction. Let's assume that talent is not segmented. Hence there exist two individuals (x'', i'') and (x', i') such that $x'' > x'$ and $\tilde{\omega} (x'', i'') = 1$ and $\tilde{\mu} (x', i') = 1$. By Lemma 2 and 3, it must be that $m_x^* (x'', i'') \geq x''$ and $w_x^* (x', i') \leq x'$. Call $x''' = m_x^* (x'', i'')$ and $x = w_x^* (x', i')$. We are thus considering four types $x''' \geq x'' > x' \geq x$. In order for this to be an optimal allocation, the following inequality must be satisfied

$$v (x''', x'') + v (x', x) > v (x''', x') + v (x'', x) .$$

We can rewrite the inequality above following usual steps as

$$\int_{x'}^{x''} v_2 (x''', z) \, dz > \int_{x'}^{x''} v_1 (z, x) \, dz$$

which, substituting the previous expression (10) from the envelope condition becomes

$$\int_{x'}^{x''} [\alpha^* (x''', z) f_2 (x''', z) - c_2 (x''', z, \alpha^* (x''', z))] \, dz \geq \int_{x'}^{x''} [\alpha^* (z, x) f_1 (z, x) - c_1 (z, x, \alpha^* (z, x))] \, dz. \tag{11}$$

The left hand side can be rewritten as

$$\int_{x'}^{x''} [\alpha^* (z, x) f_2 (x, z) - c_2 (x, z, \alpha^* (z, x))] \, zd + \int_{x'}^{x''} \int_{\alpha^* (z,x)}^{\alpha^* (x''', z)} (f_2 (x, z) - c_{23} (x, z, a)) \, dadz.$$

By assumption (1.5) $\alpha^* (z, x) f_2 (x, z) < \alpha^* (z, x) f_1 (z, x)$; by assumptions (2.3) and (2.4) $-c_2 (x, z, \alpha^* (z, x)) < -c_1 (z, x, \alpha^* (z, x))$; which thus shows that for the inequality (11) to be satisfied, $\Lambda \equiv \int_{\alpha^* (z,x)}^{\alpha^* (x''', z)} (f_2 (x, z) - c_{23} (x, z, a)) \, da$ must be sufficiently large with respect to $\alpha^* (z, x)$. Last, notice that by assumption (2.6) $f_2 (x, z) - c_{23} (x, z, a) > 0$

$\forall a \in \left[\alpha^*(z,x), \alpha^*(x''',z)\right)$. Hence, for Λ to be large enough, it must be that the interval $\left[\alpha^*(z,x), \alpha^*(x''',z)\right)$ is large enough with respect to $\alpha^*(z,x)$. If $\frac{c_{33}}{c_3}$ is small enough Λ is small and thus (11) cannot be satisfied. For example, in the limit as $\frac{c_{33}}{c_3} \to \infty$, $\alpha^*(x''',z) \to \alpha^*(z,x)$, and the inequality is not satisfied. This gives a contradiction which proves that talent must be segmented by occupation.

Next, I prove the conditions for talent segregation. The proof is again by contradiction. Let's assume that talent is not segregated. Hence there exist two individuals (x'', i'') and (x', i') such that $x'' > x'$ and $A^*(x') > A^*(x'')$. Applying previous lemmas, this implies that must exist two individuals (x', i') and (x'', i'') such that $\tilde{\omega}(x', i) = 1$ and $\tilde{\mu}(x'', i') = 1$ and $m^*_x(x', i') > x''$ and $w^*_x(x'', i'') < x'$. Call $x''' = m^*_x(x', i')$ and $x = w^*_x(x'', i'')$. We are thus considering four types $x''' > x'' > x' > x$. In order for this to be an optimal allocation, the following inequality must be satisfied

$$v(x''', x') + v(x'', x) > v(x''', x'') + v(x', x).$$

Notice that this inequality is just as the one for talent segmentation, but with opposite sign. We can thus follow the previous steps, and show that in order for this to be satisfied, it must be that Λ is small enough with respect to $\alpha^*(z,x)$. However, notice that

$$\frac{\int_{\alpha^*(z,x)}^{\alpha^*(x''',z)} \left(f_2(x,z) - c_{23}(x,z,a)\right) da}{\alpha^*(z,x)} \to \infty$$

as $\frac{c_{33}}{c_3} \to 0$. Hence, there exist a value of $\frac{c_{33}}{c_3}$ small enough such that the inequality cannot be satisfied and as a result we reach a contradiction. □

Proof of Proposition 2 for $\mathbb{X} = [0,1]$ First, I write the necessary conditions for optimality, exploiting the formulation used in the main text, and proved in Lemma 0, and noticing that optimal technology is maximized for each team independently on the occupational choice function, and can thus be solved for ex-ante, we get that $\{\omega^*(x), \mu^*(x), m^*(x), w^*(x)\}$ are the solution to the problem

$$\max_{\omega,\mu,m,w} \int_{\mathbb{X}} v(m(x), x)\, \omega(x)\, dx$$

$s.t.$

$$\begin{aligned}
\mu(x) + \omega(x) &= 1 \quad \forall x \\
\omega(x) &= \mu(m(x)) \quad \forall x \ s.t.\ \omega(x) > 0 \\
\mu(x) &= \omega(w(x)) \quad \forall x \ s.t.\ \mu(x) > 0 \\
x &= w(m(x)) \quad \forall x \ s.t.\ \omega(x) > 0 \\
x &= m(w(x)) \quad \forall x \ s.t.\ \mu(x) > 0 \\
\omega(x) &\in [0,1] \quad \forall x \\
\mu(x) &\in [0,1] \quad \forall x
\end{aligned}$$

From the first order conditions of this problem with respect to the function ω and μ evaluated for each type x we get that:

1. $\forall x$ s.t. $\omega^*(x) > 0$ and $\mu^*(x) > 0$ then $v_1(x, w^*(x)) = v_2(m^*(x), x)$;

2. $\forall x$ s.t. $\mu^*(x-\epsilon)=0$ and $\mu^*(x)>0$ then $v_1(x,w^*(x))\geq v_2(m^*(x),x)$;

3. $\forall x$ s.t. $\mu^*(x-\epsilon)>0$ and $\mu^*(x)=0$ then $v_1(x,w^*(x))\leq v_2(m^*(x),x)$.

If any of the three conditions are not satisfied, it is possible to construct an optimal deviation. As an example, consider the case of an allocation φ around a type $\tilde{x}$ that does not satisfy (2) . Then we could construct an alternative allocation $\tilde{\varphi}$ such that

$$\begin{aligned}\tilde{\mu}(\tilde{x}) &= \mu(\tilde{x})-\Delta \\ \tilde{\mu}(\tilde{x}-\epsilon) &= \mu(\tilde{x}-\epsilon)+\Delta\end{aligned} \tag{12}$$

for a small Δ. This deviation satisfies market clearing by construction and increases total output if $v_1(x,w(x))<v_2(m(x),x)$, as assumed. Intuitively then, if equation (12) is not satisfied, there exist a marginal individual $\tilde{x}$ such that some of those at least as skilled as $\tilde{x}$ are managers and those less skilled than $\tilde{x}$ are all workers,. However, at $\tilde{x}$, the marginal product of increasing the ability of workers is higher than that of increasing the ability of managers. As a result, this cannot be an optimal allocation, since it would be optimal to swap a worker and a manager around $\tilde{x}$. I next use this necessary conditions to prove the proposition.

I start by the case with segmentation. In order to prove that segmentation must be satisfied it is sufficient to show that $v_1(x,w(x))>v_2(m(x),x)\ \forall x\in\mathbb{X}$ and for any m and w that satisfies the properties described in Lemmas 1-4 and market clearing. If this condition is satisfied, it is not possible to have a worker equally or less skilled than a manager, since that would require, according to conditions (1) and (3) above, that $v_1(m^*(x),x)\leq v_2(x,w^*(x))$. Using the familiar envelope condition (10) we can rewrite $v_1(x,w(x))>v_2(m(x),x)$ as

$$\alpha^*(x,w(x))f_1(x,w(x))-c_1(x,w(x),\alpha^*(x,w(x))) > \alpha^*(m(x),x)f_2(m(x),x)-c_2(m(x),x,\alpha^*(m(x),x)) \tag{13}$$

which can be rewritten as

$$\frac{f_1(x,w(x))}{f_2(m(x),x)} > \frac{\alpha^*(m(x),x)}{\alpha^*(x,w(x))}+\frac{c_1(x,w(x),\alpha^*(x,x'))-c_2(m(x),x,\alpha^*(m(x),x))}{\alpha^*(x,w(x))f_2(m(x),x)}. \tag{14}$$

Next, by assumption (2.5) $\frac{f_1(x,w(x))}{f_2(m(x),x)}>1$, and thus if $\frac{c_{33}}{c_3}$ is large enough this inequality is satisfied. To see why this is the case, consider that in the limit as $\frac{c_{33}}{c_3}\to\infty$, then $\frac{\alpha^*(m(x),x)}{\alpha^*(x,w(x))}\to 1$ and $\frac{c_1(x,w(x),\alpha^*(x,x'))-c_2(m(x),x,\alpha^*(m(x),x))}{\alpha^*(x,w(x))f_2(m(x),x)}\to\delta$ where $\delta<0$ by assumption (2.3) and (2.4). Then, by usual continuity argument, there exists a scalar κ_2 such that if $\frac{c_{33}}{c_3}\geq\kappa_2$, then (14) is satisfied. This proves that talent is segmented by occupation if $\frac{c_{33}}{c_3}\geq\kappa_2$. Last, notice that I have been omitting throughout this discussion the support of a over which the conditions on c hold. It is immediate to see that it is sufficient that they hold over the set $\mathbb{A}_{c,\mathbb{X}}=\{\alpha^*(x',x):\ x',x\in\mathbb{X}\}$. In fact, throughout the proof, c is not evaluated at values of a outside of $\mathbb{A}_{c,\mathbb{X}}$. Additionally, it has of course to hold for any $(x',x)\in\mathbb{X}\times\mathbb{X}$. This same comment holds for the case of talent segregation below.

Next, let's consider the case of talent segregation. Consider again the inequality $v_1(x,w(x))>v_2(m(x),x)$ as shown in equation (13) I rewrite the left hand side (LHS) as

$$\alpha^*(x,w(x))f_1(x,w(x))-c_1(x,w(x),\alpha^*(x,w(x))) = \left[\alpha^*(x,x)f_1(x,w(x))-c_1(x,w(x),\alpha^*(x,x)) - \int_{\alpha^*(x,w(x))}^{\alpha^*(x,x)}(f_1(x,w(x))-c_{13}(x,w(x),a))\,da\right]$$

and the right hand side (RHS) as

$$\alpha^*\left(m\left(x\right),x\right)f_2\left(m\left(x\right),x\right)-c_2\left(m\left(x\right),x,\alpha^*\left(m\left(x\right),x\right)\right) = \left[\alpha^*\left(x,x\right)f_2\left(m\left(x\right),x\right)-c_2\left(m\left(x\right),x,\alpha^*\left(xx\right)\right) + \int_{\alpha^*(x,x)}^{\alpha^*(m(x),x)}\left(f_2\left(m\left(x\right),x\right)-c_{23}\left(m\left(x\right),x\right)\right)da\right]$$

Rearranging we get

$$\left[f_1\left(x,w\left(x\right)\right)-f_2\left(m\left(x\right),x\right) - \frac{\left[c_1\left(x,w\left(x\right),\alpha^*\left(x,x\right)\right)-c_2\left(m\left(x\right),x,\alpha^*\left(xx\right)\right)\right]}{\alpha^*\left(x,x\right)}\right] > \left[\frac{\int_{\alpha^*(x,x)}^{\alpha^*(m(x),x)}\left(f_2\left(m\left(x\right),x\right)-c_{23}\left(m\left(x\right),x\right)\right)da}{\alpha^*\left(x,x\right)} + \frac{\int_{\alpha^*(x,w(x))}^{\alpha^*(x,x)}\left(f_1\left(x,w\left(x\right)\right)-c_{13}\left(x,w\left(x\right),a\right)\right)da}{\alpha^*\left(x,x\right)}\right]. \quad (15)$$

Then, as $\frac{c_{33}}{c_3} \to 0$, the RHS diverges as long as $m\left(x\right) > x$, so that there exists a small value ϵ, such that if $m\left(x\right) > x+\epsilon$, then $v_1\left(x,w\left(x\right)\right) < v_2\left(m\left(x\right),x\right)$. Where, $\epsilon \to 0$ if $\frac{c_{33}}{c_3} \to 0$.

Next, let's assume that there is not segregation of talent by technology, hence there exist $x' > x$ such that $A^*\left(x'\right) < A^*\left(x\right)$. In order for this to be the case, it must be that x is a worker, x' is a manager, and the manager matched with x is more skilled than x', thus $m^*\left(x\right) > x'$. I next build a contradiction when $\frac{c_{33}}{c_3} \to 0$. Let's take the smallest x in the interval $[x,x']$ such that $\mu\left(x\right) > 0$. Call it $\hat{x}$. By the necessary conditions above, $v_1\left(\hat{x},w^*\left(\hat{x}\right)\right) \geq v_2\left(m^*\left(\hat{x}\right),\hat{x}\right)$ must hold. For the previous argument, this implies that as $\frac{c_{33}}{c_3} \to 0$ $m^*\left(\hat{x}\right)-\hat{x} < \epsilon$ with $\epsilon \to 0$. By the definition of $\hat{x}$, Lemma 2, and market clearing, we get that $m^*\left(x\right) \leq m^*\left(\hat{x}\right)-\left(\hat{x}-x\right)$. This in turn implies that $m^*\left(x\right)-x < \epsilon$ with $\epsilon \to 0$ which contradicts $m^*\left(x\right) > x'$. This concludes the proof. □

Remark It can be instructive to consider the case in which $c_1 = c_2 = 0$. In this case, the condition $v_1\left(x,w\left(x\right)\right) > v_2\left(m\left(x\right),x\right)$ simplifies to

$$\frac{f_1\left(x,w\left(x\right)\right)}{f_2\left(m\left(x\right),x\right)} > \frac{\alpha^*\left(m\left(x\right),x\right)}{\alpha^*\left(x,w\left(x\right)\right)}$$

which is the inequality I discuss in the main text. This inequality is easy to interpret - and prevents the need to go through the derivations of (14) and (15). It shows that if the gap in optimal technology is always small enough - then talent segmentation is optimal. Otherwise, if instead the cost function is close to linear, the RHS diverges for any $m\left(x\right) > x$ - since two similarly skilled teams would pick very different technologies - and thus the previous limit argument shows that talent segregation is optimal.

Heuristic Argument for Special Case It can be instructive to consider the case in which $c_1 = c_2 = 0$, and consider a simple heuristic argument that highlights how the convexity of c maps into stronger complementarity. Let's consider the value of a team $v\left(x',x\right)$. For standard arguments, if a Spence-Mirrless condition holds on v, then talent is segmented by occupation. This conditions reads as

$$\min_{y\in\mathbb{X}} \quad v_1\left(x,y\right) \geq \max_{z\in\mathbb{X}} v_2\left(z,x\right)$$

which can be further rewritten as

$$\begin{aligned} v_1(x, x_{min}) &\geq v_2(x_{max}, x) \\ \alpha^*(x, x_{min}) f_1(x, x_{min}) &\geq \alpha^*(x_{max}, x) f_2(x_{max}, x) \end{aligned}$$

We can then do a first order Taylor approximateion around x for both the left and the right hand sides to obtain

$$\begin{aligned} v_1(x,x) - (x - x_{min}) v_{12}(x,x) &\geq v_2(x,x) + (x_{max} - x) v_{12}(x,x) \\ \alpha^* f_1 - (x - x_{min}) \left(\alpha^* f_{12} + \frac{f_1 f_2}{c_{33}} \right) &\geq \alpha^* f_2 + (x_{max} - x) \left(\alpha^* f_{12} + \frac{f_1 f_2}{c_{33}} \right) \\ f_1 - f_2 &\geq (x_{max} - x_{min}) \left(f_{12} + \frac{f_1 f_2}{c_3^{-1} c_{33}} \right) \end{aligned} \tag{16}$$

where in the second line I used the envelope theorem, and in the third line I rearranged and used the fact that $\alpha^*(x,x) = c_3^{-1}(x, x, f(x,x))$. I also omitted the arguments, since they are all equal to (x,x).

We can then compare equations 16 and 6 in the previous appendix. Equation 6 holds by assumption. However, notice that equation 16 has an additional positive term on the right hand side, $\frac{f_1 f_2}{c_3^{-1} c_{33}}$. In fact, the overall complementarity is given by the one in the production of labor, f_{12}, and the one through endogenous technology choice, $\frac{f_1 f_2}{c_3^{-1} c_{33}}$. This last term is small if the convexity of the cost of technology - thus $c_3^{-1} c_{33}$ - is large, as a result the single-crossing holds on v and talent is segmented by occupation. Instead, when the convexity is very low, then this second term goes to infinite. As a result, the single-crossing holds within small partition of the sample space - thus for $x_{max} - x_{min}$ very small. This in turn implies that individuals must match with others of very similar type.

A.8 Characterization and Proofs for the Quasi-Linear Model

I here completely characterize the optimal allocation for the quasi-linear model as a function of $\eta = \frac{\bar{a}}{\underline{a}}$. I also provide the proof of Proposition 3.

The allocation can take one of five shapes, that are described in words in the propositions and are graphically represented in the Figures A.15, A.16, A.17, A.18, and A.19. Within the proof of this propositions I also provide the proofs of Lemmas 6, 7, and 8. Note that in the proposition I omit for brevity description of w^* since it can be directly imputed through market clearing, that is $w^* = m^{*-1}$ and the definition of w itself.

Proposition 3.b (Characterization of Quasi-Linear Model). *The optimal allocation $\{\omega^*, \mu^*, \alpha^*, m^*, w^*\}$ takes one of five shapes depending on the value of $\eta = \frac{\bar{a}}{\underline{a}}$. In each shape, the optimal technology α^* is given by the cutoff policy shown in Lemma 4. There exists four finite constants bigger than one, $\eta_1 < \eta_2 < \eta_3 < \eta_4$, such that*

(i) If $\eta \leq \eta_1$. *Skills are segmented by occupation: low skilled individuals are workers and high skilled ones are managers. Specifically,*

$$\omega^*(x) = \begin{cases} 1 & \text{if } x \leq \frac{1}{2} \\ 0 & \text{if } x > \frac{1}{2} \end{cases}, \qquad \mu^*(x) = \begin{cases} 0 & \text{if } x \leq \frac{1}{2} \\ 1 & \text{if } x > \frac{1}{2} \end{cases}$$

$$A^*(x) = \begin{cases} \bar{a} & \text{if } x \in \left[\frac{1}{2} - (1-\zeta), \frac{1}{2}\right] \cup [\zeta, 1] \\ \underline{a} & \text{if } x \in \left[0, \frac{1}{2} - (1-\zeta)\right] \cup \left[\frac{1}{2}, \zeta\right] \end{cases}$$

$$m^*(x) = \begin{cases} \frac{1}{2} + x & \text{if } x \leq \frac{1}{2} \\ 1 & \text{if } x > \frac{1}{2} \end{cases}.$$

(ii) If $\eta \in [\eta_1, \eta_2]$. Skills are imperfectly segmented by occupation: for a set of middle skilled individuals the optimal allocation dictates a mixed strategy in which for each type x some individuals are going to be workers and some managers. In this set, the workers use the frontier technology, $\bar{a}$, while the managers use the traditional one, $\underline{a}$. Specifically there exists constants $\hat{x}_1$, $\hat{x}_2$ such that

$$\omega^*(x) = \begin{cases} 1 & \text{if } x < \hat{x}_1 \\ \frac{\eta}{\eta+1} & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\ 0 & \text{if } x > \hat{x}_2 \end{cases}, \qquad \mu^*(x) = \begin{cases} 0 & \text{if } x < \hat{x}_1 \\ 1 - \frac{\eta}{\eta+1} & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\ 1 & \text{if } x > \hat{x}_2 \end{cases}$$

$$A^*(x) = \begin{cases} \bar{a} & \text{if } x \in [\hat{x}_1 - (1 - \zeta - \lambda(\hat{x}_2 - \hat{x}_1)), \hat{x}_1] \cup [\zeta, 1] \\ \bar{a}\frac{\eta}{\eta+1} + \left(1 - \frac{\eta}{\eta+1}\right)\underline{a} & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\ \underline{a} & \text{if } x \in \left[0, \hat{x}_1 - \left(1 - \zeta - \frac{\eta}{\eta+1}(\hat{x}_2 - \hat{x}_1)\right)\right] \cup [\hat{x}_2, \zeta] \end{cases}$$

$$m^*(x) = \begin{cases} \hat{x}_1 + \frac{1}{1 - \frac{\eta}{\eta+1}} x & \text{if } x \in \left[0, \left(1 - \frac{\eta}{\eta+1}\right)(\hat{x}_2 - \hat{x}_1)\right] \\ \hat{x}_2 + x - \left(1 - \frac{\eta}{\eta+1}\right)(\hat{x}_2 - \hat{x}_1) & \text{if } x \in \left[\left(1 - \frac{\eta}{\eta+1}\right)(\hat{x}_2 - \hat{x}_1), \hat{x}_1\right] \\ \hat{x}_2 + \hat{x}_1 - \left(1 - \frac{\eta}{\eta+1}\right)(\hat{x}_2 - \hat{x}_1) + \lambda(x - \hat{x}_1) & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\ 1 & \text{if } x \geq \hat{x}_2 \end{cases}.$$

(iii) If $\eta \in [\eta_2, \eta_3]$. Skills are even less segmented by occupation: there is a set of workers matched with the most skilled managers and using the frontier technology $\bar{a}$ that are more skilled than some managers which use the traditional technology. Specifically there exists constants $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$ such that[73]

$$\omega^*(x) = \begin{cases} 1 & \text{if } x \in [0, \hat{x}_1] \cup [\hat{x}_2, \hat{x}_3] \\ \frac{\eta}{\eta+1} & \text{if } x \in (\hat{x}_1, \hat{x}_2) \\ 0 & \text{if } x > \hat{x}_3 \end{cases}, \qquad \mu^*(x) = \begin{cases} 0 & \text{if } x \in [0, \hat{x}_1] \cup [\hat{x}_2, \hat{x}_3] \\ 1 - \frac{\eta}{\eta+1} & \text{if } x \in (\hat{x}_1, \hat{x}_2) \\ 1 & \text{if } x > \hat{x}_3 \end{cases}$$

$$A^*(x) = \begin{cases} \bar{a} & \text{if } x \in [\hat{x}_2, 1] \\ \frac{\eta}{\eta+1}\bar{a} + \left(1 - \frac{\eta}{\eta+1}\right)\underline{a} & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\ \underline{a} & \text{if } x \in [0, \hat{x}_1] \end{cases}$$

$$m^*(x) = \begin{cases} \hat{x}_1 + \frac{1}{1 - \frac{\eta}{\eta+1}} x & \text{if } x \in \left[0, \left(1 - \frac{\eta}{\eta+1}\right)(\hat{x}_2 - \hat{x}_1)\right] \\ \hat{x}_3 + x - \left(1 - \frac{\eta}{\eta+1}\right)(\hat{x}_2 - \hat{x}_1) & \text{if } x \in \left[\left(1 - \frac{\eta}{\eta+1}\right)(\hat{x}_2 - \hat{x}_1), \hat{x}_1\right] \\ \hat{x}_3 + \hat{x}_1 - \left(1 - \frac{\eta}{\eta+1}\right)(\hat{x}_2 - \hat{x}_1) + \lambda(x - \hat{x}_1) & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\ \hat{x}_3 + \hat{x}_1 - \left(1 - \frac{\eta}{\eta+1}\right)(\hat{x}_2 - \hat{x}_1) + \lambda(\hat{x}_2 - \hat{x}_1) + (x - \hat{x}_2) & \text{if } x \in [\hat{x}_2, \hat{x}_3] \\ 1 & \text{if } x \geq \hat{x}_3 \end{cases}.$$

(iv) If $\eta \in [\eta_3, \eta_4]$. Skills are almost perfectly segregated by technology: most of the individuals which use the traditional technology $\underline{a}$ are less skilled than those that use the frontier one $\bar{a}$, there is however a group of managers that use $\underline{a}$ which are more skilled than the lowest skilled workers among those that use $\bar{a}$. Specifically there exists

[73] *In order to ease notation I am using again $\hat{x}_1$, and $\hat{x}_2$. However notice that these constant are not necessarily identical to the ones defined for the case $\frac{\bar{a}}{\underline{a}} \in [\eta_1, \eta_2]$. In fact in general have a different value. This is true for all constants defined below.*

constants $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$, $\hat{x}_4$ such that

$$\omega^*(x) = \begin{cases} 1 & x \in [0,\hat{x}_1] \cup [\hat{x}_3,\hat{x}_4] \\ \frac{\eta}{\eta+1} & x \in (\hat{x}_2,\hat{x}_3) \\ 0 & x \in \left(\hat{x}_1,\hat{x}_2\right] \cup \left(\hat{x}_4,1\right] \end{cases}, \qquad \mu^*(x) = \begin{cases} 0 & x \in [0,\hat{x}_1] \cup [\hat{x}_3,\hat{x}_4] \\ 1-\frac{\eta}{\eta+1} & x \in (\hat{x}_2,\hat{x}_3) \\ 1 & x \in \left(\hat{x}_1,\hat{x}_2\right] \cup \left(\hat{x}_4,1\right] \end{cases}$$

$$A^*(x) = \begin{cases} \bar{a} & \textit{if } x \in [\hat{x}_3,1] \\ \frac{\eta}{\eta+1}\bar{a} + \left(1-\frac{\eta}{\eta+1}\right)\underline{a} & \textit{if } x \in [\hat{x}_2,\hat{x}_3] \\ \underline{a} & \textit{if } x \in [0,\hat{x}_2] \end{cases}$$

$$m^*(x) = \begin{cases} \hat{x}_1 + x & \textit{if } x \in [0,(\hat{x}_2-\hat{x}_1)] \\ \hat{x}_2 + \frac{1}{(1-\lambda)}(x-(\hat{x}_2-\hat{x}_1)) & \textit{if } x \in [(\hat{x}_2-\hat{x}_1),\hat{x}_1] \\ \hat{x}_3 & \textit{if } x \in [\hat{x}_1,\hat{x}_2] \\ \hat{x}_4 + \lambda(x-\hat{x}_2) & \textit{if } x \in [\hat{x}_2,\hat{x}_3] \\ \hat{x}_4 + \lambda(\hat{x}_3-\hat{x}_2) + x - \hat{x}_3 & \textit{if } x \in [\hat{x}_3,\hat{x}_4] \\ 1 & \textit{if } x \geq \hat{x}_4 \end{cases}.$$

(v) If $\eta \geq \eta_4$. Skills are segregated by technology: low skilled individuals use the traditional technology $\underline{a}$, and high skilled ones use the frontier technology $\bar{a}$. Among the individuals that use $\underline{a}$, the most skilled are managers. Similarly, among the individuals that use $\bar{a}$, the most skilled are managers. Specifically, there exists constants $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$ such that

$$\omega^*(x) = \begin{cases} 1 & x \in [0,\hat{x}_1] \cup [\hat{x}_2,\hat{x}_3] \\ 0 & x \in (\hat{x}_1,\hat{x}_2) \cup \left(\hat{x}_3,1\right] \end{cases}, \qquad \mu^*(x) = \begin{cases} 0 & x \in [0,\hat{x}_1] \cup [\hat{x}_2,\hat{x}_3] \\ 1 & x \in (\hat{x}_1,\hat{x}_2) \cup \left(\hat{x}_3,1\right] \end{cases}$$

$$A^*(x) = \begin{cases} \bar{a} & \textit{if } x \in [\hat{x}_2,1] \\ \underline{a} & \textit{if } x \in [0,\hat{x}_2] \end{cases}$$

$$m^*(x) = \begin{cases} \hat{x}_1 + x & \textit{if } x \in [0,\hat{x}_1] \\ \hat{x}_2 & \textit{if } x \in [\hat{x}_1,\hat{x}_2] \\ \hat{x}_3 + (x-\hat{x}_2) & \textit{if } x \in [\hat{x}_2,\hat{x}_3] \\ 1 & \textit{if } x \geq \hat{x}_3 \end{cases}.$$

Proof. Let's first consider the optimal technology choice. Consider any team (x',x). Given the step-wise linearity of the cost function and the linearity in a of the marginal return from technology, that are given by $x'(1+\gamma x)$, a team picks either $\bar{a}$ or $\underline{a}$. Hence a team picks $\bar{a}$ if and only if

$$\begin{aligned} \bar{a}x'(1+\gamma x) - \zeta(\bar{a}-\underline{a})(1+\gamma x) &> \underline{a}x'(1+\gamma x) \\ x' &> \zeta \end{aligned}$$

this proves Lemma 5. Also, this gives the usual function $v(x',x;\bar{a},\underline{a})$ that takes into consideration optimal technology choice and that is given by

$$v\left(x',x;\bar{a},\underline{a}\right)=\begin{cases}\left(\bar{a}\left(x'-\zeta\right)+\underline{a}\zeta\right)\left(1+\gamma x\right) & \text{if } x'\geq\zeta\\ \underline{a}x'\left(1+\gamma x\right) & \text{if } x'<\zeta\end{cases}$$

Next, let me prove that the problem is homogeneous. Using Lemma 1, total output is given by

$$Y=\int\limits_{\mathbb{X}}v\left(m\left(x\right),x;\bar{a},\underline{a}\right)\omega\left(x\right)dx=\underline{a}\int\limits_{\mathbb{X}}v\left(m\left(x\right),x;\eta,1\right)\omega\left(x\right)dx$$

where the first equality comes from the definition of total output, and the second one from the functional form of v shown above. Lemma 6 follows immediately. To ease notation I next drop the explicit dependence of $v\left(x',x;\eta,1\right)$ on η.

I now prove the main part of the proposition. Let's consider the marginal of v for types x (either when workers or managers), while matched with their corresponding partners

$$\begin{aligned}v_1\left(x,w\left(x\right)\right) &= \begin{cases}\eta\left(1+\gamma w\left(x\right)\right) & \text{if } x\geq\zeta\\ \left(1+\gamma w\left(x\right)\right) & \text{if } x<\zeta\end{cases}\\ v_2\left(m\left(x\right),x\right) &= \begin{cases}\gamma\left(\eta\left(m\left(x\right)-\zeta\right)+\zeta\right) & \text{if } m\left(x\right)\geq\zeta\\ \gamma m\left(x\right) & \text{if } m\left(x\right)<\zeta\end{cases}\end{aligned}$$

also notice that

$$\begin{aligned}v_{12}\left(x,w\left(x\right)\right) &= \begin{cases}\gamma\eta & \text{if } x\geq\zeta\\ \gamma & \text{if } x<\zeta\end{cases}\\ v_{12}\left(m\left(x\right),x\right) &= \begin{cases}\gamma\eta & \text{if } m\left(x\right)\geq\zeta\\ \gamma & \text{if } m\left(x\right)<\zeta\end{cases}\end{aligned}$$

as a result we get that, since $m\left(x\right),w\left(x\right),x\in\left[0,1\right]$

$$\begin{aligned}v_1\left(x,w\left(x\right)\right) &> v_2\left(m\left(x\right),x\right)\ \text{if } x\geq\zeta\\ v_1\left(x,w\left(x\right)\right) &> v_2\left(m\left(x\right),x\right)\ \text{if } m\left(x\right)<\zeta\\ v_1\left(x,w\left(x\right)\right) &< v_2\left(m\left(x\right),x\right)\ \text{if } x<\zeta\text{and } m\left(x\right)>\kappa\left(\eta,w\left(x\right)\right)\end{aligned}$$

where

$$\kappa\left(\eta,w\left(x\right)\right)=\zeta+\frac{1+\gamma w\left(x\right)-\zeta\gamma}{\gamma\eta}$$

so that $\lim_{\eta\rightarrow\infty}\kappa\left(\eta,w\left(x\right)\right)=\zeta$ and $\kappa_1<0$. Next, we can use the first order conditions and the above marginal values to characterize the shape that the optimal allocation must have. As mentioned, with shape I refer to the number of cutoff in which occupation change. First, let me recall that the necessary first order conditions are

1. $\forall x$ s.t. $\omega^*\left(x\right)>0$ and $\mu^*\left(x\right)>0$ then $v_1\left(x,w^*\left(x\right)\right)=v_2\left(m^*\left(x\right),x\right)$;
2. $\forall x$ s.t. $\mu^*\left(x-\epsilon\right)=0$ and $\mu^*\left(x\right)>0$ then $v_1\left(x,w^*\left(x\right)\right)\geq v_2\left(m^*\left(x\right),x\right)$;
3. $\forall x$ s.t. $\mu^*\left(x-\epsilon\right)>0$ and $\mu^*\left(x\right)=0$ then $v_1\left(x,w^*\left(x\right)\right)\leq v_2\left(m^*\left(x\right),x\right)$.

Next, I show two Lemmas that highlights two properties that must hold in order for the necessary conditions to be satisfied. These are useful for the rest of the proof.

Lemma A.7.1. In the optimal allocation, within the interval $x \in [\zeta, 1]$, there cannot be a worker more skilled than a manager. This result holds immediately due to the fact that $v_1(x, w(x)) > v_2(m(x), x)$ if $x \geq \zeta$.

Lemma A.7.1. In the optimal allocation, there cannot exists (x_1, x_2, x_3) such that $x_1 < x_2 < x_3 < \zeta$ and $\mu^*(x_1) = 1$, $\omega^*(x_2) = 1$, and $\mu^*(x_3) = 1$. Therefore in the optimal allocation there cannot be a set of workers in between two sets of managers, as long as the lowest skill of the most skilled set of managers is less skilled than ζ. Let me now prove this result. In order for the one described to be an optimal allocation, it must be that there exist (x_1', x_2'), with $x_1 \leq x_1' \leq x_2 \leq x_2' \leq x_3$ such that $v_1(x_1', w^*(x_1')) \leq v_2(m^*(x_1'), x_1')$ and $v_1(x_2', w^*(x_2')) \geq v_2(m^*(x_2'), x_2')$. But this two inequalities cannot hold simultaneously according to the definition of v_1, v_2, and v_{12}. In fact, in order for $v_1(x_1', w^*(x_1')) \leq v_2(m^*(x_1'), x_1')$ to hold, it must be that $m^*(x_1') > \zeta$, but then since $x_2' < \zeta$, $v_{12}(m^*(x), x) > v_{12}(x, w^*(x))$ $\forall x \in [x_1', x_2']$ and thus $v_1(x_1', w^*(x_1')) < v_2(m^*(x_1'), x_1')$, contradicting the second inequality. This result turns out to be extremely useful in the characterization, because it limits the number of cases that we have to consider.

Next, I characterize each case as η increases.

Case 1: $\eta \in [1, \eta_1]$. Consider first $\eta = 1$. It is simple to verify that this implies that $v_1(x, w(x)) > v_2(m(x), x)$ for any m. Hence, according to the conditions above, there cannot be a worker more skilled than a manager, since that would require that there exist an x such that $v_1(x, w(x)) < v_2(m(x), x)$. Next, let's consider η_1 that solves

$$\kappa(\eta_1, 0) = 1.$$

For any $\eta < \eta_1$, $\kappa(\eta, 0) > 1$, hence over the relevant support, $v_1(x, w(x)) > v_2(m(x), x)$. Therefore we have shown that talent is segmented by occupation for any $\eta \in [1, \eta_1]$. The used technology and the matching function follows directly from the characterization of the shape of the optimal assignment, hence I discuss them only when some properties are useful to be emphasized.

Case 2: $\eta \in \left(\eta_1, \eta_2\right]$, where η_1 is as previously defined and η_2 is such that $\kappa(\eta_2, \hat{x}_0) = 1$, where $\hat{x}_0$ is defined as the type such that $m^*(\hat{x}_0) = \zeta$. For this values of η the optimal allocation is given by

$$\omega^*(x) = \begin{cases} 1 & \textit{if } x < \hat{x}_1 \\ \lambda & \textit{if } x \in [\hat{x}_1, \hat{x}_2], \\ 0 & \textit{if } x > \hat{x}_2 \end{cases} \qquad \mu^*(x) = \begin{cases} 0 & \textit{if } x < \hat{x}_1 \\ 1-\lambda & \textit{if } x \in [\hat{x}_1, \hat{x}_2] \\ 1 & \textit{if } x > \hat{x}_2 \end{cases}$$

where $\hat{x}_1$, $\hat{x}_2$, and λ must satisfy - by the first order conditions -

$$v_1(x, w(x)) = v_2(m(x), x) \;\; \forall x \in [\hat{x}_1, \hat{x}_1]$$

and since by market clearing $m(x) = \hat{x}_1 + \lambda x$ $\forall x \in [\hat{x}_1, \hat{x}_2]$ and $w(x) = 0 + (1-\lambda)x$, then we can rewrite the previous equality as

$$(1 + \gamma(1-\lambda)x) = \gamma(\eta(\hat{x}_1 + \lambda x - \zeta) + \zeta) \;\; \forall x \in [\hat{x}_1, \hat{x}_2]$$

which immediately implies that

$$\lambda = \frac{\eta}{\eta + 1}.$$

λ gives the fraction of managers and workers such that the equality between the marginal product is satisfied.

Notice that this allocation satisfies, by construction, the necessary conditions for optimality. Also, notice that no other allocation might satisfy them. Due to Lemma A.7.2, the only other alternative allocation would be to have a set of only workers that must expand until $x > \zeta$. However, in order for that to be optimal, it should be that $v_1(\zeta, \hat{x}_0) \leq v_2(1, \zeta)$ but this is not satisfied, due to the fact that $\eta \leq \eta_2$.[74]

Case 3: $\eta \in \left(\eta_2, \eta_3\right]$, where η_3 is defined such that $\kappa(\eta_3, 0) = \hat{x}_3$. With this parameter value, the previous allocation is not optimal anymore. The reason being that for $\eta > \eta_2$ at ζ

$$v_2(1, \zeta) > v_1(\zeta, \hat{x}_0)$$

with $\hat{x}_0$ defined above. In fact the optimal allocation is given by

$$\omega^*(x) = \begin{cases} 1 & \text{if } x \in [0, \hat{x}_1] \cup [\hat{x}_2, \hat{x}_3] \\ \lambda & \text{if } x \in (\hat{x}_1, \hat{x}_2) \\ 0 & \text{if } x > \hat{x}_3 \end{cases}, \qquad \mu^*(x) = \begin{cases} 0 & \text{if } x \in [0, \hat{x}_1] \cup [\hat{x}_2, \hat{x}_3] \\ 1-\lambda & \text{if } x \in (\hat{x}_1, \hat{x}_2) \\ 1 & \text{if } x > \hat{x}_3 \end{cases}$$

where λ as before is set to guarantee

$$v_1(x, w(x)) = v_2(m(x), x) \;\; \forall x \in [\hat{x}_1, \hat{x}_1],$$

hence $\lambda = \frac{\eta}{\eta+1}$ and additionally $\hat{x}_3 > \zeta$, or otherwise Lemma A.7.1 would be violated. This allocation satisfies by construction the necessary conditions. The only other allocation that could satisfy the necessary conditions would be one in which there is a set of managers before the mixed region $(\hat{x}_1, \hat{x}_2)$. The condition that $\kappa(\eta_3, 0) = \hat{x}_3$ prevents this to be optimal. In fact, for this to be the case, it should be that for some x

$$v_2(\hat{x}_3, x) > v_1(x, 0),$$

but this is excluded due to the fact that $\eta \leq \eta_3$.

Case 4: $\eta \in \left(\eta_3, \eta_4\right]$, where η_4 solves $\kappa(\eta_4, \hat{x}_1) = \hat{x}_3$. Now consider a marginal increase in η, then the previously described deviation (a set of managers before the mixed region $(\hat{x}_1, \hat{x}_2)$) becomes optimal. Hence the optimal allocation is described by

$$\omega^*(x) = \begin{cases} 1 & x \in [0, \hat{x}_1] \cup [\hat{x}_3, \hat{x}_4] \\ \lambda & x \in (\hat{x}_2, \hat{x}_3) \\ 0 & x \in \left(\hat{x}_1, \hat{x}_2\right] \cup \left(\hat{x}_4, 1\right] \end{cases}, \qquad \mu^*(x) = \begin{cases} 0 & x \in [0, \hat{x}_1] \cup [\hat{x}_3, \hat{x}_4] \\ 1-\lambda & x \in (\hat{x}_2, \hat{x}_3) \\ 1 & x \in \left(\hat{x}_1, \hat{x}_2\right] \cup \left(\hat{x}_4, 1\right] \end{cases}$$

where as usual $\lambda = \frac{\eta}{\eta+1}$. Additionally, notice that this allocation is optimal as long as $\hat{x}_2 < \hat{x}_3$. In fact the only possible deviation that could satisfy the necessary conditions would be to have an empty region of mixing. This is

[74] I can also show how to solve explicitly for $\hat{x}_0$ and η_2, so that they are not expressed in terms of endogenous objects. We do this, by using market clearing, that gives

$$(1 - \hat{x}_0) + (\zeta - \hat{x}_0)\left(\frac{\eta_2}{\eta_2 + 1}\right) = \frac{1}{2}$$

and the definition of κ itself.

$$\zeta + \frac{1 + \gamma\hat{x}_0 - \zeta\gamma}{\gamma\eta_2} = 1.$$

not optimal as long as

$$v_1(\hat{x}_1, w(\hat{x}_1)) \geq v_2(\hat{x}_3, \hat{x}_1)$$

which is satisfied by the fact that $\eta \geq \eta_4$.

Case 5: $\eta \in (\eta_4, \infty)$. Now, the optimal allocation is given by

$$\omega^*(x) = \begin{cases} 1 & x \in [0, \hat{x}_1] \cup [\hat{x}_2, \hat{x}_3] \\ 0 & x \in (\hat{x}_1, \hat{x}_2) \cup \left(\hat{x}_3, 1\right] \end{cases}, \qquad \mu^*(x) = \begin{cases} 0 & x \in [0, \hat{x}_1] \cup [\hat{x}_2, \hat{x}_3] \\ 1 & x \in (\hat{x}_1, \hat{x}_2) \cup \left(\hat{x}_3, 1\right] \end{cases}$$

This hold immediately from the fact that $\eta > \eta_4$, hence the previously discussed deviation would be optimal and thus there cannot be any mixing region. Usual arguments show that this is the only possible allocation that satisfies the necessary conditions and conclude the proof. □

Remark (Solving for the Cutoffs $\hat{x}_s$) Given the equilibrium shape, as shown in the previous proposition, we can solve for the optimal cutoffs by rewriting the planner problem incorporating the shape described and then simply taking the first order condition with respect to the cutoffs. As an illustrative example, consider the Case 5, $\eta \in (\eta_4, \infty)$.

The planner problem becomes

$$\max_{\hat{x}_1, \hat{x}_2, \hat{x}_3} \int_0^{\hat{x}_1} v(m^*(x), x)\, dx + \int_{\hat{x}_2}^{\hat{x}_3} v(m^*(x), x)\, dx$$

subject to the market clearing constraints

$$\begin{aligned} \hat{x}_2 - \hat{x}_1 &= \hat{x}_1 \\ 1 - \hat{x}_3 &= \hat{x}_3 - \hat{x}_2 \end{aligned}$$

where

$$m^*(x) = \begin{cases} \hat{x}_1 + x & \text{if } x \in [0, \hat{x}_1] \\ \hat{x}_2 & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\ \hat{x}_3 + (x - \hat{x}_2) & \text{if } x \in [\hat{x}_2, \hat{x}_3] \\ 1 & \text{if } x \geq \hat{x}_3 \end{cases}$$

and

$$\alpha^*(x) = \begin{cases} \bar{a} & \text{if } x \in [\hat{x}_2, 1] \\ \underline{a} & \text{if } x \in [0, \hat{x}_2] \end{cases}$$

We can then just take the first order conditions using Leibnitz lemma to find a system of three equations in three unknowns that can be solved numerically.

Proof of Proposition 3 First, let's use Proposition 3.b to calculate $\pi(\eta)$ for Case 1 and Case 5. For Case 1, which corresponds to talent segmentation, it is given by

$$\begin{aligned} \pi(\eta) &= \frac{\frac{(1-\zeta)}{2} - \frac{\left(\zeta + \frac{1}{2}\right)}{2}}{\frac{1}{2}} \\ &= \frac{1}{2}. \end{aligned}$$

For Case 5, which corresponds to talent segregation, it is given by

$$\pi(\eta) = 1$$

since $A^*(x) = A_p^*(x)$.

Next, I show that $\pi(\eta)$ is non-decreasing in η for values of $\eta \in [\eta_1, \eta_4]$. Let $\varphi^*(\eta')$ be the optimal allocation for η', with associated $\pi(\eta')$. Next, assume that exists $\eta > \eta'$ such that the optimal allocation $\varphi^*(\eta)$ is such that $\pi(\eta) < \pi(\eta')$. I show that this would lead to a contradiction. Total output in the allocation $\varphi^*(\eta')$ can be rewritten as

$$\underline{a}\left[\eta' \int\limits_{\Omega(\bar{a};\eta')} v(m^*(x;\eta'), x)\,\omega^*(x;\eta')\,dx + \int\limits_{\Omega(\underline{a};\eta')} v(m^*(x;\eta'), x)\,\omega^*(x;\eta')\,dx\right]$$

where

$$\begin{aligned} \Omega(\bar{a};\eta') &\equiv \{x : \alpha^*(m^*(x;\eta'), x) = \bar{a}\} \\ \Omega(\underline{a};\eta') &\equiv \{x : \alpha^*(m^*(x;\eta'), x) = \underline{a}\} \end{aligned}$$

next let's define

$$\begin{aligned} B(\eta') &\equiv \int\limits_{\Omega(\bar{a};\eta')} v(m^*(x;\eta'), x)\,\omega^*(x;\eta')\,dx \\ C(\eta') &\equiv \int\limits_{\Omega(\bar{a};\eta')} v(m^*(x;\eta'), x)\,\omega^*(x;\eta')\,dx \end{aligned}$$

thus the output of the allocation $\varphi^*(\eta)$ can be written as

$$\underline{a}\,[\eta B(\eta) + C(\eta)] \,.$$

Next, $\pi(\eta) < \pi(\eta')$ implies that

$$\begin{aligned} B(\eta') &> B(\eta) \\ C(\eta') &< C(\eta) \end{aligned}$$

then optimality of $\varphi^*(\eta')$ and $\varphi^*(\eta)$ requires respectively that

$$\begin{aligned} \eta' B(\eta') + C(\eta') &> B(\eta) + C(\eta) \\ \eta B(\eta) + C(\eta) &> B(\eta') + C(\eta') \end{aligned}$$

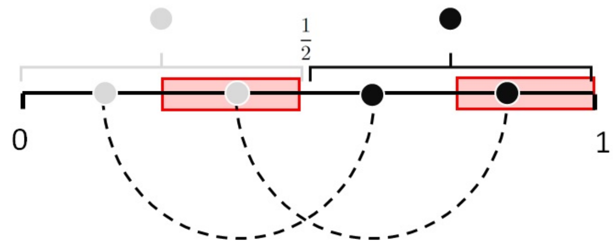
which thus gives

$$\begin{aligned} \eta' &> \frac{C(\eta) - C(\eta')}{B(\eta) - B(\eta')} \\ \eta &< \frac{C(\eta) - C(\eta')}{B(\eta) - B(\eta')} \end{aligned}$$

which contradicts the fact that $\eta > \eta'$ and shows that $\pi(\eta)$ cannot be decreasing in η.

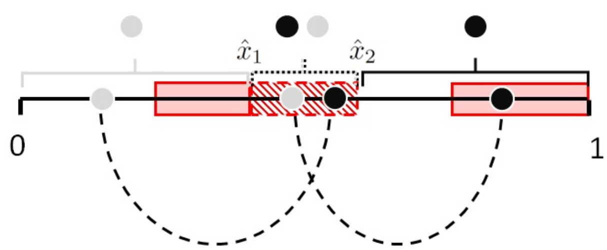
Last, notice that $\forall\eta \in (\eta_1, \eta_4)$, as shown in Proposition 3.b talent is nor segmented nor segregated, which gives $\pi(\eta) \in \left(\frac{1}{2}, 1\right)$ $\forall\eta \in (\eta_1, \eta_4)$ and concludes the proof. $\square$

Figure A.15: Quasi-Linear Model: Segmentation $\frac{\bar{a}}{\underline{a}} \leq \eta_1$



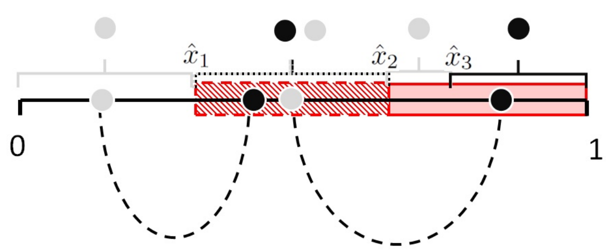
Notes: I plot the occupational choice for each individual. The squared brackets put together individuals with the same occupation. If the square brackets is grey with a grey dot on top of it means that individuals in that interval are workers. If instead is black it means that are managers. Dotted square brackets with one grey and one black dot on top are regions in which there is mixing, to the extent that some individuals are managers and some others workers. The red regions highlight the individuals using the frontier technology $\bar{a}$. The red striped regions are present in mixing area on which the workers use $\bar{a}$, while the managers use $\underline{a}$. A grey dot represents a worker, while a black dot represents a manager. Dotted lines connect a worker with his manager.

Figure A.16: Quasi-Linear Model: $\frac{\bar{a}}{\underline{a}} \in [\eta_1, \eta_2]$



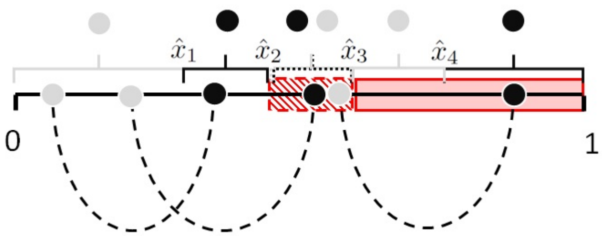
Notes: See Figure A.15.

Figure A.17: Quasi-Linear Model: $\frac{\bar{a}}{\underline{a}} \in [\eta_2, \eta_3]$



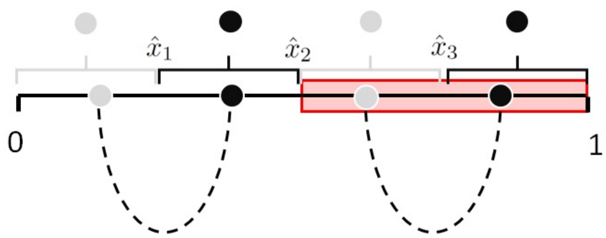
Notes: See Figure A.15.

Figure A.18: Quasi-Linear Model: $\frac{\bar{a}}{\underline{a}} \in [\eta_3, \eta_4]$



Notes: See Figure A.15.

Figure A.19: Quasi-Linear Model: Segregation, $\frac{\bar{a}}{\underline{a}} \geq \eta_4$



Notes: See Figure A.15.

A.9 Details on Self-Employment

I here define the planner problem with self-employment and provide a proof of Proposition 4. I also discuss how the mass of self-employed individuals changes as η varies.

First let me summarize the assumptions on the environment. I work with the case $\mathbb{X} = [0,1]$, and the quasi-linear model. I define the production function of a self-employed of ability x with technology a as $g_s(x,a)$. The next assumption summarizes the environment

Assumption A.9.1 (Self-Employment Model) *For $\gamma < \frac{1}{2}$ and $\zeta \in \left(\frac{1}{2} + \frac{\gamma+\sqrt{\gamma^2+\gamma(2-\gamma)}}{4-2\gamma}, 1\right)$*

$$
\begin{aligned}
\mathbb{X} &= [0,1] \\
f(x',x) &= x'(2-\gamma+\gamma x) \\
c(x',x,a;\underline{a},\bar{a}) &= \tilde{c}(a;\underline{a},\bar{a})(2-\gamma+\gamma x) \\
\tilde{c}(a;\underline{a},\bar{a}) &= \begin{cases} 0 & \text{if } a < \underline{a} \\ \zeta(a-\underline{a}) & \text{if } a \in [\underline{a},\bar{a}] \\ \infty & \text{if } a > \bar{a} \end{cases}
\end{aligned}
$$

$$
g_s(x,a) = ax - \tilde{c}(a;\underline{a},\bar{a}).
$$

Notice that the functional form for f and the conditions on γ and ζ are changed with respect to the quasi-linear model in the main text. These functional forms guarantee that in the optimal allocation not everyone is self-employed, but at least someone is. Additionally, they guarantee that at least some manager use the traditional technology $\underline{a}$.

Definition A.9.1 (Allocation). *An allocation φ is a septuple of functions $\{\tilde{\omega}, \tilde{\mu}, \tilde{\sigma}, \tilde{m}, \tilde{w}, \alpha, \alpha_s\}$, where (i) $\tilde{\omega} : \mathbb{X} \times \mathbb{I} \rightarrow \{0,1\}$ is a function that takes value one if an individual is a worker; (i) $\tilde{\mu} : \mathbb{X} \times \mathbb{I} \rightarrow \{0,1\}$ is a function that takes value one if an individual is a manager; (iii) $\tilde{\sigma} : \mathbb{X} \times \mathbb{I} \rightarrow \{0,1\}$ is a function that takes value one if an individual is a self-employed; (iv) $\tilde{m} : \mathbb{W} \rightarrow \mathbb{M}$ assigns each individual in the set $\mathbb{W}$, that is closure of the set of workers $\{(x,i) : \tilde{\omega}(x,i) = 1\}$, to a partner in the set $\mathbb{M}$, that is the closure of the set of managers $\{(x,i) : \tilde{\mu}(x,i) = 1\}$; (v) $\tilde{w} : \mathbb{M} \rightarrow \mathbb{W}$ assigns to each manager a corresponding worker; (vi) $\alpha : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{A}$ assigns to each possible pair of types a production technology within the available set $\mathbb{A}$; and (vii) $\alpha_s : \mathbb{X} \rightarrow \mathbb{A}$ assigns to each self-employed type a production technology.*

The set $\mathbb{A}$ is the set of available technologies, which, if not otherwise specified, is equal to the set of positive real numbers.

Definition A.9.2 (Feasible Allocation). *An allocation φ is feasible if (i) $\tilde{\mu}(x,i) + \tilde{\omega}(x,i) + \tilde{\sigma}(x,i) = 1$ $\forall (x,i) \in \mathbb{X} \times \mathbb{I}$; and (ii) $\hat{m}$ and $\hat{w}$ are measure preserving bijection; (iii) $m(x,i) = (x',i')$ implies $w(x',i') = (x,i)$ and $w(x,i) = (x',i')$ implies $m(x',i') = (x,i)$ $\forall (x,i) \in \mathbb{X} \times \mathbb{I}$. I call $\mathbb{F}$ the set of all feasible allocations.*

It is also useful to define few auxiliary functions which I use in the definition of the planner's problem.

Definition A.9.3 (Occupational Choice Functions). *Define (i) $\omega(x) = \int_{\mathbb{I}} \tilde{\omega}(x,i)\, di$ the function that describes the fractions of individuals of type x who are workers; (i) $\sigma(x) = \int_{\mathbb{I}} \tilde{\sigma}(x,i)\, di$ the function that describes the fractions of individuals of type x who are self-employed; and (ii) $\mu(x) = 1 - \omega(x) - \sigma(x)$ the function that describes the fractions of individuals of type x who are managers.*

I call ω and μ the occupational choice functions.

Definition A.9.4 (Matching Functions). *I define (i) $m_x : \mathbb{W} \rightarrow \mathbb{X}$ to be the function that assigns to each worker the type of his manager, hence if $\tilde{m}(x,i) = (x',i')$ then $m_x(x,i) = x'$; (ii) $m : \mathbb{X} \rightarrow \mathbb{X}$ to be the function that assigns to each type of workers the average type of its managers, hence $m(x) = \int_{\mathbb{I}} m_x(x,i)\, di$ if $\omega(x) > 0$ and $m(x) = \int_{\mathbb{I}} m_x(\hat{x},i)\, di$ where $\hat{x} = \max\{x' \in \mathbb{X} : x' < x \text{ and } \omega(x) > 0\}$ otherwise; (iii) $w_x : \mathbb{M} \rightarrow \mathbb{X}$ to be the function that assigns to each manager the type of his worker, hence if $\tilde{w}(x,i) = (x',i')$ then $w_x(x,i) = x'$; (ii) $w : \mathbb{X} \rightarrow \mathbb{X}$ to be the function that assigns to each type of manager the average type of its workers, hence $w(x) = \int_{\mathbb{I}} w_x(x,i)\, di$ if $\mu(x) > 0$ and $w(x) = \int_{\mathbb{I}} w_x(\hat{x},i)\, di$ where $\hat{x} = \max\{x' \in \mathbb{X} : x' < x \text{ and } \mu(x) > 0\}$ otherwise.*

Planner's Problem. The planner picks the feasible allocation which maximizes total output:

$$\max_{\varphi \in \mathbb{F}} Y(\varphi)$$

where output is given by

$$Y(\varphi) = \int_{\mathbb{X}} \int_{\mathbb{I}} g(m_x(x,i), x, \alpha(m_x(x,i), x))\, \tilde{\omega}(x,i)\, dxdi + \int_{\mathbb{X}} \int_{\mathbb{I}} g_s(x, \alpha_s(x))\, \tilde{\sigma}(x,i)\, dxdi.$$

The only difference with respect to the quasi-linear problem is the additional term that sums the output produced by self-employed.

For brevity, I omit the complete characterization of the problem, which follows the same steps of the one for the quasi-linear problem shown in Proposition 3.b. It is available upon request. I nonetheless provide a proof of Proposition 4 that does not require the complete characterization.

Proof of Proposition 4. Lemmas 0 and 2-3 holds trivially for this setting as well. Also, the introduction of self-employment does not affect the homogeneity in $\underline{a}$ (Lemma 6). As a preliminary step, let me define, given the new functional forms, the output of a team (x', x) and the one of a self-employed that takes into account optimal technology. They are

$$v(x', x) = \begin{cases} (\eta(x'-\zeta)+\zeta)(2-\gamma+\gamma x) & \text{if } x' \geq \zeta \\ x'(2-\gamma+\gamma x) & \text{if } x' < \zeta \end{cases}$$

$$v_s(x) = \begin{cases} \eta(x-\zeta)+\zeta & \text{if } x \geq \zeta \\ x & \text{if } x < \zeta \end{cases}$$

where I have already used the reformulation of the problem in terms of η, that is possible due to homogeneity. The marginal products are given by

$$v_1(x', x) = \begin{cases} \eta(2-\gamma+\gamma x) & \text{if } x' \geq \zeta \\ 1(2-\gamma+\gamma x) & \text{if } x' < \zeta \end{cases}$$

$$v_2(x', x) = \begin{cases} \gamma(\eta(x'-\zeta)+\zeta) & \text{if } x' \geq \zeta \\ \gamma x' & \text{if } x' < \zeta \end{cases}$$

$$v_s'(x) = \begin{cases} \eta & \text{if } x \geq \zeta \\ 1 & \text{if } x < \zeta \end{cases}$$

I next write the necessary first order conditions. They are in the same spirit as the ones of the model without self-employment, but with additional conditions resulting from the choice of σ^*.

1. $\forall x$ s.t. $\omega^*(x) > 0$ and $\mu^*(x) > 0$ then $v_1(x, w^*(x)) = v_2(m^*(x), x)$;
2. $\forall x$ s.t. $\omega^*(x) > 0$ and $\sigma^*(x) > 0$ then $v_s'(x) = v_2(m^*(x), x)$;
3. $\forall x$ s.t. $\mu^*(x) > 0$ and $\sigma^*(x) > 0$ then $v_s'(x) = v_1(x, w^*(x))$;
4. $\forall x$ s.t. $\omega^*(x-\epsilon) = 0$; $\mu^*(x-\epsilon) > 0$; and $\omega^*(x) > 0$ then $v_1(x, w^*(x)) \leq v_2(m^*(x), x)$;
5. $\forall x$ s.t. $\omega^*(x-\epsilon) = 0$; $\sigma^*(x-\epsilon) > 0$; and $\omega^*(x) > 0$ then $v_s'(x) \leq v_2(m^*(x), x)$;
6. $\forall x$ s.t. $\mu^*(x-\epsilon) = 0$; $\omega^*(x-\epsilon) > 0$; and $\mu^*(x) > 0$ then $v_2(m^*(x), x) \leq v_1(x, w^*(x))$;
7. $\forall x$ s.t. $\mu^*(x-\epsilon) = 0$; $\sigma^*(x-\epsilon) > 0$; and $\mu^*(x) > 0$ then $v_s'(x) \leq v_1(x, w^*(x))$;
8. $\forall x$ s.t. $\sigma^*(x-\epsilon) = 0$; $\omega^*(x-\epsilon) > 0$; and $\sigma^*(x) > 0$ then $v_1(x, w^*(x)) \leq v_s'(x)$;
9. $\forall x$ s.t. $\sigma^*(x-\epsilon) = 0$; $\mu^*(x-\epsilon) > 0$; and $\sigma^*(x) > 0$ then $v_2(m^*(x), x) \leq v_s'(x)$;

Using condition (9) and the fact that, given the chosen functional form, $v_s'(x) < v_1(x, w^*(x))$ $\forall x$, we get that in the optimal allocation the most skilled self-employed must be less skilled than the least skilled manager. As a result, since we work with parameter values (high enough ζ) such that at least some managers use the low technology $\underline{a}$, we get that self-employed use only $\underline{a}$. Therefore no self-employed is more skilled than ζ, and this implies that for all self-employed $v_s'(x) = 1$. This in turn, using again the first order conditions, implies that the set of self-employed is connected, and that $\sigma^*(x) \in \{0,1\}$ almost everywhere. Then, in order to calculate the expected ability of self-employed is sufficient to calculate the difference between the mass of individuals which are more skilled than self-employed and those that are less skilled than them. In fact, let $\sigma_{max} = \max_x \{x: \ \sigma^*(x) = 1\}$ and $\sigma_{min} = \min_x \{x: \ \sigma^*(x) = 1\}$, then the average ability of self-employed is given by $\vartheta = \frac{\sigma_{max}+\sigma_{min}}{2}$. Where the set $[\sigma_{max}, 1]$ is the set of individuals more skilled than self-employed and $[0, \sigma_{min}]$ the set of types less skilled. They are respectively of masses (defined h and l) $h = (1-\sigma_{max})$ and $l = \sigma_{min}$. We can then rewrite $\vartheta = \frac{1-(h-l)}{2}$, that shows that ϑ is decreasing in $(h-l)$. Next, let's notice that since $v_2(m^*(x), x)$ is increasing in x, while $v_s'(x)$ is constant, it must be that

$$\begin{aligned} h &= \int_0^1 \mu^*(x)\, dx + \int_{x:\, v_2(m^*(x),x)=v_s'(x)}^1 \omega^*(x)\, dx \\ l &= \int_0^{x:\, v_2(m^*(x),x)=v_s'(x)} \omega^*(x)\, dx, \end{aligned}$$

and then by market clearing we get

$$h - l = 2 \int_{x:\, v_2(m^*(x),x)=v_s'(x)}^1 \omega^*(x)\, dx.$$

Last, we need to show that $2\int_{x:\, v_2(m^*(x),x)=v_s'(x)}^1 \omega^*(x)\, dx$ increases in η. Notice that this is the set of workers that is matched with managers more skilled than ζ and such that $v_2(m^*(x), x) \geq v_s'(x)$. This last inequality can be rewritten, using the functional forms, as

$$m(x) \geq \zeta + \frac{1-\zeta\gamma}{\gamma\eta},$$

we know that the set of manager more skilled than ζ is connected as well (again using the first order conditions), as a result

$$\int_{x:\, v_2(m^*(x),x)=v_s'(x)}^1 \omega^*(x)\, dx = \left[\min\left\{1, \zeta + \frac{1-\zeta\gamma}{\gamma\eta}\right\}, 1\right]$$

which shows that the selection of self-employed is equal to 0 for low enough η, and monotonically decreasing after that. This thus concludes the proof. □

Remark: Number of Self-Employed It may also be interesting to ask what are the determinants of the number of individuals which are optimally allocated to self-employment, and how it changes as a function of distance from the frontier. Let's consider two individuals x and x', with $x' > x$. It is optimal to organize them in a pair if and only if

$$g(x', x\alpha^*(x')) > g_s(x', \alpha^*(x')) + g_s(x, \alpha^*(x))$$

where I've used the fact that in the quasi-linear setting the optimal technology depends only on the type of the manager. From the expression above we can notice that forming a pair allows to leverage the technology of the most skilled individuals, since both workers then use $\alpha^*(x')$. However there is also a cost to form a pair, since - due to the chosen functional forms - $g(x, x, a) < g_s(x, a) + g_s(x, a)$. The result is that it is never optimal to form a pair with two identical individuals, but they would rather work both as self-employed. Intuitively, a firm is optimal only to the extent that the cost in terms of transaction costs within the firms is compensated by the fact that the manager can leverage his higher talent. Next, with this trade-off in mind we can explore the role of distance to the frontier.

When a country is farther from the frontier, there is more scope to leverage the talent, since the gap between the frontier and the traditional technology is higher, this reduces the number of self-employed. However, if we focus within the traditional sector, then we are left with individuals which use the same technology, and which are more similar between each other, since all the high skilled ones are allocated to the modern sector. The compression of the skill distribution in the traditional sector increases the number of self-employed. The overall effect depends on parameter values, but let me stress that the model is able to replicate, as in the example shown in figures 4a and 4b, the fact that in countries far from the frontier we observe a very large number of low skilled self-employed.

A.10 Model Extension with Endogenous Education

I here show, in a simple environment, that the relationship between the convexity of cost of technology and the pattern of matching is robust to the inclusion of an endogenous education choice. Additionally, allowing individuals to choose an optimal investment in education shows that the relationship between education and ability is positive, as assumed in the empirical application, and provides additional implications that are supported in the data.

I make the following assumptions on the environment. $\mathbb{X} = \{x_1, x_2, x_3, x_4\}$, with $x_1 < x_2 < x_3 < x_4$ and $\mathbb{I} = \{1\}$. Production function for a team of individuals of ability (x', x) that choose education investment (s', s) is given by

$$\hat{g}(x', x, s', s, a) = g(x' + s', x + s, a) - \kappa(s') - \kappa(s)$$

where

$$\begin{aligned} g(x', x, a) &= (ax' - c(a))(1 + \gamma x) \\ c(a) &= \frac{a^{1+\psi}}{1+\psi} \\ \kappa(s) &= \varphi_0 \frac{s^{1+\varphi_1}}{1+\varphi_1} \end{aligned}$$

The setting is familiar, with the exception of the convex costs of education $\kappa(s')$ and $\kappa(s)$. Also notice that education and skills are perfect substitutes. I omit any direct complementarity in order to highlight that this setting implies naturally a positive correlation between ability and education, even if everyone has identical production function of ability from school (education).

I need a single crossing condition that takes into consideration the endogenous choice of education. In particular I impose that γ is small enough such that at least in the case in which $\psi \to \infty$ is optimal to have an allocation with full skill-segmentation. This assumption is similar to the single-crossing condition, but takes into account also the endogenous choice of education.

Assumption A.10.1 (Single-Crossing with Endogenous Education)

$$\min_{y \in \mathbb{X}} 1 + \gamma \left[y + \left(\frac{\gamma}{\varphi_0} \right)^{\frac{1}{\varphi_1}} \right] > \gamma$$

and notice that if $\varphi_0 \to \infty$, hence if education becomes infinitely costly, hence is effectively fixed and equal to 0, this condition reduces to the same one used in the linear model, hence that $\gamma < 1$. Also, notice that, this condition is always satisfied if γ is small enough.

Lemma A.10.1 (Talent Segmentation or Segregation) *There exist two scalars κ_1 and κ_2 such that if $\psi \geq \kappa_1$ talent is segmented, and if $\psi \leq \kappa_2$ talent is segregated.*

Proof. The planner problem is to pick an optimal allocation $\varphi^* = \{(a_1^*, m_1^*, w_1^*, s_{m1}^*, s_{w1}^*), (a_2^*, m_2^*, w_2^*, s_{m2}^*, s_{w2}^*)\}$ to maximize total output:

$$\varphi^* \in \arg\max g(a_1, m_1, w_1, s_{m1}, s_{w1}) + g(a_2, m_2, w_2, s_{m1}, s_{w1})$$

$$\text{s.t}$$
$$\{m_1, m_2, w_1, w_2\} = \{x_1, x_2, x_3, x_4\}$$

I first show that the optimal allocation displays talent segmentation if ψ is large enough, and talent segregation if ψ is small enough, where notice that $\psi = \frac{c''}{c'a}$.

Similar to the choice of technology, the choice of education can be solved as a function of the a pair (x', x). Recall that I defined

$$v(x' + s', x + s) = g(x' + s', x + s, a^*(x' + s', x + s))$$

I can similarly define

$$\hat{v}(x', x) = \hat{g}(x', x, s_m^*(x', x), s_w^*(x', x), a^*(x' + s_m^*(x', x), x + s_w^*(x', x)))$$

where the optimal schooling of managers and workers are given by

$$\begin{aligned} s_m^*(x', x) &\in \arg\max_{s'} g(x' + s', x + s_w^*(x', x), a^*(x' + s', x + s_w^*(x', x))) \\ s_w^*(x', x) &\in \arg\max_{s} g(x' + s_m^*(x', x), x + s, a^*(x' + s_m^*(x', x), x + s)) \end{aligned}$$

the problem of the planner then becomes to maximize

$$\hat{v}(m_1, w_1) + \hat{v}(m_2, w_2)$$

$$\text{s.t}$$
$$\{m_1, m_2, w_1, w_2\} = \{x_1, x_2, x_3, x_4\}$$

applying the same arguments as in Lemmas 2-4, it is easy to see that the only two possible allocations are $\{(x_4, x_2), (x_3, x_1)\}$ and $\{(x_4, x_3), (x_2, x_1)\}$. The latter is optimal if and only if

$$\int_{x_2}^{x_3} \hat{v}_1(y, x_1)\, dy \geq \int_{x_2}^{x_3} \hat{v}_2(x_4, y)\, dy$$

next, the envelope theorem gives

$$\begin{aligned} \hat{v}_1(x', x) &= v_1(x' + s_m^*(x', x), x + s_w^*(x', x)) \\ \hat{v}_2(x', x) &= v_2(x' + s_m^*(x', x), x + s_w^*(x', x)) \end{aligned}$$

which can be used to rewrite the condition as

$$\int_{x_2}^{x_3} v_1\left(y+s_m^*\left(y,x_1\right),x_1+s_w^*\left(y,x_1\right)\right)dy \geq \int_{x_2}^{x_3} v_2\left(x_4+s_m^*\left(x_4,y\right),y+s_w^*\left(x_4,y\right)\right)dy$$

and, using the functional form assumptions, as

$$\int_{x_2}^{x_3}\left(y+s_m^*\left(y,x_1\right)\right)^{\frac{1}{\psi}}\left(1+\gamma\left(x_1+s_w^*\left(y,x_1\right)\right)\right)dy \geq \gamma\int_{x_2}^{x_3}\left(x_4+s_m^*\left(x_4,y\right)\right)^{\frac{1}{\psi}}dy.$$

This last inequality is satisfied if

$$\left(x_2+s_m^*\left(x_2,x_1\right)\right)^{\frac{1}{\psi}}\left(1+\gamma\left(x_1+s_w^*\left(x_2,x_1\right)\right)\right) \geq \gamma\left(x_4+s_m^*\left(x_4,x_3\right)\right)^{\frac{1}{\psi}}.$$

Next, notice that, from the first order conditions for optimal schooling choice we get optimal schooling of managers and workers to be respectively

$$\begin{aligned} s_m^*\left(x',x\right) &= \kappa_1^{-1}\left(v_1\left(x'+s_m^*\left(x',x\right),x+s_w^*\left(x',x\right)\right)\right) \\ s_w^*\left(x',x\right) &= \kappa_1^{-1}\left(v_2\left(x'+s_m^*\left(x',x\right),x+s_w^*\left(x',x\right)\right)\right) \end{aligned}$$

Using the functional form assumption and substituting into the last inequality we get

$$\left(x_2+\left[\frac{1}{\varphi_0}x_2^{\frac{1}{\psi}}\left(1+\gamma x_1\right)\right]^{\frac{1}{\varphi_1}}\right)^{\frac{1}{\psi}}\left(1+\gamma\left(x_1+\left[\frac{1}{\varphi_0}x_2^{\frac{1}{\psi}}\gamma\right]^{\frac{1}{\varphi_1}}\right)\right) \geq \gamma\left(x_4+\left[\frac{1}{\varphi_0}x_4^{\frac{1}{\psi}}\left(1+\gamma x_3\right)\right]^{\frac{1}{\varphi_1}}\right)^{\frac{1}{\psi}}$$

that holds as long as ψ is high enough, due to the single crossing assumption that modulates the value of γ.

Similarly, the allocation $\{(x_4,x_3),(x_2,x_1)\}$ is optimal if and only if

$$\int_{x_2}^{x_3}\hat{v}_1\left(y,x_1\right)dy \leq \int_{x_2}^{x_3}\hat{v}_2\left(x_4,y\right)dy$$

which can be shown, following similar steps to the ones above is satisfied if

$$\left(x_3+\left[\frac{1}{\varphi_0}x_3^{\frac{1}{\psi}}\left(1+\gamma x_1\right)\right]^{\frac{1}{\varphi_1}}\right)^{\frac{1}{\psi}}\left(1+\gamma\left(x_1+\left[\frac{1}{\varphi_0}x_3^{\frac{1}{\psi}}\gamma\right]^{\frac{1}{\varphi_1}}\right)\right) \leq \gamma\left(x_4+\left[\frac{1}{\varphi_0}x_4^{\frac{1}{\psi}}\left(1+\gamma x_2\right)\right]^{\frac{1}{\varphi_1}}\right)^{\frac{1}{\psi}}$$

that holds as long as ψ is low enough. This concludes the proof. □

Lemma A.10.2 (Education Skill-Relationship) *Optimal education is increasing in individual skills.*

Proof. First notice that using the chosen functional form we can solve explicitly for $s_m^*(x',x)$ and $s_w^*(x',x)$, they are given by

$$\begin{aligned} s_m^*\left(x',x\right) &= \left[\frac{1}{\varphi_0}x'^{\frac{1}{\psi}}\left(1+\gamma x\right)\right]^{\frac{1}{\varphi_1}} \\ s_w^*\left(x',x\right) &= \left[\frac{1}{\varphi_0}x'^{\frac{1}{\psi}}\gamma\right]^{\frac{1}{\varphi_1}}. \end{aligned}$$

This shows that $\frac{\partial}{\partial x'} s_m^* (x', w(x')) > 0$ and $\frac{\partial}{\partial x'} s_m^* (m(x), x) \geq 0$. Additionally, $s_m^* (x', x) > s_w^* (x', x)$. This is sufficient to show that education is increasing in skills in the allocation $\{(x_4, x_2), (x_3, x_1)\}$. However, in the allocation $\{(x_4, x_3), (x_2, x_1)\}$ a manager, x_2 has lower ability than a worker, x_3. Thus we have to show that $s_w^* (x_4, x_3) > s_m^* (x_2, x_1)$. To do so, notice the previous solution of s_m^* and s_w^*, that is

$$\begin{aligned} s_m^* (x_2, x_1) &= \kappa_1^{-1} (v_1 (x_1 + s_m^* (x_2, x_1), x_1 + s_w^* (x_2, x_1))) \\ s_w^* (x_4, x_3) &= \kappa_1^{-1} (v_2 (x_4 + s_m^* (x_4, x_3), x_3 + s_w^* (x_4, x_3))). \end{aligned}$$

In order for $\{(x_4, x_3), (x_2, x_1)\}$ to be optimal, $v_2 > v_1$ in the equation above, which thus implies - using implicit differentiation - that $s_w^* (x_4, x_3) > s_m^* (x_2, x_1)$. □

Lemma A.10.3 (Cross-sectional Dispersion of education) *Define the variable $s^* (x)$ that is equal to $s_m^* (x, w(x))$ if x is a manager and $s_w^* (m(x), x)$ if x is a worker. Var$(\log s^* (x))$ is higher when talent is segregated than when talent is segmented.*

Proof. First notice that the ratio

$$\frac{s_m^* (x', x)}{s_w^* (x', x)}$$

does not depend on ψ, which is the measure of the convexity of $c(a)$, however, if we take $x'' > x'$ we get that

$$\frac{s_w^* (x'', x)}{s_w^* (x', x)} = \frac{s_m^* (x'', x)}{s_m^* (x', x)} = \left(\frac{x''}{x'}\right)^{\frac{1}{\psi}}$$

thus the gap in education, for both managers and workers, is higher the smaller is ψ. Hence, dispersion of education across workers within the team does not depend on ψ. While dispersion of education of individuals in different teams is decreasing in ψ. Additionally, when ψ is low enough, we have a reallocation of workers, which farther increase dispersion of education. Next, notice that we can rewrite $\text{Var}(\log s^* (x))$ as

$$\text{Var}(\log s^* (x)) = \frac{1}{\varphi_1} \left[\frac{1}{\psi} \text{Var}(\log M(x)) + \text{Var}(\log L(x)) + \frac{1}{\psi} \text{Cov}(M(x), L(x)) \right]$$

where

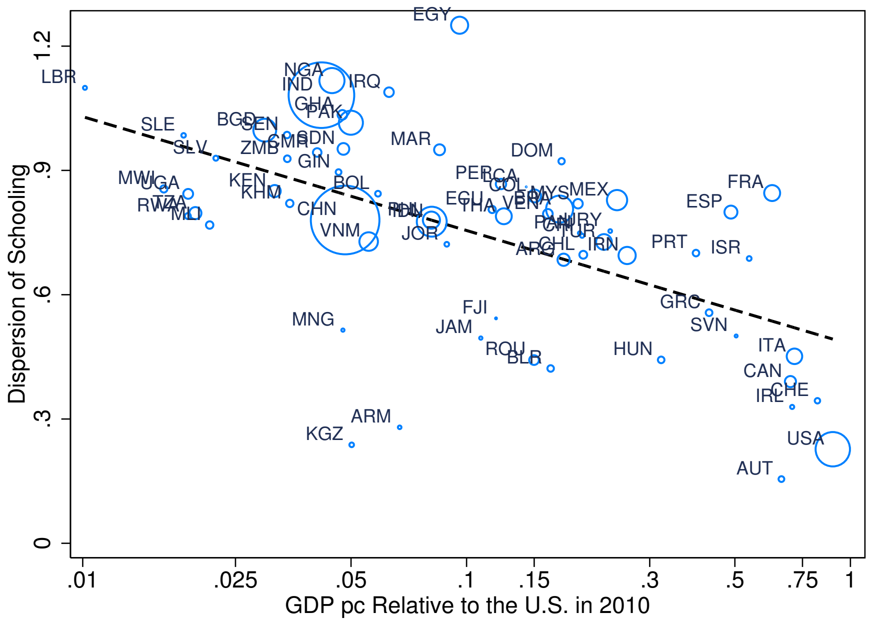
$$\begin{aligned} M(x) &= \begin{cases} x & \text{if } x \text{ is a manager} \\ m(x) & \text{if } x \text{ is a worker} \end{cases} \\ L(x) &= \begin{cases} (1 + \gamma w(x)) & \text{if } x \text{ is a manager} \\ \gamma & \text{if } x \text{ is a worker} \end{cases}. \end{aligned}$$

$M(x)$ and $L(x)$ depend on the optimal allocation. The dispersion of schooling increases in the distance from the frontier for two reasons: (i) for fixed allocation, mechanically a lower ψ increases the variance; (ii) when the allocation changes, this increases both $\text{Var}(\log M(x))$, $\text{Var}(\log L(x))$ and $\text{Cov}(M(x), L(x))$ and thus farther increases the variance. As a result, when ψ is low enough, so that there is talent segregation, the dispersion of education is larger. □

Remark: Multiple Equilibria. In this setting the Proposition 1 on decentralization does not hold anymore. There are in fact strategic complementarities in the choice of education. As a result, if individuals are not able to coordinate on the output maximizing equilibrium, there are in general multiple equilibria, and only one among them solves the planner problem.

Empirical Prediction Last, recall that I argued that countries far from the frontier have an allocation of talent that is closer to segregation by technology. Therefore, according to Lemma A.9.3, they should have larger dispersion of education (in logs)[75]. I use the previously described micro data from IPUMS international to show that this prediction is supported in the data. As can be seen in Figures A.20 and A.21, poor countries have larger dispersion of schooling than rich ones today and than the United States when was at similar level of development.

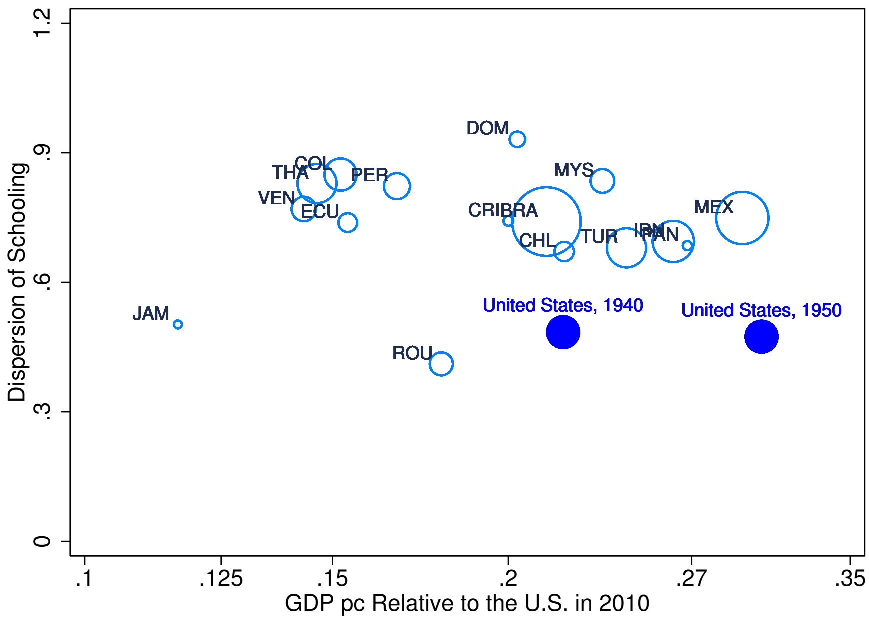
Figure A.20: Cross-county Differences in Dispersion of Schooling



Notes: this Figure is identical to Figure 6, with the difference that I plot the measure of dispersion of schooling, rather than the concentration of talent.

[75]In order to not drop individuals with zero schooling I assign 1 year of schooling to all individuals that report zero education. Additionally, I have computed robustness checks using, as a measure of dispersion, the coefficient of variation rather than the standard deviation of log years of schooling. Results are available upon request.

Figure A.21: Dispersion of Schooling in Developing Countries today compared to the U.S. in the past



Notes: this Figure is identical to Figure 7, with the difference that I plot the measure of dispersion of schooling, rather than the concentration of talent.

A.11 Additional Details on the Data

A.11.1 Household Level Data

Data Sources. The IPUMS data are available online at https://international.ipums.org/international/, through the Minnesota Population Center (2011). KLOSA dataset are available online at http://survey.keis.or.kr. KLIPS dataset are available from Cornell University through the Cross National Equivalent File project, see https://cnef.ehe.osu.edu. I use the version 8.0 of the Penn World Table, see Feenstra et al. (2013), available online.

Variable Construction and Remarks. Education years are imputed from educational attainment.

The industry variable is INDGEN in the IPUMS dataset. As described by Ipums: INDGEN recodes the industrial classifications of the various samples into twelve groups that can be fairly consistently identified across all available samples. The groupings roughly conform to the International Standard Industrial Classification (ISIC). IPUMS data also report information on the individual occupation, coded following the International Standard Classification of Occupations. However, this information is not useful to test the predictions of my model due to the fact that the occupation classification does not correspond to the notion of managers and workers in my model, but in fact depends on the technology used. As an example, a manager according to ISCO is an occupation with skill level 4, hence citing from their report available at ilo.org: "Occupations at this skill level generally require extended levels of literacy and numeracy, sometimes at very high level.." and also "typically involve the performance of tasks that require complex problem-solving, decision-making and creativity based on an extensive body of theoretical and factual knowledge..". For example, the manager of a small and low productivity firm - which is a manager according to the model language - would most likely not be classified as such in the data.

The variable sector is constructed by aggregating INDGEN into three sectors: agriculture, manufacturing, and services.

For each country I also use - when available - the non-harmonized industry variable, that varies from 1-digit to 4-digit in different countries. This is the variable IND in the IPUMS.

Self-employment is coded using the variables CLASSWK and CLASSWK detailed. For almost all countries, additional details are available and it is possible to distinguish whether a self-employed person is an employer or an own account worker. In the model, the definition of self-employed is equivalent to the definition of own account worker in the data. Hence, when available, I use this finer distinction. Since this detailed information is sometimes missing in few countries, I also computed the results using the coarser definition of self-employed or dropping countries for which the detailed information was rarely available. Results are very similar and available upon request. In the main analysis I use the self-employment variable to compute selection using the normalized skill x as described in the main text. I have explored alternative measures of selection. For example, using simple differences in average education, or this same differences weighted by the within country standard deviation of education. Results are robust to these alternative measures and are available upon request.

A.11.2 Firm Level Data

Data Sources. I use the round 2002-2006 and 2007-2014 of World Bank Enterprise Survey (WBES), available online at http://www.enterprisesurveys.org.

Variable Construction and Remarks. For the WBES 2006, I use the answer to the variable "What percent of the workforce at your establishment have the following education levels?" to code education. The possible choices are: *"less than 6, 6 to 9, 9 to 12, more than 12"*. I use mid point of the interval, and impute 14 years of education for the group "more than 12". I also use, as a robustness check, two alternative variables that report respectively the share of firm's workers with a high school degree and with some college. Results are similar and available upon requests.

The variable "top managers" in WBES 2006 does not correspond to the managers in the model, which should include also lower tier management. In fact, taking the model literally managers are the top half individuals in terms of skills within a firm.

The education variable in WBES 2014 is available only for manufacturing firms.

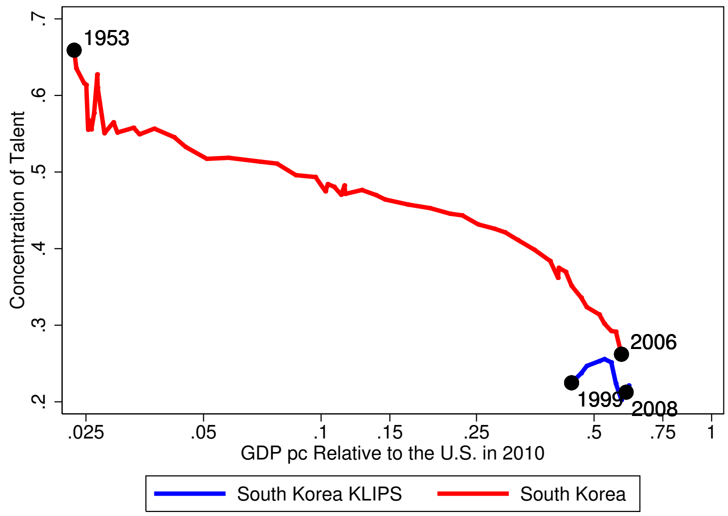
In order to compute the variance decomposition exercise shown in Figure A.12a, I compute the cross-sectional variance of education of the individuals hired at the firms in the sample. I can compute the country specific distribution of education using a variable that asks what is the fraction of the above described variable that asks to each firm the fraction of labor force with less than 6 years of education, between 6 and 9, between 9 and 12 and more than 12. I drop countries for which I have less than 10,000 total individuals (7 countries). I then decompose this total variance in the variance of education within firm and the variance of education across firms. For each country I compute the ratio between variance of education across firms and total variance and I plot it as a function of the country GDP per capita.

A.12 Robustness Checks and Additional Results for South Korea Data

A concern with the South Korea data is that we observe only one cross-section. In order to alleviate this concern, I compare the measure of concentration of talent, for the years for which is available, with data from KLIPS, that cover the whole population. Results are shown in Figures A.22 and A.23. They show that both the level of concentration of talent and the morphology of the data are similar in the two datasets. I also compute the structural transformation path implied by the KLOSA microdata and compare it with aggregate statistics from the World Bank Development Indicators. This is done in left panel of Figure A.24 and shows that the patterns are similar.

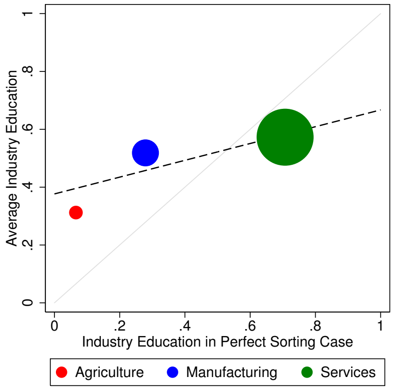
Last, in A.25 I report the disaggregated data from which the measure of concentration of talent is computed from 1960 to 2005. The process of structural transformation is evident in the figure, but we can notice that, throughout the growth miracle, the linear measure of concentration of talent consistently provides a reasonably good fit.

Figure A.22: Comparison between KLOSA and KLIPS dataset

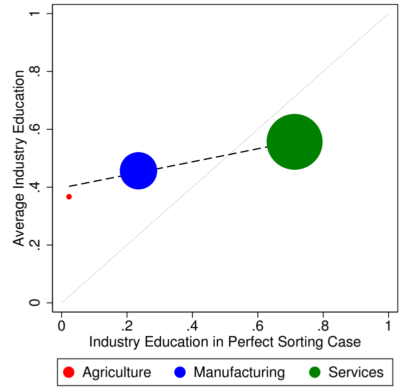


Notes: in the Figure I plot the growth path of concentration of talent across sectors computed from KLOSA data, as shown in Figure 8, and I compare it with the value of concentration of talent computed from the KLIPS data for the period 1999 to 2008.

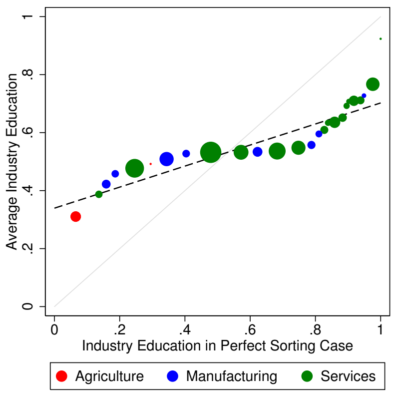
Figure A.23: Concentration of Talent, Comparison between KLOSA and KLIPS dataset in 2005



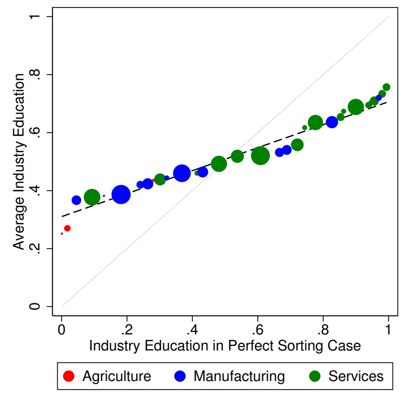
(a) KLOSA, Across Sectors



(b) KLIPS, Across Sectors



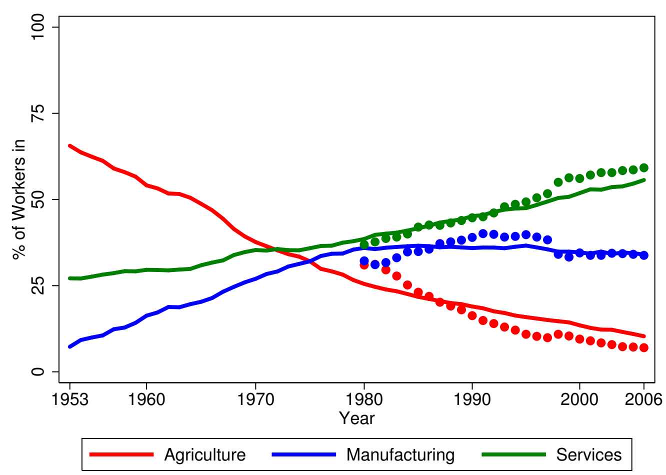
(c) KLOSA, Across Industries



(d) KLIPS, Across Industries

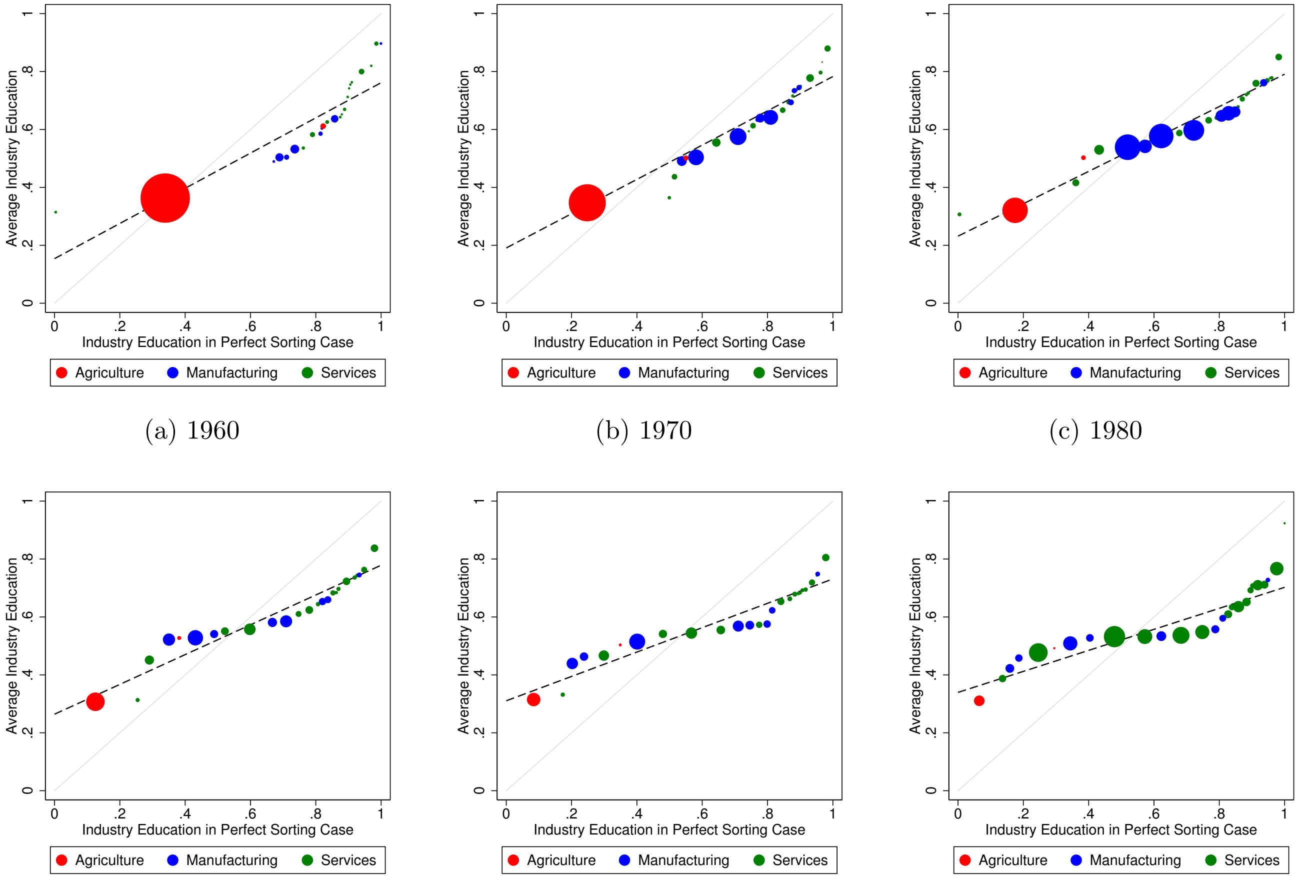
Notes: this figure replicates figures A.2 using both the benchmark KLOSA data and the KLIPS. The slopes of the regressions lines are, for sectors, 0.29 and 0.22 and, for industries, 0.37 and 0.40.

Figure A.24: Structural Transformation of South Korea



Notes: I plot the fraction of male population employed in either sector. Dots are values from (constructed) cross sections in KLOSA. The solid line are instead aggregate data for the male population from the World Bank Development Indicators.

Figure A.25: Concentration of Talent Across Industries, South Korea Growth Path



(a) 1960 (b) 1970 (c) 1980

(d) 1990 (e) 2000 (f) 2005

Notes: I apply, across industries, the same procedure as described in Figure A.2. Data are from KLOSA 2007. The slopes of the regression lines, which measure the concentration of talent, are, from 1960 to 2005: 0.61, 0.58, 0.56, 0.51, 0.42, 0.36.

A.13 Sequential Formulation of Dynamic Model and Proofs of Lemmas 9 and 10

I here describe the sequential formulation of the planner problem. The planner chooses an allocation φ, given by

$$\varphi = \{(m_{1t}, w_{1t}, a_{1t}), (m_{2t}, w_{2t}, a_{2t})\}_{t=t_0}^{\infty}.$$

The planner thus chooses for each sector, in $\{1, 2\}$, and each period after t_0 (time of take-off) a manager, a worker, and a technology. Feasibility of the allocation requires that all individuals are assigned to a sector and a partner, and that technology is chosen within the set of available technologies. The problem thus read as

$$\max_{\varphi} \sum_{t=t_0}^{\infty} \beta^t \left[g(m_{1t}, w_{1t}, a_{1t}; a_{1t-1}, \bar{a}_t) + g(m_{2t}, w_{2t}, a_{2t}; a_{2t-1}, \bar{a}_t)\right]$$

$$s.t.$$

$$\begin{aligned}
\{m_{1t}, w_{1t}, m_{2t}, w_{2t}\} &= \{x_1, x_2, x_3, x_4\} \ \forall t \\
a_{1t}, a_{2t} &\in \mathbb{A}_t = \{a_0, a_1, ..., \bar{a}_t\}. \\
\bar{a}_t &= \lambda^t \\
a_{1t_0} &= 1 \\
a_{2t_0} &= 1
\end{aligned}$$

where the planner takes as given the exogenous time of take-off, the growth path of the frontier technology, and the initial values of technology in each sector. Also, recall that

$$\begin{aligned}
g(m_{jt}, w_{jt}, a_{jt}; a_{jt-1}, \bar{a}_t) &= a_{jt} f(m_{jt}, w_{jt}) - c(m_{jt}, w_{jt}, a_{jt}; a_{jt-1}, \bar{a}_t) \\
f(m_{jt}, w_{jt}) &= m_{jt}(1 + \gamma w_{jt}) \\
c(m_{jt}, w_{jt}, a_{jt}; a_{jt-1}, \bar{a}_t) &= \tilde{c}(a; a_{jt-1}, \bar{a}_t)(1 + \gamma w_{jt}) \\
\tilde{c}(a_{jt}; a_{jt-1}, \bar{a}_t) &= \frac{a_{jt}}{\gamma^{\frac{\log \bar{a}_t}{\log \lambda} - \frac{\log a_{jt}}{\log \lambda}}} + a_{jt} \kappa \gamma^{\frac{\log \bar{a}_t}{\log \lambda} - \frac{\log a_{jt-1}}{\log \lambda}}
\end{aligned}$$

Next, I prove Lemma 9.

Proof of Lemma 9. First, recall the definition of $d_{jt} \equiv \frac{\log \bar{a}_t - \log a_{jt}}{\log \lambda}$. We can use it to rewrite $a_{jt} = \lambda^{t - d_{jt}}$. Substituting this into g as previously defined, we can rewrite

$$g(m_{jt}, w_{jt}, a_{jt}; a_{jt-1}, \bar{a}_t) = \lambda^t \hat{g}(m_{jt}, w_{jt}, d_{jt}; d_{jt-1})$$

where

$$\hat{g}(m_{jt}, w_{jt}, d_{jt}; d_{jt-1}) = \lambda^{-d_{jt}} \left(m_{jt} - (\lambda\gamma)^{-d_{jt}} - \kappa \lambda^{-d_{jt}} \gamma^{-d_{jt-1}} \right) (1 + \gamma w_{jt}).$$

Next, defining $\hat{\beta} \equiv \lambda \beta$, the sequential problem can be rewritten as choosing an allocation

$$\hat{\varphi} = \{(m_{1t}, w_{1t}, d_{1t}), (m_{2t}, w_{2t}, d_{2t})\}_{t=t_0}^{\infty}$$

that solves

$$\max_{\hat{\varphi}} \sum_{t=t_0}^{\infty} \hat{\beta}^t \left[\hat{g}(m_{1t}, w_{1t}, d_{1t}; d_{1t-1}) + \hat{g}(m_{2t}, w_{2t}, d_{2t}; d_{2t-1}) \right]$$

$$s.t.$$

$$\begin{aligned}
\{m_{1t}, w_{1t}, m_{2t}, w_{2t}\} &= \{x_1, x_2, x_3, x_4\} \ \forall t \\
d_{1t}, d_{2t} &\in \{0, 1, ..., t\}. \\
d_{1t_0} &= t_0 \\
d_{2t_0} &= t_0
\end{aligned}$$

and, as long as t is large enough, so that the constraint $d_{1t}, d_{2t} \in \{0, 1, ..., t\}$ does not bind, it is immediate to see that the problem can be written recursively as

$$v(d_1, d_2) = \max_{\{m_1, m_2, w_1, w_2, d_1', d_2'\}} \hat{g}(m_{1t}, w_{1t}, d_{1t}; d_{1t-1}) + \hat{g}(m_{2t}, w_{2t}, d_{2t}; d_{2t-1}) + \hat{\beta} v(d_1', d_2')$$

$$s.t$$

$$\{m_1, w_1, m_2, w_2\} = \{x_1, x_2, x_3, x_4\}$$

Last, notice that if $\hat{\beta} = 0$, due to Lemmas 1-4 in Section 2 only two allocation may be optimal $\{(x_4, x_2), (x_3, x_1)\}$ or $\{(x_4, x_3), (x_2, x_1)\}$. Therefore, by a standard continuity argument, this holds as long as $\hat{\beta}$ is low enough and thus we can rewrite the problem as

$$v(d_1, d_2) = \max\{v_A(d_1, d_2), v_B(d_1, d_2)\}$$

$$\begin{aligned} v_A(d_1, d_2) &= \max_{d_1', d_2'} g\left(x_4, x_2, d_1'; d_1\right) + g\left(x_3, x_1, d_2'; d_2\right) + \tilde{\beta} v\left(d_1', d_2',\right) \\ v_B(d_1, d_2) &= \max_{d_1', d_2'} g\left(x_4, x_3, d_1'; d_1\right) + g\left(x_2, x_1, d_2'; d_2\right) + \tilde{\beta} v\left(d_1', d_2',\right), \end{aligned}$$

which concludes the proof. □

Next, I prove the Lemma 10.

Proof of Lemma 10. If $\beta = 1$, eventually any country would turn to an allocation A in order to converge to the frontier, the reason being that along allocation B future output is lower. Hence, by a standard continuity argument, for any t_0 there exists a sufficiently high discount rate such that the planner values sufficiently the future and thus reallocate workers as in allocation A , possibly enjoying current losses, but being compensated by a future higher output. □

A.14 Cost of Technology, Details

I describe a microfoundation for the cost of technology used in the dynamic model. This microfoundation is useful to understand which assumptions are needed in order to have that countries farther from the technology frontier have a lower convexity of the cost of technology. The cost functions drawn in Figure A.7 and A.8 are computed from the setting described below.

The cost to use technology a is given by the sum of two components, one that is time dependent, $c_t(a)$, and one that depends on the distance from the frontier d and is thus country (or sector in the dynamic example) specific, $c_d(a)$

$$c_{td}(a) = c_t(a) + c_d(a).$$

I now describe each cost.

Time Dependent Cost $c_t(a)$ I interpret the cost $c_t(a)$ as the price to purchase the capital associated to any given technology. For example, $c_t(a)$ could be the price to purchase a computer, or a pencil and an accounting book. This price is pinned down by the cost of production. Cost of producing available technologies and their set, $\mathbb{A}_t$, evolve as follows. At time $t = 1$ the most basic technology, $a_1 = \lambda > 1$, is discovered. Production of this technology costs, at time $t = 1$, $c_1(a_1) = \lambda$. In each period $t > 1$ a new technology $a_t = \lambda a_{t-1}$, with associated cost $c_t(a_t) = \lambda c_{t-1}(a_{t-1})$, is invented. Additionally, production of previous technologies becomes cheaper: $c_t(a_j) = \frac{1}{\gamma} c_{t-1}(a_j)$ for all $j < t$, where $\gamma > 1$ modulates the efficiency improvement over time. As a result of this process, at each time t the technological environment is characterized by a set of available technologies $\mathbb{A}_t$ and their cost $c_t(\cdot)$ given by

$$\mathbb{A}_t \equiv \{a_1, a_2, ..., a_t\} = \left\{\lambda^1, \lambda^2, ..., \lambda^t\right\}$$

$$c_t(a_j) = \frac{a_j}{\gamma^{t-j}}.$$

Country Specific Cost $c_d(a)$ I index a country d by the number of steps between the country and the technology frontier. The index d must lie within the set of integers $\{0,...,t\}$.[76] For example, $d=0$ is a country on the technology frontier, while $d=5$ is a country five periods behind the frontier. I interpret the cost $c_d(a)$ as a direct cost associated with using the technology a. I interpret the ability x as an index of the relative skill of an individual within his country. In fact, I've assumed the distribution of x to be constant across countries. For this reason, one natural way to think at $c_d(a)$ is that it captures cross-country differences in the return to use each technology. The net return to use technology a for an individual x in country d is in fact given by $ax - c_d(a)$. The cost $c_d(a)$ is linear in a with marginal cost given by $\kappa\gamma^d$

$$c_d(a_j) = \kappa\gamma^d a_j,$$

where $\kappa \leq 1$ modulates the intensity of the cost and γ^d captures the idea that the farther a country is from the frontier the costlier is to use any given technology. For example due to the fact that technologies produced on the frontier are targeted for individuals with the skill-set available in those countries or due to the lack of complementary infrastructure, such as a reliable electrical grid.

However, despite this higher marginal cost, the assumption $\kappa \leq 1$ guarantees that countries far from the frontier gain from the access to frontier technology. If we compare a country that is d steps behind the frontier with a country that was on the frontier d periods ago, the former faces not only a larger set of available technologies, since $\mathbb{A}_{t-d} \subset \mathbb{A}_t$, but also a a cheaper cost for each technology. In fact $\kappa \leq 1$ guarantees that $c_{t,d}(a) \leq c_{t-d,0}(a)$ $\forall a$.

Interpretation in terms of race between human capital and technology

It can be useful to notice an alternative environment that provides the same cost of technology, but has a role for the distribution of human capital. Let each country start with human capital distributed as $x \sim U[0,1]$. Notice that here (departing from the rest of the paper) I use x as a cardinal and not an ordinal variable. Let the technology frontier evolve exogenously as before, with the addition that over time technologies are more expensive, and thus require an additional linear cost equal to $\kappa\gamma^t a_j$. Next, let the human capital of a country grow constantly, of course depending on the time in which a country opens up to the frontier. Specifically, let a country that has been growing for $t-d$ periods have $x \sim U[0,1] + \kappa\left(\gamma^t - \gamma^d\right)$. Human capital increases over time - and in such a way to perfectly compensate the cost of technology, however, if a country is farther from the frontier, human capital has increased by less. The functional forms are chosen ad hoc for this environment to yield an identical cost of technology as the one above.

This functional form has very useful properties that I now characterize. In order to make progress is useful to keep the assumptions of the quasi-linear model, and consider the problem of a manager x that must choose the optimal technology. This problem is

$$V_{td}(x) = \max_{j\in\{0,1..,t\}} a_j x - c_{td}(a_j).$$

Proofs are trivial and omitted for brevity.

[76] A country cannot be more than t steps away from the frontier, since otherwise should have negative time. I throughout assume that t is a very large number so that all countries would choose $d < t$.

Lemma A.14.1 (Homogeneity in a_t) For t sufficiently large, the value $V_{td}(x)$ can be rewritten as $V_{td}(x) = a_t v_d(x)$ where

$$\begin{aligned} v_d(x) &= \max_{\delta \in \{0,-1,..\}} a_\delta x - \tilde{c}_d(a_\delta) \\ \delta &= j - t \\ \tilde{c}_d(a_\delta) &= \left(\gamma^\delta + \kappa \gamma^d\right) \lambda^\delta \end{aligned}$$

The optimal technology choice, relatively to the frontier one, is thus constant over time, and is only a function of the distance, d, of a country from the frontier.

Optimal Technology Choice I define $\alpha_d^*(x)$ to be the optimal technology choice, relative to the frontier one, of a manager of type x in country d:

$$\alpha_d^*(x) \equiv \arg \max_{\delta \in \{0,-1,..\}} a_\delta x - \tilde{c}_d(a_\delta).$$

I next show the properties of $\alpha_d^*(x)$ as a function of d.

Lemma A.14.2 (Level of Optimal Technology) For any type x, the optimal technology $\alpha_d^*(x)$ is decreasing in d.

This lemma shows that countries far from the technology frontier have on average lower level of technology. This comes directly from the fact that they face a higher cost.

Lemma A.14.3 (Dispersion of Optimal Technology) For any two types $x' > x$, the ratio between their optimal technologies, $\frac{\alpha_d^*(x')}{\alpha_d^*(x)}$, is increasing in d.

This lemma shows that in countries far from the technology frontier the gap in optimal technology for any two types is larger than in countries close to the frontier. This comes from the fact that the cost of technology is less convex[77] the farther a country is from the frontier.

Next, let's assume that the distribution of types x is identical across countries, with support $[x_{\min}, x_{\max}]$ and positive density everywhere. Under this assumption, the set of technologies used in a country d, which I define $\mathbb{A}_d^*$, is simply given by

$$\mathbb{A}_d^* = \{\alpha_d(x_{min}), .., \alpha_d(x_{max})\}.$$

As shown in the next lemma, the number of technologies used is a function itself of the distance of a country from the frontier, with countries farther from it using a larger number of technologies for a given distribution of ability.

Lemma A.14.4 (Set of Used Technologies) The cardinality of $\mathbb{A}_d^*$ is increasing in d.

Remark: a Differentiable Version Last, let me notice that this cost function has been defined over a discrete set of technologies. However, we can define a cost over a continuous a such that at the discrete values $\lambda, \lambda^2, ..., \lambda^t$ takes the value of $c_{td}(a_j)$. This function, which I define $\tilde{c}_{td}$ is

$$\tilde{c}_{td} = \gamma^{-t} a \gamma^{\frac{\log a}{\log \lambda}} + \kappa \gamma^d a.$$

[77] Since $\tilde{c}_d(a_\delta) = \left(\gamma^\delta + \kappa\gamma^d\right)\lambda^\delta$ is a weighted sum of a convex and a linear cost in technology, with the weight on the linear part being larger the farther a country is from the frontier.