Cross-Country Differences in the Optimal Allocation of Talent and Technology*

Tommaso Porzio†

UCSD

Click here for most current version

September 22, 2016

[PRELIMINARY AND INCOMPLETE]

Abstract

I model an economy inhabited by heterogeneous individuals that form teams and choose an appropriate production technology. The model characterizes how the technological environments shapes the assignment of individuals into teams. I apply the theoretical insights to study cross-country differences in the allocation of talent and technology. Their low endowment of technology, coupled with the possibility of importing advanced one from the frontier, naturally leads poor countries to a different economic structure, with stronger concentration of talent and larger cross-sectional productivity dispersion. As a result, the efficient equilibrium in poor countries displays economic features, such as larger productivity gaps across sectors, that are often cited as evidence of misallocation. Micro data from countries of all income levels documents cross-country differences in the allocation of talent that support the theoretical predictions. A quantitative application of the model suggests that a sizable fraction of the larger productivity dispersion documented in poor countries is due to differences in the efficient allocation.

---

*This paper is based on my PhD dissertation at Yale University. This project deeply benefited from numerous conversations with colleagues and advisors. First of all, 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, and Yu Jung Whang. Last, but not least, I received excellent comments from seminar participants at Yale and several other universities.

†9500 Gilman Drive, La Jolla, CA 92037, email: tporzio@ucsd.edu.
1 Introduction

Poor countries have low labor productivity. They produce lower output for the same amount of input. At the same time, cross-sectional productivity dispersion is larger in less developed countries: most individuals see their labor produce very little output, but some others enjoy a fairly high productivity.\(^1\) Usually, this excess productivity dispersion is interpreted as evidence of larger frictions in developing countries, which prevent many people from moving to more productive firms and sectors, or to adopt the best available technologies.\(^2\)

In this paper, I explore an alternative view. I argue that the larger productivity dispersion may be a natural side effect of poor countries having access to advanced technologies imported from the distant technology frontier. The economic mechanism I consider entails two insights.

First, low income countries have the unique opportunity to directly adopt technologies, discovered elsewhere, that are much more advanced relative to their current level of development. As long as individuals differ in their returns from adoption, for example due to heterogenous ability, only some would take advantage of this opportunity, thus naturally leading to dispersion of technology, hence likely of productivity.\(^3\) As an example, consider India. It has approximately the same level of GDP per capita of the United States at the beginning of the past century. Nonetheless, we would not be surprised to spot in the streets of Kolkata someone using his cellphone to shop online from Flipkart, the Indian alternative to Amazon. And that same person, could be cruising around on a pulled rickshaw, just as he was doing more than 100 years ago. A wide range of technologies coexist in India today. Arguably, many more than in the United States at the beginning of the past century.

Second, I show that the coexistence of different technologies in developing countries shapes the allocation of talent and is itself affected by it. High skilled individuals have higher returns from accessing more advanced technology, and thus they cluster together in the few economic niches where modern technology is pervasive. In equilibrium, if all high skilled individuals form teams with each other to produce, the rest of the economy is left mostly with low skilled ones, that have low inherent productivity and no incentives to adopt advanced technology. The concentration of similarly talented individuals within the same teams is thus both a consequence and a catalyst of productivity and technology dispersion.

These economic forces are not mere theoretical possibilities. In this paper, I use micro data from several countries to show that talent is more concentrated in less developed ones, as predicted. Moreover, I show, through the lens of a quantitative version of the model, that these

\(^1\)See Caselli (2005), and Hsieh and Klenow (2009) (Figure I of TFPQ, not the famous Figure II of TFPR).

\(^2\)Two notable exceptions are? and Lagakos and Waugh (2012).

\(^3\)The idea that individuals of different abilities may have different appropriate technology dates as back as the “New view of technological change” of Atkinson and Stiglitz (1969). The prediction that only some individuals in poor countries adopt advanced technology frames well with empirical evidence of Comin and Mestieri (2013), that shows that modern technologies are used everywhere, but with different penetration rates.
differences are sizable, and that the mechanism outlined can significantly contribute to explain
the larger productivity dispersion in poor countries.

**Overview.** The paper is organized in four sections.

In Section 2, I develop a theoretical framework to analyze the joint determination of the
allocation of talent and technology within an economy inhabited by individuals of heterogenous
ability. Production requires three inputs: a manager, a worker, and a technology. The produc-
tion function satisfies two key assumptions. First, output is more sensitive to the ability of the
manager than to the ability of the worker. Second, there is complementarity between technology
and the ability of both the manager and the worker. Each individual chooses his occupation:
whether to be a worker or a manager. Managers choose the ability of the worker to hire, and
the level of technology to operate. More advanced technologies are costlier, but allows to achieve
higher labor productivity. I show that the competitive equilibrium of this setting decentralizes the
Pareto efficient allocation. I then characterize it. The main object of interest are the endogenous
sets of managers and workers. Given these two sets, complementarity dictates positive assorta-
tive matching - more able managers are matched with more able workers - and so the unique
equilibrium is pinned down. As standard, each individual chooses the occupation where he has
a comparative advantage. The main difficulty of this setting is that the comparative advantages
are endogenous and depend on the technology choice of each team.\footnote{Usually, comparative advantages are exogenous. For example, in Roy models of occupational choice individuals have a managerial and a worker specific ability. See for an application Hsieh et al. (2013). Another set of occupational choice models are based on Lucas (1978). Here individuals have unidimensional skills, but the assumption of the production function are such that the highest skilled have a comparative advantage in being managers.} If every team would pick
identical technology, the higher skill-sensitivity of output dictates that the most skilled individu-
als have a comparative advantage in being managers. Endogenous technology choice introduces
a new trade-off. In equilibrium, each individual uses a higher technology if he decides to be a
worker. Hence the trade-off: to use a more advanced technology, which increases the marginal
value of your skills, you need to choose a lower skill-intensive occupation. In the main theoretical
results of the paper, I formally show that the equilibrium assignment, and specifically the ability
gap between a worker and his manager, depends on the relative strength of the asymmetry in
skill-sensitivity of the production function and the elasticity of the optimal technology choice to
the ability of the team members. The allocation of talent and technology (and thus productivity)
are tied together: when technology is dispersed there is a lower ability gap between managers and
workers. In this case, talent is concentrated: high ability individuals pair together and use the
most advanced technology; low ability ones also pair with each others and use lower technology.

In Section 3, I explicitly model cross-country differences in the cost of technology. I distin-
guish between technology and technology vintage. I call technology the productivity term that
multiples labor input. I argue that what I call technology of a team depends on two elements: (i) the choice of which vintage of technology to use, for example relying on animal or electric power; and (ii) the amount of capital of the used vintage, for example how many bullocks are purchased. Newer technology vintages have a lower marginal cost for each technology level, but a larger fixed cost. Countries differ in their local technology vintage. Individuals can use the local vintage without the need to pay any fixed cost. Individuals can also import more modern vintages from other countries, but at a cost. I derive from this environment a country specific cost of technology and show that the elasticity of technology to the ability of team members depends on the distance of a country from the technology frontier, defined as the gap between the local vintage and the most advanced one available. Intuitively, this depends on the fact that only the most able teams use modern vintages, while less able ones find optimal to rely on the local ones. Combining the insights of Sections 2 and 3 the theory provides sharp predictions on cross-country differences in the economic structure. First of all, the model predicts that the efficient competitive equilibrium generates larger technology and labor productivity dispersion in poor countries. A feature often associated with misallocation is generated, in this framework, through differences in endowments, captured by the lower available local technology vintage. Second, the theory predicts that talent should be more concentrated in countries far from the frontier.

In Section 4, I 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 higher, as predicted by the model. In the data, we do not directly observe either teams or individual skill. I show that under two assumptions - (i) education is correlated with ability and (ii) individuals within the same industry use a more similar technology than those in different industries - we can construct, using the distribution of schooling within and across industries, an empirical measure of concentration of talent that aligns well with the one 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

---

5I use technology to distinguish from productivity, that is output divided by the number of workers, and thus takes into account the worker ability.
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.

Finally, in Section 5 I write a computational version of the model, and show that the cross-country differences in concentration of talent are consistent with sizable differences in within country productivity dispersion. Specifically, I estimate the parameters of the model to target cross-country differences in the concentration of talent and the agricultural productivity gap in rich countries. I then use the model to predict agricultural productivity gaps in poor countries and show that the model, once disciplined with the cross-country differences in the allocation of talent, accounts for approximately 40% of the larger agricultural productivity gap in poor countries.

**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

---

6Labor productivity in agriculture in the model is defined as the average labor productivity of the teams who use a technology that employs individuals whose average ability is the same as the one empirically measured for agriculture in countries of the relevant income level. I use the same procedure to calculate non agricultural labor productivity.
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.

2 A Model of Technology Choice and Allocation of Talent

This section develops a matching model in which heterogeneous individuals form production teams and choose an appropriate and costly technology. The model highlights that the equilibrium technology distribution and allocation of talent are intertwined.

2.1 Environment

The economy is inhabited by a continuum of mass one of individuals, indexed by their ability $x \sim U [0, 1]$. Individuals with higher $x$ are more able. Each individual supplies inelastically one unit of labor and has a non-satiated and increasing utility for the unique final good produced in
the economy.

**Production.** Production of labor input requires two individuals to join in a team: a manager and a worker. A manager of ability $x'$ paired with a worker of ability $x$ produces $f(x', x)$ units of labor inputs, where $f(x', x)$ is strictly increasing in both arguments and twice continuously differentiable.\(^7\) Production of the final good require the labor input to be combined with a production technology $a \in \mathbb{R}$, that multiplies labor input. There is a continuum of available technologies in the economy, and technology $a$ has a cost $c(a)$, that is increasing, convex, and twice continuously differentiable. $c(a)$ captures the cost, in terms of output, to purchase the vintage of capital associated with a given technology and to learn how to use it. Higher technologies are costlier, but yield higher returns by combining more capital to the same amount of labor input.

A production team $(a, x', x)$ where $a$ is the technology, $x'$ is the ability of the manager, and $x$ is the ability of the worker produces $g(a, x', x)$ units of the final output

$$g(a, x', x) = a f(x', x) - c(a).$$

Since $f$ is increasing in both arguments, technology is complementary with the skills of both the manager and the worker: $g_{12}(a, x', x) > 0$ and $g_{13}(a, x', x) \geq 0$.

**Assumptions on Production of Labor Input.** I assume that the production function of labor input satisfies the following three properties

- (1): $f_1(x, x) > f_2(x, x) \forall x \in [0, 1]$
- (2): $f_{12}(x', x) > 0 \forall (x', x) \in [0, 1] \times [0, 1]$
- (3): $f_1(x, y) > f_2(z, x) \forall (x, y, z) \in [0, 1] \times [0, 1] \times [0, 1]$.

(1) captures the fact that, for given technology and partner, the individual’s skills are 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. (2) captures complementarity in production between tasks. This is a natural assumption, pervasive in the literature. (3) is a Spence-Mirrlees condition that separates types into occupations, by imposing that, if all teams use the same technology, the complementarity between types (2) is sufficiently weak relatively to the skill asymmetry (1) and thus high skilled individuals have a comparative advantage in the

\(^7\)Please notice the slight abuse of notation, I refer to $x$ as the ability of an individual in general, however, when I want to distinguish explicitly between the ability of the manager and the worker I refer to the former as $x'$ and to the latter as $x$. 

6
managerial occupation. As I will show, (3) is not sufficient to separate types if technology choice is endogenous.

**Complementarity Skill-Technology.** Assumption (3) implies that

\[
(4) \ g_{12}(a, x, y) > g_{13}(a, z, x) \geq 0 \ \forall (a, x, y, z) \in \mathbb{R} \times [0, 1] \times [0, 1] \times [0, 1].
\]

Technology and skills are complementary, consistent with most evidence for both developed and developing countries. (See, for example, Goldin and Katz (1998), and Foster and Rosenzweig (1995)). Further, the ability of the manager is more relevant to generate high returns from technology than the ability of the worker, as emphasized in recent studies that highlighted the role of managers in technology adoption (see Bloom and van Reenen (2007) and Gennaioli et al. (2013)).

**Assignment of Talent to Teams.** Production requires one manager and one worker. Individuals ability \(x\) is observable, and individuals are not restricted ex-ante to belong to either group. As a result an allocation in this setting comprehends an occupational choice function, \(\omega(x) : [0, 1] \Rightarrow [0, 1]\), that defines the probability that an individual \(x\) is a worker, and a matching function, \(m(x) : [0, 1] \Rightarrow [0, 1]\), that assigns to each worker of ability \(x\) the ability type of his manager.\(^{10}\)

### 2.2 Competitive Equilibrium

The competitive equilibrium of this economy is given by five functions: optimal technology \(\alpha(x', x) : [0, 1] \times [0, 1] \Rightarrow \mathbb{R}\); profit \(\pi(x) : [0, 1] \Rightarrow \mathbb{R}\); wage \(w(x) : [0, 1] \Rightarrow \mathbb{R}\); occupational choice \(\omega(x) : [0, 1] \Rightarrow [0, 1]\); and matching \(m(x) : [0, 1] \Rightarrow [0, 1]\) such that

1. each team chooses the optimal technology

\[
\alpha(x', x) = \arg \max_{a \in \mathbb{R}} af(x', x) - c(a).
\]

2. Each manager chooses the type of worker to hire taking into account the optimal technology the pair would choose and taking as given the equilibrium wage schedule

\[
\pi(x) = \max_{z \in [0, 1]} \alpha(x, z) f(x, z) - w(z) - c(\alpha(x, z)).
\]

\(^8\)Assumption (3) is stronger than (1). Nonetheless I find it useful for exposition to conceptually separates the three assumptions. Formally, (1) is a redundant assumption.

\(^9\)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.

\(^{10}\)The definition of \(m(x)\) as a function rather than a correspondence might seem restrictive. However, it is not. In fact, I prove in a previous version of this paper, (Porzio (2016)), that for the currently considered case in which \(x \sim U[0, 1]\), \(m(x)\) is indeed a function and not a correspondence.
3. The matching function is consistent with manager’s optimality

\[ m(z^*(x)) = x \]

where \( z^*(x) \) is the solution of the manager’s problem.

4. Each individual chooses the occupation that pays him the higher income, or randomizes among them if \( \pi(x) = w(x) \): \( \omega(x) \) satisfies

\[ \omega(x) \in \arg\max_{z \in [0,1]} (1-z)\pi(x) + zw(x). \]

5. The sum of wage and profit of a team equals their produced output, \( \forall x \)

\[ \pi(m(x)) + w(x) = g(\alpha(m(x), x), m(x), x). \]

This restriction also guarantees that the goods market clear.

6. Labor market clears for each individual type \( x \). That is, for each type the probability that an individual of type \( x \) is a worker must equal the probability that an individual of type \( m(x) \) is a manager:

\[ \omega(x) = 1 - \omega(m(x)). \]

**Proposition 1: Existence and Pareto Efficiency.**

A competitive equilibrium exists and is Pareto Efficient.

*Proof.* See appendix. □

Equilibrium uniqueness is not guaranteed in this environment. Nonetheless, this does not represent a concern, since the properties of the equilibrium that I characterize in the next section hold for any equilibrium. In Section 3.2, I provide a tractable case for which I prove uniqueness by equilibrium construction.

### 2.3 Equilibrium Characterization

I characterize the equilibrium and show how the assignment of talent to teams and the choice of technology are related.

#### 2.3.1 Choice of Technology

A team of a manager of ability \( x' \) and a worker \( x \) picks the technology that maximizes their net output, that is

\[ \alpha(x', x) = c^{-1}(f(x', x)). \]
The complementarity between technology and labor input and the assumptions on the functional form of $f$ give the following result.

**Lemma 1: Appropriate Technology**

The appropriate technology of a team increases in the skills of both the manager and the worker, but more so in the skills of the manager: $\alpha_1 > \alpha_2 \geq 0$.

*Proof.* See appendix. □

The manager and the worker agree on the choice of technology, since it does not affect how they share output between each other.

**2.3.2 Manager Problem**

Consider a manager of ability $x$. He picks the optimal type of workers to maximize his profit $\pi(x)$, that is

$$
\pi(x) = \max_{z \in [0,1]} \alpha(x, z) f(x, z) - w(z) - c(\alpha(x, z)).
$$

The solution of this maximization problem yield the matching function, that assigns to each worker his manager, and satisfies $m(z^*(x)) = x$. The skill complementarity between managers and workers ($f_{12} > 0$) implies that the matching function is increasing.

**Lemma 2: Matching Function**

The matching function $m^*$ is increasing almost everywhere: if $x' > x$ then $m(x') > m^*(x)$.

*Proof.* See appendix. □

The envelope and first order conditions give the slopes of the profit and wage functions

$$
\begin{align*}
\pi'(x) &= \alpha(x, m^{-1}(x)) f_1(x, m^{-1}(x)) \\
w'(x) &= \alpha(m(x), x) f_2(m(x), x)
\end{align*}
$$

where $m^{-1}(x)$ is the inverse of the matching function, and thus assigns to each manager the type of his worker. By definition, $m^{-1}(x) = z^*(x)$ and the function $m$ is invertible due to the fact that is strictly increasing almost everywhere. The slopes of the profit and wage functions determine the marginal values of skills in each occupation, which, as I will show, drive the occupational choice. These values depend on the used technology and the occupation specific skill-sensitivity. In general, an individual has different partners, and thus different appropriate technology whether he is a worker or a manager. The higher the used technology, the more skills are valued, due to skill-technology complementarity. At the same time, the tasks of a manager are different from the task of a worker, and thus each occupation is going to have a specific skill-sensitivity that affects its overall marginal value of skills.
2.3.3 Occupational Choice

I first discuss the optimal assignment to occupations within a team. Two individuals that work together use, by assumption, identical technology. As a result, the relative marginal value of skills in either occupation depends only on the asymmetry in skill-sensitivity. Due to the Spence-Mirlees assumption, the managerial task uses skills more efficiently, and thus the most skilled individual of the team must be the manager.

Lemma 3: Occupational Choice within a Team

The manager of the team is more skilled than the worker of the team: \( m(x) \geq x \forall x \in [0, 1] \).

Proof. See appendix. \( \square \)

The Technology-Occupation Tradeoff. Lemma 3 shows that each individual \( x \) is matched with a more skilled partner, and thus uses a more advanced technology if he decides to be a worker rather than a manager. There is a technology-occupation trade-off: an individual would use a higher technology, which gives, ceteris paribus, a higher marginal value to of his skills, if he selects into the less-skill intensive occupation. As a result, an individual \( x \) sees his skills having a higher marginal value as a manager - i.e. \( \pi'(x) \geq w'(x) \) - if and only if the gap in skill-sensitivity across occupations is sufficiently large with respect to the technology gap across them:

\[
\frac{f_1(x, m^{-1}(x))}{f_2(m(x), x)} \geq \frac{\alpha^*(m(x), x)}{\alpha^*(x, m^{-1}(x))}.
\]

Skill-Sensitivity Gap Technology Gap

Individuals, as usual, select into the occupation where they have their comparative advantage. High skilled individuals have a comparative advantage in the occupation that has the highest marginal value of skills. Most of the literature\(^{11}\) focuses on functional form assumptions such that management is more skill-intensive, that is \( \pi'(x) \geq w'(x) \forall x \). For example, Lucas (1978) uses a production function that gives \( w'(x) = 0 \): high skilled have a comparative advantage in being managers, and thus the shape of the occupational choice function is known ex-ante and is given by a cutoff policy that separates types into two connected sets of managers and workers. In this setting, instead, the comparative advantage of each individual \( x \) is endogenous, since it depends on the optimal technology choice of each team.\(^{12}\) For example, some high skilled individuals may find their skills more rewarded by being workers with a high technology rather than managers with a lower one. How to solve for this complex fixed point problem? I develop a method to use the necessary conditions for optimality, together with market clearing, to characterize the overall equilibrium assignment and show how it depends on the shape of \( f \) and \( \alpha \).

\(^{11}\)The one notable exception that I am aware of is Kremer and Maskin (1996).

\(^{12}\)Since matching is one-to-one, the 50\% most skilled individuals have a comparative advantage in the occupation that has higher marginal value of skills (note: this occupation might be different for each \( x \)). The 50\% less skilled, instead, have a comparative advantage in the occupation that has lower marginal value of skills.
Necessary Conditions. The occupational choice function \( \omega(x) \) maximizes individual income:

\[
\omega(x) \in \arg\max_{z \in [0,1]} (1 - z) \pi(x) + zw(x). \tag{2}
\]

\( \omega(x) \) divides the type space into subsets in which individuals are either managers, or workers, or randomize between the two occupations. Since the type space is a compact set, individuals that are at the boundaries between two subsets in which two different occupations are picked, must be indifferent between being a manager or a worker. Similarly, individuals that randomize between the two occupations must be indifferent as well. For these individuals, the maximization problem (2) provides useful necessary conditions that link the occupational choice to the marginal values of skills in either occupation.

**Lemma 4: Necessary Conditions of Occupational Choice Function**

The occupational choice function satisfies, for \( \epsilon \to 0 \):

1. \( \forall x \text{ such that } \omega(x) \in (0,1): \pi'(x) = w'(x) \)
2. \( \forall x \text{ such that } \omega(x - \epsilon) = 0 \text{ and } \omega(x) > 0: \pi'(x) \leq w'(x) \)
3. \( \forall x \text{ such that } \omega(x - \epsilon) = 1 \text{ and } \omega(x) < 1: \pi'(x) \geq w'(x) \).

**Proof.** See appendix. □

These conditions are derived from the fact that the wage and profit functions must cross in proximity of ability type \( x \) that are indifferent between either occupation. I use them to characterize the equilibrium. All proofs are left to the appendix. But as a simple example of the method I use, I provide here the proof of Lemma 3.

**Proof of Lemma 3.** Let \( x > x' \), with \( m(x) = x' \). Then let \( \hat{x} \) be the lowest type larger than \( x' \) with \( \omega(\hat{x}) > 0 \). Since \( \omega(x) > 0 \) and \( \omega(x') < 1, \hat{x} \in [x', x] \). By the necessary conditions, \( w'(\hat{x}) \geq \pi'(\hat{x}) \). Substituting in Equation (1), \( w'(\hat{x}) \geq \pi'(\hat{x}) \) becomes

\[
\frac{\alpha(m(\hat{x}), \hat{x})}{\alpha(x, m^{-1}(\hat{x}))} \geq \frac{f_1(\hat{x}, m^{-1}(\hat{x}))}{f_2(m(x), \hat{x})}.
\]

Lemma 1 showed that \( m \) and \( m^{-1} \) are increasing. As a result, since \( x \geq \hat{x} \), then \( m(\hat{x}) \leq x' \leq \hat{x} \), and also \( m^{-1}(\hat{x}) \geq x \geq \hat{x} \). Therefore \( \frac{\alpha(m(\hat{x}), \hat{x})}{\alpha(x, m^{-1}(\hat{x}))} \leq 1 \), since \( \alpha \) is increasing in both his arguments.

The Spence-Mirlees assumption (3) guarantees that \( \frac{f_1(\hat{x}, m^{-1}(\hat{x}))}{f_2(m(x), \hat{x})} > 1 \). This lead to a contradiction, and thus to \( m(x) \geq x \). □

### 2.3.4 Technology Choice and Equilibrium Assignment of Talent to Teams

The main characterization result ties the properties of the optimal technology choice \( \alpha \) and the production function \( f \) to the assignment of talent to teams.
**Proposition 2: Assignment of Talent Across Teams**

In a competitive equilibrium, for any worker \( t \in [0, 1] \) the ability gap between him and his manager, \( m(t) - t \), is bounded above by \( \Upsilon (t) \) and below by \( \Lambda (t) \), where \( \Upsilon (t) \) and \( \Lambda (t) \) depend on \( f \) and \( \alpha \) as follows

1. consider two functions \( f \) and \( \tilde{f} \), if \( \forall (x, y, z) \in [0, 1] \times [0, 1] \times [0, 1] \) such that \( x \geq y \geq z \)

\[
\frac{f_1(x,y)}{f_2(y,z)} > \frac{\tilde{f}_1(x,y)}{\tilde{f}_2(y,z)} \text{ then } \Upsilon (t) \geq \tilde{\Upsilon} (t) \text{ and } \Lambda (t) \geq \tilde{\Lambda} (t) \forall t \in [0, 1];
\]

2. consider two functions \( \alpha \) and \( \tilde{\alpha} \), if \( \forall (x, y, z) \in [0, 1] \times [0, 1] \times [0, 1] \) such that \( x \geq y \geq z \)

\[
\frac{\alpha(x,y)}{\tilde{\alpha}(y,z)} \geq \frac{\tilde{\alpha}(x,y)}{\tilde{\alpha}(y,z)} \text{ then } \Upsilon (t) \leq \tilde{\Upsilon} (t) \text{ and } \Lambda (t) \leq \tilde{\Lambda} (t) \forall t \in [0, 1].
\]

**Proof.** See appendix. \( \square \)

In the appendix I include the implicit functions that defines each bound. The proposition shows that \( \Lambda (x) \leq m(x) - x \leq \Upsilon (x) \) and that these upper and lower bounds depends on the shape of \( f \) and \( \alpha \). \( f \) and \( \alpha \) modulates the technology-occupation trade-off and thus change the equilibrium assignment of talent into teams. When there is stronger asymmetry in skill-sensitivity, so that everything else equal, high skilled individuals have a stronger comparative advantage in being managers, then the gap between workers and managers widens, since more high skilled find it optimal to be managers. When instead the elasticity of optimal productivity choice to the ability of team members increases, thus when teams formed by individuals of different ability choose very different optimal technologies, then it becomes less important your occupation and more the technology that you choose to produce. Individuals then match with other of similar abilities and the gap between a worker and his manager reduces.

It is simple to see that when \( m(x) - x \) is smaller, than talent is more concentrated, since some teams gather together the high skilled, and thus in equilibrium other teams are left with low skilled managers and workers. Formally, I define the concentration of talent as follows.

**Definition 1: Concentration of Talent**

Consider two matching function \( m \) and \( \tilde{m} \). Talent is more concentrated according to \( m \) if

\[
\int [m(x) - x] \omega (x) dx < \int [\tilde{m}(x) - x] \omega (x) dx.
\]

I define \( 1 - \int [m(x) - x] dx \) the concentration of talent.

Proposition 2 showed, the concentration of talent is tightly linked to the strength of either side of the technology-occupation trade-off. I also consider two polar cases with respect to it, defined below, and next show under which conditions they emerge.

**Definition 2: Segmentation by Occupation**

Talent is segmented by occupation if \( x' > x \) and \( \omega(x) < 1 \) imply \( \omega(x') = 0 \).
**Definition 3: Segregation by Technology**

Talent is segregated by technology if \( x' > x \) implies \( \alpha(x', m^{-1}(x')) \geq \alpha(m(x), x) \) with probability 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. When talent is segregated by technology, more skilled individuals use a higher technology than lower skilled ones. When \( \alpha(x', x) \) strictly increases in both his arguments, this requires that \( m(x) \to x \). This second case is illustrated in Figure 1b.

**Corollary 1: Conditions for Segregation and Segmentation**

If \( \alpha \) and \( f \) satisfy
\[
\frac{\alpha(x,y)}{\alpha(y,z)} < \frac{f_t(x,y)}{f_t(y,z)} \quad \forall (x, y, z) \in [0, 1] \times [0, 1] \times [0, 1], \text{ with } x > y > z,
\]
then talent is segmented by occupation and \( m(x) = \frac{1}{2} \). If \( \alpha \) satisfies
\[
\frac{\alpha(x,y)}{\alpha(y,z)} \to \infty \quad \forall (x, y, z) \in [0, 1] \times [0, 1] \times [0, 1], \text{ with } x > y > z,
\]
then talent is segregated by technology and \( m(x) \to x \).

**Proof.** See appendix. \( \square \)

Proposition 2 bounds the concentration of talent, while Corollary 1 shows two limit cases of it. It is natural to wonder how wide are the bounds provided. The answer depends on the specific choice of the functional form. To explore whether the bounds can be informative, I provide here a numerical example for simple functional forms. Let \( f(x', x) = x' (1 + \lambda x) \), with \( \lambda \leq 1 \), and
\[
c(a) = \frac{a^{1+\eta}}{1+\eta}, \text{ with } \eta \geq 0, \text{ which implies } \alpha(x', x) = (x' (1 + \lambda x))^{\frac{\eta}{c(a)}}.
\]
This system of functional forms has two parameters: lower \( \lambda \) generates stronger skill asymmetry, and lower \( \eta \) generates larger technology gaps. In Figure 2 I plot the upper and lower bounds for the median individual \( x = \frac{1}{2} \), as a function of \( \eta \) and for two different values of \( \lambda \). At least in this case, the bounds are informative. We can also note that they converge to the values provided in Corollary 1.

---

13To see why this is the case, consider that if \( \exists x \in (x, m(x)) \), and \( \hat{x} \) is a worker, then \( \alpha(m(\hat{x}), \hat{x}) > \alpha(m(x), x) \), thus violating the definition of segregation by technology. If instead \( \hat{x} \) is a manager, then \( \alpha(m(x), x) > \alpha(\hat{x}, m^{-1}(\hat{x})) \), again violating the definition.
This section has characterized how the optimal technology choice and the allocation of talent are linked. The results are characterized in terms of one property of \( \alpha \), the technology gap, rather than in terms of the primitive \( c(a) \).\(^{14}\) In the next section, I fill this gap, and describe a technological environment from which I derive a cost of technology for each country and shows how it varies depending on their level of development.

Figure 2: Lower and Upper Bounds of Ability Gap between Managers and Workers

3 Cross-Country Differences in the Allocation of Talent

In this section, I describe a technological environment and derive from it a functional form for the cost of technology that is country specific. I characterize the properties of the cost function and obtain sharp predictions for the relationship between the distance of a country from the technology frontier and its allocation of talent.

3.1 Distance to the Technology Frontier and Cost of Technology

Recall that output of a team \((x', x)\), gross of the cost of technology, is given by \( af(x', x) \). The technology \( a \) is a multiplicative productivity term. I call it technology to distinguish it from labor productivity, that is simply defined as output divided by labor. More specifically, I think of \( a \) as a reduced form term that captures two separate elements that define how productive a team is: (i) the technology vintage that a team uses and (ii) the capital intensity at which the specific technology vintage is operated. Let me make a trivial example. In order to grind one hundred

\(^{14}\)If the reader prefers to see the results in terms of primitives, he can let \( c(a) = a^{1+\eta} \), and then Proposition 2 and Corollary 1 can be rewritten in terms of \( \eta \): (i) both bounds are increasing in \( \eta \); (ii) if \( \eta \to 0 \), \( m(x) \to x \); (iii) if \( \eta > \eta_1 \), with \( \eta_1 \) being a constant bigger than 1, then \( m(x) = \frac{1}{2} \). The technology gap is closely linked to the elasticity of the cost of technology, that in this simple functional form is given by \( \eta \).
pounds of grain into flour in an hour, a team may either use five slow horse powered mills, or one faster electric mill. A team may achieve the same level of technology/productivity $a$ - one hundred of pounds of grain per hour - with either many units of an old vintage of technology or fewer units of a newer vintage of technology. According to this interpretation, the cost of achieving productivity $a$ needs to take into consideration (i) the optimal choice of technology vintage, and the fixed cost associated with using it, that could derive for example from the cost of learning the blueprint of the specific vintage; and (ii) the amount of capital of that technology vintage that is necessary to achieve $a$. Following this view, I next introduce a technological environment and its associated cost function $c(a)$.

**The Technological Environment.** The most advanced vintage of available technology is $\bar{t}$. $\bar{t}$ represents the technology frontier at a given point in time. Each country is indexed by its level of development $t \leq \bar{t}$. All individuals in country $t$ know how to use technology vintage $t$, hence they do not need to pay any fixed cost to use it. I will refer to $t$ as the local technology vintage. The distance of a country from the technology frontier is given by $\bar{t} - t$, that is the distance at a given point in time between that country level of development and world knowledge. Individual in each country can decide to import and learn how to use better vintages of technology, up the frontier one $\bar{t}$, paying a fixed cost. The lower the level of development of a country, the higher the fixed cost. More modern technology vintages allow to achieve the same level of productivity $a$ at a lower marginal cost, since they require less capital, but they have a higher associated fixed cost. There is thus a trade-off that leads to optimal choice of technology vintage.

**Cost of Technology.** I next derive a functional form for the cost of technology $a$ in country $t$ when the frontier is $\bar{t}$. The choice of the functional form has been dictated by the need to captures in a parsimonious way the technological environment just described, while allowing for tractable solution of the optimal choice of technology and technology vintage.

The cost of achieving productivity level $a$ in country $t$ using the local technology vintage is

$$
c_L(a; t) = \gamma^{-\eta t} \eta \left( a - a_t \right)^{1+\eta} 1 + \eta
$$

where $\gamma$ captures the efficiency difference between each technology vintage, $\eta$ captures the within vintage cost elasticity of technology, and $a_t = \nu \gamma^t \nu \in [0, 1]$ is a non-homotheticity term that becomes useful in the limit case described below, but in general can be assumed to be 0. Individuals already know how to use the local vintage - ther is no learning fixed cost - hence we should think at $c_L(a; t)$ as the variable cost component linked to the chosen capital intensity. $\kappa_1$ is a constant term that scales the cost of technology. I pick $\kappa_1 \equiv \frac{\eta e - 1}{\eta e}$ as a function of the other parameters to guarantee that the marginal type that decides not to use the local technology, does not depend on either $t$ or $\bar{t}$. While restrictive, this provides tractability.

The cost of achieving productivity level $a$ in country $t$ using an imported technology vintage
\( \tilde{t} \) is given by

\[
c_I (a, \tilde{t}; t) = \kappa_1^\eta c_L (a; \tilde{t}) + \kappa_2^{-(\eta \varepsilon - 1)} \gamma^\varepsilon n(\tilde{t} - t) \varepsilon (1 + \eta)
\]

where \( \kappa_2^{-(\eta \varepsilon - 1)} \gamma^\varepsilon n(\tilde{t} - t) \) is the fixed cost associated with learning technology vintage \( \tilde{t} \), \( \varepsilon \) captures the elasticity of the cost of technology across vintages, \( \kappa_2 \equiv \chi_0^{-\frac{\eta + 1}{\eta - 1}} \) is defined so that the marginal team that decides to import technology rather than using the local one is given by \( \chi_0 \).

The fixed cost is increasing in the gap between current level of development and the vintage of technology \( \tilde{t} \) to capture that individuals in less developed countries may have a harder time learnings, using and extracting proper returns from more advanced technologies.\(^{15}\)

Last, the cost of achieving productivity level \( a \) in country \( t \), taking into consideration the optimal vintage choice is given by

\[
c (a; t, \tilde{t}) = \min \left\{ c_L (a; t), \min_{\tilde{t} \leq t} c_I (a, \tilde{t}; t) \right\}.
\]

In order for this problem to have an interior solution, I assume that \( \eta \varepsilon < 1 \).

---

\(^{15}\)This assumption is coherent with the view that frontier technologies are targeted to the level of development of frontier countries, and thus might be costlier to use for less developed countries. See for example Acemoglu and Zilibotti (2001).
Optimal Technology. I next show that this system of functional forms provides simple solutions for the optimal technology $\alpha$ and satisfies properties that are consistent with the existing empirical evidence on technology usage across countries.

The cost function $c(a; t, \bar{t})$ is the lower envelope of the costs of technology for each vintage. More modern vintages have lower marginal cost of productivity, but a higher fixed cost, as a result, as can be seen in Figure (3) teams that choose a higher productivity minimize costs by choosing a more advanced vintage. The next three Lemmas summarize the cost of technology, the resulting optimal technology choice, and its properties.

Lemma 5: Cost of Technology
The cost of a technology $c(a; t, \bar{t})$ in country $t$ when the frontier is $\bar{t}$ is given by

$$c(a; t, \bar{t}) = \begin{cases} \gamma^{-\eta} \kappa_1^{-\eta} \frac{(a-a_1)^{1+\eta}}{1+\eta} & \text{if } a \leq \hat{a}_{I,t} \\ \gamma^{-\eta} \kappa_2^{-\eta} \frac{(a-a_2)^{1+\eta}}{1+\eta} & \text{if } a \in (\hat{a}_{I,t}, \bar{a}_{I,t}) \\ \gamma^{-\eta} \frac{(a-a_1)^{1+\eta} + \kappa_2^{-\eta} \gamma^{2+\eta}(\bar{t}-t)}{2+\eta} & \text{if } a \geq \bar{a}_{I,t} \end{cases}$$

where $\eta \equiv \frac{\eta_{c-1}}{\eta+1} < \eta$, $\hat{a}_{I,t} = a_t + \gamma^t \kappa_1 x_{t,0}$ and $\bar{a}_{I,t} = a_t + \gamma^t \kappa_1 \chi_0^{1+\eta}$. Proof. See appendix. \square

Lemma 6: Optimal Technology Choice
The optimal technology choice of a team $(x', x)$ in country $t$ when the frontier is $\bar{t}$ is given by

$$\alpha (x', x; t, \bar{t}) = \begin{cases} a_t + \gamma^t \kappa_1 x_{t,0}^{\frac{1}{\gamma}} & \text{if } f(x', x) \leq \chi_0 \\ a_t + \gamma^t \kappa_2 x_{t,0}^{\frac{2+\eta}{\eta+1}} & \text{if } f(x', x) \in (\chi_0, \chi_1(t, \bar{t})) \\ a_t + \gamma^t x_{t,0}^{\frac{1}{\gamma}} & \text{if } f(x', x) \geq \chi_1(t, \bar{t}) \end{cases}$$

where $\chi_1(t, \bar{t}) = \chi_0^{\frac{\eta(\bar{t}-t)}{\eta+1}}$. Also, teams with $f(x', x) \leq \chi_0$ use the local vintage and teams with $f(x', x) \geq \chi_1(t, \bar{t})$ use the frontier vintage. Proof. See appendix. \square

Lemma 7: Properties of Optimal Technology
The optimal technology choice of a team $(x', x)$ in country $t$ when the frontier is $\bar{t}$ satisfies the following properties

1. individuals in countries closer to the frontier use better technology: $\forall (x', x, t) \alpha (x', x; t', \bar{t}) \geq \alpha (x', x; t, \bar{t})$ if and only if $t' \geq t$

2. for a given level of development, the more advanced is the frontier, the higher the technology used: $\forall (x', x) \alpha (x', x; t, \bar{t}') \geq \alpha (x', x; t, \bar{t})$ if and only if $\bar{t}' \geq \bar{t}$
3. the technology gap between teams depends only on the distance from the frontier: \( \frac{\alpha(x,y;\bar{t})}{\alpha(y,z;\bar{t})} = \frac{\tilde{\alpha}(x,y;\bar{t}-t)}{\tilde{\alpha}(y,z;\bar{t}-t)} \), where \( \tilde{\alpha}(x',x;\bar{t}-t) = \alpha(x',x;\bar{t}-t) \gamma^{-t} \);

4. the cutoff \( \chi_1(t,\bar{t}) \) increases in the distance from the frontier;

Proof. See appendix. □

The first two properties are reassuring. In more developed countries everyone should use a more advanced technology, and at the same time, the presence of modern technology vintages at the frontier benefits all countries. The third property shows that is not the absolute level of development that matters for the technology gap, but rather the distance from the technology frontier. The fourth property says that, everything else equal, the number of individuals using modern vintage is increasing in the level of development of a country. Nonetheless, also some teams in less developed countries use the most advanced vintage available. This property of the cost function is consistent with the fact that most modern technologies are used even in the poorest regions of the world, with the main difference across countries being that in rich ones many more individuals use them.\(^{16}\) Cellphones and computers are two salient example.

Last, and most importantly, this technological environment implies that in countries far from the frontier the technology gaps across teams are higher.

**Proposition 3: Technology Gap and Distance to the Technology Frontier**

The optimal technology function \( \alpha(x',x;\bar{t}) \) is such that the technology gap is higher further from the technology frontier: \( \forall (x,y,z) \in [0,1] \times [0,1] \times [0,1] \) such that \( x \geq y \geq z \), \( \frac{\alpha(x,y;\bar{t})}{\alpha(y,z;\bar{t})} \geq \frac{\alpha'(x,y';\bar{t}')}{\alpha'(y,z';\bar{t})} \) if and only if \( \bar{t} - t \geq \bar{t}' - t' \).

Proof. See appendix. □

The proposition says that if we take any two teams, the gap in their optimal technology choice is larger the further a country is from the technology frontier. The possibility in countries far away from the frontier to choose whether to import more advanced technology vintages from abroad leads naturally to larger gaps in optimal technology across teams. This result emerges because not everyone finds it optimal to choose the same technology vintage. At the frontier instead, all individuals use the same vintage, the frontier one, since there is nothing better available.

Proposition 3, together with Proposition 2 in the previous section, implies that the concentration of talent is stronger in countries far from the frontier.

**Corollary 2: Assignment of Talent and Distance to the Technology Frontier**

Both the upper and lower bound of the ability gap between a worker and his manager are decreasing in the distance to the technology frontier.

Proof. See appendix. □

\(^{16}\)A more refined, and empirically documented, version of this argument is present in the work of Comin and Mestieri (2013).
Lemma 8 concludes the description of the properties of this technology function and shows that at any point in time, we can use the GDP per capita of a country as a proxy of its distance from the technology frontier.

**Lemma 8: Distance to the Technology Frontier and GDP per capita**

The GDP per capita of a country is given by

\[
Y(t, \bar{t}) = \int_0^1 g(\alpha(m(x), x; t, \bar{t}), m(x), x) \omega(x) \, dz
\]

and satisfies

\[
Y(t, \bar{t}) = \gamma \tilde{Y}(\bar{t} - t)
\]

where \( \frac{\partial \tilde{Y}(\bar{t} - t)}{\partial (t - \bar{t})} < 0 \).

**Proof.** See appendix. □

### 3.2 A Tractable Case

For the general functional forms, I could only characterize the equilibrium in terms of bounds to the matching function. I here consider a functional form for \( g(a, x', x) \) that, combined with a limit case for \( c(a; t, \bar{t}) \), allows a complete analytical solution of the equilibrium.

I use

\[
g(a, x', x) = ax' (1 + \lambda x) - c(a; t, \bar{t}) (1 + \lambda x),
\]

where \( \lambda \leq 1 \), and \( c(a; t, \bar{t}) \) is the cost function defined in the previous section, with \( a_t = \gamma_t \) and in the limit case when \( \eta \to \infty \) and \( \varepsilon \eta \to 1 \), so that \( \varepsilon \to 0 \).\(^{17}\) The within vintage cost elasticity of technology \( (\eta) \) goes to infinite: conditional on a given technology vintage, all teams would pick identical productivity \( a \). The vintage cost elasticity \( (\varepsilon) \) goes instead to 0: conditional on deciding to use a foreign vintage, everyone would choose the frontier one. As a result, only two technologies are chosen. The functional form of \( g \) gives two other convenient properties: (i) the optimal technology choice depends only on the ability of the manager; (ii) the marginal value of either managers and workers depend only on their partner ability and not on their own, that is, \( \forall (x', x'') \) \( f_1(x', x) = f_1(x'', x) \) and \( f_2(x', x) = f_2(x'', x) \). Notice that, in equilibrium, your own type \( x \) does affect the marginal value of your skills, but only trough the matching function that assigns more skilled managers to more skilled workers due to complementarity (Lemma 2).

Lemmas 1-4 and Proposition 2 holds for this functional form. Lemma 9 characterizes the optimal choice of technology.

\(^{17}\)Two features of \( g \) are worth of notice. First, \( f(x', x) = x' (1 + \lambda x) \) satisfied the Spence-Mirlees assumption described in Subsection (2.1) as long as \( \lambda \leq 1 \). Second, I am departing from the production function defined in (2.1) to the extent that I am now allowing the cost of technology, here given by \( c(a; t, \bar{t}) (1 + \lambda x) \) to depend on the type of workers. As I discuss below, this change is convenient for tractability. All the results of Section (2) holds for this functional form. Results are available upon request.
Lemma 9: Optimal Technology for the Limit Case
The optimal technology for a team \((x', x)\) is given by

\[
\alpha(x', x) = \begin{cases} 
\gamma' & x' < \chi_0 \\
\gamma' + \gamma & x' \geq \chi_0 
\end{cases}
\]

Proof. See appendix. □

Lemma 9 implies that \(\frac{\alpha(m(x), x)}{\alpha(x, m^{-1}(x))} \in \{1; 1 + \gamma^{\bar{t} - t}\}\): the technology gap is equal to 1 if an individual would use the same technology if a manager or a worker, or equal to \(1 + \gamma^{\bar{t} - t}\) if an individual is itself less skilled than \(\chi_0\), but by being a worker is matched with a manager \(m(x) \geq \chi_0\), thus getting access to the frontier technology vintage. This functional form allows to solve analytically the equilibrium.

Talent Segregation and Segmentation. Let me first consider the conditions that guarantee talent segmentation by occupation - i.e. higher skilled are managers - and talent segregation by technology - i.e. higher skilled use the frontier technology vintage. The two cases can be seen in Figures (4a) and (4e) below. As expected from an application of Corollary 1, if a country is close enough to the frontier talent is segmented, while if it is far enough talent is segregated.\(^{18}\)

Lemma 10: Segregation and Segmentation in the Limit Case
There exists two constants, \(d_1 < d_2\), such that: (i) if \(\bar{t} - t \leq d_1\), talent is segmented by occupation; (ii) if \(\bar{t} - t \geq d_2\), talent is segregated by technology. The two constants are known functions of parameters: \(d_1 = \frac{\log(1 - \lambda) - \log(1 - \chi_0)}{\log \gamma}\) and \(d_2 = \frac{\log(2 - \lambda(1 - \chi_0)) - \log(\lambda(1 - \chi_0))}{\log \gamma}\).

Proof. See appendix. □

The cutoffs for segmentation and segregation depend on the skill-sensitivity asymmetry parameter \(\lambda\). The higher is \(\lambda\), thus the lower the difference in skill-sensitivity between managers and workers, the closer a country has to be to the frontier to have talent segmentation, and not to have talent segregation.

Intermediate Cases. I next consider the intermediate cases when \(\bar{t} - t \in (d_1, d_2)\) and thus talent is nor segmented nor segregated.

Proposition 4: Matching Function in the Limit Case
For any worker \(x\) the distance between his ability and the one of his manager, \(m(x) - x\), decreases in the distance from the frontier, \(\bar{t} - t\).

Proof. See appendix. □

\(^{18}\)There is one main difference with respect to Corollary 1 in Section 2: talent segregation is obtained without the need for the technology gap to go to infinite. This result is obtained because, in this limit case, the optimal technology function, \(\alpha(x', x)\) is not differentiable. Also notice that this result is not in contradiction to Corollary 1, since Corollary 1 provides sufficient but not necessary conditions for segregation and segmentation.
Proposition 4 is the equivalent of Proposition 2 for this tractable case. It is here possible not only to bound the ability gap between managers and workers, but to solve explicitly for the unique equilibrium, and characterize analytically the matching function. The solution of the equilibrium is left to Appendix B.3, and entails two main steps. First, I show that the optimal allocation can take only one of five possible shapes, shown in Figure (4), 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. Second, I show that the slope of the matching function \( m(x) \) depends on the fraction of workers at \( x \), \( \omega(x) \), and fraction of managers at \( m(x), 1 - \omega(m(x)) \). This allows to solve for the \( \omega \) that guarantees that the marginal values of managers and workers are identical, a condition required over regions of the type space in which the optimal allocation dictates mixing, as in Figure 4c.\(^{19}\)

Figure 4: Allocation of Talent
(a) Segmentation (b) Intermediate Case 1 (c) Intermediate Case 2
(d) Intermediate Case 3 (e) Segregation

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 vintage. The red striped regions are present in mixing area on which the workers use the frontier technology, while the managers use the local one. Dotted lines connect examples of workers and managers that are together in a team.

As the distance from the frontier increases, the economic structure changes smoothly, in contrast with most frictionless matching models that feature discrete jump between two polar cases.\(^{20}\) Consider the case with segmentation of talent in Figure 4a. When a country is sufficiently close to the technology frontier, the gap between the frontier and the local technology is small, thus every team uses a similar technology and high skilled individuals are assigned to be managers. As \( \tilde{t} - t \) increases, it becomes more skill rewarding for an individual to become a worker and get access to the frontier technology, rather than being a manager and use the local one. The reward

\[^{19}\] A similar insight is present in Low (2013) for a case with bipartite matching.

\[^{20}\] 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)).
from being a manager rather than a worker depends also on the production partner, and thus on
the matching function \( m \). At first, only the lowest skilled among the managers finds it optimal to
become workers. As \( \bar{t} - t \) increases further, more and more individuals who, if they were managers
would use local technology, become workers in order to get access the frontier one. When \( \bar{t} - t \) is
sufficiently large, access to the frontier technology drives the assignment. The optimal allocation
thus resembles a dual economy: within each technology, there is talent segmentation, but skills
are segregated by technology.

### 3.3 Interpretation and Discussion

Before moving forward, I summarize the main insights of the theoretical analysis.

The model of Section 2 has proved that the equilibrium allocation of talent and the cross-
sectional distribution of technology are inherently intertwined. The technological environment,
and related cost function, introduced in Section 3 has shown that the possibility of less developed
countries to adopt frontier technology vintages naturally leads to larger technology dispersion and
consequently to a different allocation of talent. Specifically, in countries close to the technological
frontier, most teams use similar technology, and 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 specialize in the most skill-sensitive task. As a result, all teams
are fairly similar and there is low productivity dispersion across them. In countries far from
the technological frontier, instead, the allocation is asymmetric. Some teams attract skilled
individuals (both managers and workers) and use frontier technologies. Some other teams instead
are left with low skilled ones and use traditional technologies. Teams now concentrate similarly
skilled individuals to reap the benefits from the complementarity between skills and technology.
As a result, there is larger dispersion of talent, technology, and productivity in the economy.

In the limit case depicted in Figure 4e, the possibility to adopt frontier technology leads to
an endogenous formation of a dual economy in poor countries. Teams that adopt advanced
technology attract the most skilled individuals, leaving the rest of the economy with low talent
and thus lower productivity.\(^{21}\) In some cases, it is even possible that some individuals would
use a higher technology in autarchy then when countries gain access to the frontier, due to the
polarization of talent generated by the possibility of technology adoption. In fact, when talent
concentrates, low skilled workers are matched with lower ability managers, and thus - ceteris
paribus - would use a lower technology.

The model predictions on the economic structure in developing countries are qualitatively con-

\(^{21}\)This feature of the model resembles a mechanism outlined by Acemoglu (2015) (page 454) 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.”
sistent with a large body of empirical evidence. The larger productivity dispersion in poor countries has been noted among others by Caselli (2005), Hsieh and Klenow (2009), and Adamopoulos and Restuccia (2014). The model also predicts that in developing countries some very low skilled individuals are employed in managerial positions. Bloom and Van Reenen (2010) has shown the existence of a thick left tail of poorly managed firms and that firms with more educated manager 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)). Most existing theories that provide an explanation for these empirical facts attributes them to larger market frictions in developing countries. In the context of this paper, instead, they emerge as a result of differences in endowment that lead to differences in optimal allocations.\footnote{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.} This paper also departs from previous literature in linking these previously documented cross-country differences to different assignment of individuals to teams. Differences in the allocation of talent are a new feature of economic development, that has been previously overlooked. In Section 4, I show that it has empirical content.

**Cross-Country Differences in Ability Distribution.** As a last, but important, remark, I discuss the assumption that all countries have ability identically distributed, as $x \sim U[0, 1]$. This seems in contrast with the abundant empirical evidence that average schooling years are lower in poor countries. I intend $x$ to capture the relative ability rank within a country, thus comparable only within countries and not across them. The reason for this choice, is that there is an intrinsic isomorphism between the level of ability $x$ and the cost of technology. A higher ability is isomorphic to a lower cost of technology. Let me show an example. Let $h^t$ be a human capital term, that captures the average ability of a country with level of development $t$. Also, consider for simplicity the choice of technology for the tractable case of Section 3.2. Keeping constant the cost of technology, and letting ability change, is identical to keep ability fixed, and let the cost of technology change by a properly scaled factor:

$$\max_a ax h^t - \frac{a^{1+\eta}}{1+\eta} = \max_a ax - \frac{1}{h^{(1+\eta)t}} \frac{a^{1+\eta}}{1+\eta}.$$  

For this reason, the cross-country differences in the local technology vintage can be interpreted as cross-country differences in the level of human capital. I charged all cross-country differences on the cost of technology for the sake of clarity.\footnote{Allowing the distribution of ability to vary across countries would be more problematic. The reason being that the distribution of ability impacts the matching patterns. However, I don’t have any fundamental reason to think that the ability distribution should be a primitive that changes across countries, rather I show in previous version of this paper (See Porzio (2016)) that if individuals are allowed to invest in their ability, for example
4 Empirical Evidence on the Allocation of Talent

In this section, I provide evidence to support the main empirical prediction of the model: the concentration of talent is higher the further a country is from the technological frontier.

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. To directly compute the measure of concentration of talent defined in the model, we would need to observe the ability of all individuals in the economy and their production partners. Additionally, we would need such data to be comparable for several countries around the world. Unfortunately, to the best of my knowledge, such data simply does not exist.\textsuperscript{24} I take therefore an indirect approach that exploits one of the assumption of the model, the complementarity between skills and technology, which implies that more able teams use a more advanced technology. Observing the average ability of individuals that use each technology becomes then sufficient to make inference on the matching function, and thus the concentration of talent. Let me make an example using the tractable case, when only two technologies are used: the local and the frontier ones. The complementarity assumption implies that the most skilled teams use the frontier technology. Hence, among the managers, the most able ones use the frontier technology. The same for workers. Consider first the case when talent is segmented by occupation, hence when talent concentration is low. All managers are high skilled, and some of them use the local technology. As a result, the ability gap across technologies is small. Consider next the case when talent is instead segregated by technology, hence when talent concentration is high. All most skilled individuals now use the frontier technology, some of them being managers other being workers. Only the lowest skilled individuals are left using the local technology. The ability gap across technologies therefore is now large. This same insight holds for intermediate cases as well.

Concretely, I use censuses and labor forces surveys from several countries around the world and over time. In this data, I observe individuals education, that I use as a proxy of ability, and the industry in which an individual works, that I use as a proxy of the technology he uses. I discuss potential concerns of this strategy after providing details on the data and on the exact construction of the empirical measure of concentration of talent.

4.1 Data

I use 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. For most countries, the datasets have very large sample. Merging all countries and years together, there are more than 600 millions individuals through schooling, then the distribution of schooling will depend on the matching pattern and the distribution of technology. The stronger concentration of talent and dispersion of technology in poor countries implies a larger cross-sectional dispersion of education. Consistent with what we observe in the data.

\textsuperscript{24}Matched employer-employee dataset are available for few countries around the world. However, for less developed countries they are not representative of the whole economy.
in the data. 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, own-account 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 4.4. Data details are in Section E.

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 F, I perform robustness checks to address this concern.

4.2 Measure of Concentration of Talent

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 \( F, \) I don’t use any country subscript to ease notation): \( \hat{x}_i = F(s_i), \) where \( s_i \) is the schooling year of individual \( i \).\(^{25}\) Second, I compute the average skill in each industry \( j \): \( \bar{x}_j = E[\hat{x}_i | I_{ij} = 1], \) 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. According to the model assumption of skill-technology complementarity, 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

\(^{25}\)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.
the highest skilled ones among the remaining workforce are then assigned into the second highest and so on.\textsuperscript{26} Fourth, I compute the average skill in each industry under the perfect sorting counterfactual: \( \tilde{p}_j = E \left[ \tilde{x}_i | I_{Cj}^j = 1 \right] \), where \( I_{Cj}^j \) is the constructed indicator function. Fifth and last, I regress \( \tilde{x}_j = B_0 + B_1 \tilde{p}_j + \varepsilon \).\textsuperscript{27} The measure of concentration of talent is \( \hat{\pi} = \hat{B}_1 \). By the definition of the least squares estimator, \( \hat{\pi} = \frac{E[\tilde{x}_j - \tilde{\pi}_j]}{\tilde{p}_j - \tilde{p}_j'} \). This measure thus capture the expected ability gap across industries relative to the benchmark case in which workers sort perfectly across industries based on their ability, as when there is segregation by technology.

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, \( \tilde{x}_j \), as a function of the skill in the perfect sorting counterfactual, \( \tilde{p}_j \). Each dot in the figure corresponds to an industry and its size increases in the number of individuals there employed. A linear regression \( \tilde{x}_j = \alpha + \pi \tilde{p}_j + \varepsilon \) fits the data well.\textsuperscript{28} 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 build this measure of concentration of talent for each country-year

\textsuperscript{26}More formally, let \( \tilde{X}_j \) be the observed set of individuals in industry \( j \), of mass \( v(\tilde{X}_j) \), then \( \tilde{x}_j = E \left[ \tilde{x} | \tilde{x} \in \tilde{X}_j \right] \). The counterfactual sets are given by \( \tilde{P}_j \equiv \left\{ \tilde{x} : \tilde{x} \in [\tilde{P}_1(j), \tilde{P}_2(j)] \right\} \) where \( \tilde{P}_1(j) \equiv \sum_{k} \tilde{x}_k < \tilde{x}_j \) and \( \tilde{P}_2(j) \equiv \tilde{P}_1(j) + v(\tilde{X}_j) \). Then \( \tilde{y}_j = E \left[ \tilde{x} | \tilde{x} \in \tilde{P}_j \right] \).

\textsuperscript{27}I weight the regression by the number of individuals in each industry \( j \). Unweighted results are similar and available upon request.

\textsuperscript{28}The average \( R^2 \) across countries from this regression is \( \sim 0.9 \) similarly in rich and poor countries.
pair in my sample and document systematic differences between countries depending on their distance from the frontier.\textsuperscript{29}

**Link with Model Definition of Concentration of Talent.** I next discuss more formally how the empirical concentration of talent is linked to the model. I here show that, under few assumptions, the empirical measure of concentration of talent maps exactly into the model one. Assume that we observe a continuum of industries. Let $\tilde{a}$ be the technology rank of an industry, by construction, $\tilde{a} \sim U [0, 1]$. Let $\bar{x}(\tilde{a})$ be the average ability of individuals in $\tilde{a}$. Due to the definition of $\tilde{a}$, counterfactual average ability is equal to $\bar{p}(\tilde{a}) = \tilde{a}$. The empirical measure of concentration of talent is given by the slope in the regression of $\bar{x}(\tilde{a})$ on $\tilde{a}$. Next, assume that the average ability of workers in $\tilde{a}$, call it $x_w(\tilde{a})$ can be well approximated by a linear function with unknown slope: $x_w(\tilde{a}) = \beta \tilde{a}$. Last, assume that the ability gap between managers and workers is not correlated with technology. Under this set of assumptions, the empirical measure of concentration of talent as previously defined would be exactly equal to $\beta$. Let’s solve for $\beta$. Market clearing implies that

$$\frac{1}{2} \int_0^1 x_w(\tilde{a}) d\tilde{a} + \frac{1}{2} \int_0^1 x_m(\tilde{a}) d\tilde{a} = \int_0^1 x dx$$

$$\frac{1}{4} \beta + \frac{1}{2} \int_0^1 [x_m(\tilde{a}) - x_w(\tilde{a})] d\tilde{a} = \frac{1}{2}$$

$$\beta = 1 - \int_0^1 (m(x) - x) \omega(x) dx$$

where I used the fact that by market clearing and by the definition of $m(x)$, $\int_0^1 x_w(\tilde{a}) d\tilde{a} = \int_0^1 x \omega(x) dx$, and $\int_0^1 x_m(\tilde{a}) d\tilde{a} = \int_0^1 m(x) \omega(x) dx$. The empirical measure of concentration of talent is thus - in this case - exactly identical to the model one.

**Discussion of Empirical Strategy.** The empirical strategy is sound if two assumptions hold. 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.\textsuperscript{30} There

\textsuperscript{29}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.

\textsuperscript{30}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.
are few concerns. First, education might measure skill with white noise. 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, education might be less correlated with skills in poor countries. For example, if some talented individuals are credit constrained and hence do not go to school even if they would have a high return from it. First, if this lower of correlation can be mapped into higher white noise, this would not constitute a problem. Otherwise, I discuss in 4.4, that this would likely attenuate the 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 of course consistent with the model prediction that more skilled teams sort into higher production technologies. 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. 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 possibility in Section 4.4.

4.3 Results

I next compare the empirical measure of concentration across countries and show that it varies systematically as a function of the distance to the technology frontier. A straightforward application of Lemma 8 shows that there are three possible cross-country comparison to identify differences in distance from the frontier. First, keep constant the level of the frontier. Comparing countries in the same year, the poorer ones are further from the frontier. Second, keep constant the level of GDP per capita. Comparing one country in the past with another one today, thus facing a more advanced frontier, the latter is further from the frontier. Third, if we follow two countries over time, the one that grows faster is approaching the frontier.

1st Comparison: Poor and Rich Country Today. First, I compare countries with higher and lower level of GDP per capita by plotting the concentration of talent as a function of GDP

\[ Y(t, \bar{t}) = \gamma Y(\bar{t}) \]

where, \( \frac{\partial Y(t, \bar{t})}{\partial (t-\bar{t})} < 0 \).
per capita relative to one of the United States in 2010. The result is shown in Figure 6: poor countries, i.e. those farther from the technology frontier, have larger concentration of talent. 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. In Section 5 I further explore the quantitative properties of the model and show that these differences are sizable.

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

![Graph showing concentration of talent vs. GDP per capita relative to the U.S. in 2010.](image)

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.

2nd Comparison: Poor Countries Today and U.S. in the Past. 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

---

34 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.

35 The regression line has a positive slope that is significant at 1% level.
frontier. Specifically, I have comparable census data for the United States every ten years from 1940 to 2010.\(^{36}\) 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. \(^{37}\) 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.\(^{38}\)

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

![Figure 7](image-url)

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.

3\(^{rd}\) **Comparison: South Korea Convergence to 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%.\(^{39}\) In Figure 8, I plot the growth path for concentration of talent across sectors for both countries.\(^{40}\) 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.

---

\(^{36}\) Before 1940 census data did not report education years.

\(^{37}\) 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.

\(^{38}\) A permutation test of the null hypothesis that the U.S. is not different rejects the null hypothesis more than 99% of the time.

\(^{39}\) These facts can be appreciated in Figure A.5.

\(^{40}\) 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.
4.4 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.\footnote{Classic measurement error does not bias the results, as long as it averages to zero at the sectoral level.} 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.\footnote{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.}

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.
Table 1: Robustness Table

<table>
<thead>
<tr>
<th></th>
<th>Point Estimate</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Baseline</td>
<td>-0.0708</td>
<td>36.3%</td>
</tr>
<tr>
<td><strong>Level of Industry Aggregation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) Sectors (Agr, Mfg, Ser)</td>
<td>-0.0861</td>
<td>44.0%</td>
</tr>
<tr>
<td>(3) Unharmonized Industries</td>
<td>-0.0472</td>
<td>17.3%</td>
</tr>
<tr>
<td><strong>Sample Selection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Include Non Households Heads</td>
<td>-0.0681</td>
<td>35.5%</td>
</tr>
<tr>
<td>(5) Include Women</td>
<td>-0.0719</td>
<td>37.9%</td>
</tr>
<tr>
<td>(6) Only Women</td>
<td>-0.0610</td>
<td>15.0%</td>
</tr>
<tr>
<td>(7) Include Self-Employed (Own-Accounts)</td>
<td>-0.0715</td>
<td>36.4%</td>
</tr>
<tr>
<td><strong>Role of Agriculture</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) Drop Agriculture</td>
<td>-0.1710</td>
<td>55.3%</td>
</tr>
<tr>
<td>(9) Only Individuals non in Agriculture</td>
<td>-0.0266</td>
<td>8.7%</td>
</tr>
<tr>
<td><strong>Measure of Concentration of Talent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10) Correlation</td>
<td>-0.0407</td>
<td>24.9%</td>
</tr>
<tr>
<td>(11) Correlation Using Normalized Skills</td>
<td>-0.0246</td>
<td>10.3%</td>
</tr>
</tbody>
</table>

Note: All Coefficients are significant at less than 1%, with the exception of rows (9) and (11) that are significant at less than 5%.

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). 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. 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. $\bar{x}_j = B_0 + B_1\bar{p}_j + \varepsilon$. 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 consider only individuals who are

---

43Examples of Brazil 2010 and United States 1940 are in Figures A.1 and A.2 in the Section A. We can appreciate that the nice fit of the measure of concentration of talent is present at any level of aggregation.

44The 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.
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, along the lines of the one used in Kremer and Maskin (1996), namely the correlation between individuals education and the average education in an industry, that is

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

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).

### 4.5 Firm Level Evidence

[FIRM LEVEL EVIDENCE USES WORLD BANK ENTERPRISE SURVEY - SEE FIGURES IN APPENDIX]

### 5 A Quantitative Exploration

As discussed in Section 3.3 the equilibrium in countries far from the technology frontier, despite being efficient, resembles several empirical facts that are usually attributed to the presence of large frictions in developing countries. The results of this paper thus suggests that cross-country exercises that compare micro-level allocations should be cautious in assigning all the observed differences to larger frictions in developing countries. Nonetheless, whether the mechanism this paper proposes is quantitatively relevant in explaining cross-country differences remains an open question.

In this section I take a first step towards this broader goal and ask whether the model can explain a sizable fraction of the larger agricultural to non-agricultural productivity gap documented in poor countries. I write down a version of the model that is amenable to a quantitative exploration of the data, I estimate it using data on the allocation of talent in rich and poor countries and the agriculture to non agriculture productivity gap in rich countries. I then show that, once fitted into the model, the observed cross-country differences in allocation of talent explain approximately 40% of the larger agricultural productivity gap in poor countries.

### 5.1 Quantitative Model

I add one feature to the model of Section 2. Each individual is characterized not only by his ability $x$ but also by a vector of technology specific shocks, which make him more willing to use some technologies rather than others. The introduction of this additional source of noise allows
to develop an efficient algorithm to solve the numerically cumbersome problem of Section 2.\textsuperscript{45} The model with technology specific Frechet shocks mirrors the one presented in Section 2. I here discuss how I introduce the shocks into the model and why they allow to efficiently compute the equilibrium.

**Adding Frechet Shocks.** Each individual $x$ chooses the technology that gives him the highest income. He thus solves

$$\max_{a \in \mathbb{A}} v_a y(a, x)$$

where $\mathbb{A}$ is a discrete approximation\textsuperscript{46} of the set of available technologies, $v_a$ is distributed according to a Frechet with dispersion parameter $\theta$ (dispersion of shocks increases as $\theta$ decreases)

$$v_a \sim e^{-\frac{1}{\theta} a},$$

and $y(a, x)$ is the income of an individual of ability $x$ that uses technology $a$ and picks his occupation optimally

$$y(a, x) = \max_{z \in [0,1]} z \pi(a, x) + (1 - z) w(a, x).$$

Managers choose the optimal worker among the set $\Omega_a$ that gathers the individuals that are using technology $a$:

$$\pi(a, x) = \max_{z \in \Omega_a} a f(x, z) - c(a) - w(a, z).$$

The Spence-Mirlees assumption (Assumption 3 in Section 2) guarantees that, as long as individual use the same technology, the most skilled ones are managers. Market clearing dictates that the mass of manager and workers for each technology should be identical. Combining the cutoff policy implied by the Spence-Mirlees assumption with market clearing allows to solve for the cutoff type $\hat{x}(a)$ that characterizes the occupational choice. Let $\phi(a, x)$ be the joint distribution of individuals over technology $a$ and ability $x$. The cutoff $\hat{x}(a)$ solves

$$\int_{0}^{\hat{x}(a)} \phi(a, x) \, dx = \int_{\hat{x}(a)}^{1} \phi(a, x) \, dx.$$

\textsuperscript{45}Why is the problem numerically cumbersome? Consider a typical matching problem in which you have $N$ males and $N$ females that must be matched. There are exactly $N!$ possible matches to be evaluated. Now consider the problem in which there are $2N$ individuals and we have to solve for optimal teams of two members. We now need to consider that we can generate $\frac{1}{2} \frac{(2N)!}{(N!)^2}$ possible partitions in two groups, and then we need to evaluate each possible matching for each partition. There is thus a total of $\frac{1}{2} \frac{(2N)!}{(N!)^2}$ possibilities to evaluate: several order of magnitude more than $N!$. The fact that individuals are not ex-ante restricted to be either managers or workers make the computation more complex.

\textsuperscript{46}I describe in details in the Appendix C how this set is constructed for the computation.
The income of an individual $x$ is then given by

$$y(a, x) = \begin{cases} w(a, x) & \text{if } x \leq \hat{x}(a) \\ \pi(a, x) & \text{if } x > \hat{x}(a) \end{cases}$$

The matching function within each technology, $m(a, x)$, is derived using the fact that, due to complementarity of $f$, there is positive assortative matching between managers and workers, as proved in Lemma 2. $m(a, x)$ solves

$$m(a, x) = \int_{\hat{x}(a)}^{x} \phi(a, z) \, dz = \int_{0}^{\hat{x}(a)} \phi(a, z) \, dz.$$

Wages and profits are calculated using the first order and envelope conditions of the manager problem, together with market clearing

$$\pi(a, m(x)) + w(a, x) = af(m(x), x) - c(a).$$

That is

$$w(a, x) = \kappa + \int_{0}^{x} w_2(a, z) \, dz$$

$$\pi(a, x) = w(a, \hat{x}(a)) + \int_{\hat{x}(a)}^{x} \pi_2(a, z) \, dz$$

where $\kappa$ satisfies

$$\int_{0}^{\hat{x}(a)} w(a, x) \phi(a, x) \, dx + \int_{\hat{x}(a)}^{x} \pi(a, x) \phi(a, x) \, dx = \int_{0}^{\hat{x}(a)} (af(m(a, x), x) - c(a)) \phi(a, x) \, dx.$$

The last equilibrium object to describe is the distribution of individuals over technologies, $\phi(a, x)$. The properties of the Frechet, as has been shown in the literature (see for example Hsieh et al. (2013)), imply that, $\frac{\phi(a, x)}{\sum_{a} \phi(a, x)}$ - i.e. probability that an individual $x$ selects into technology $a$ - is given by

$$\frac{\phi(a, x)}{\sum_{a} \phi(a, x)} = \frac{\tilde{y}(a, x)^{\theta}}{\sum_{a} \tilde{y}(a, x)^{\theta}}$$

where $\tilde{y}(a, x) = \max\{0, y(a, x)\}$. Last, the GDP per capita is

$$Y = \sum_{a} \int_{0}^{\hat{x}(a)} (af(m(a, x), x) - c(a)) \phi(a, x) \, dx. \quad (3)$$

**Computing Algorithm.** This model can be computed for any functional form of $f$ and $c$ that satisfies the assumptions in Section 2. The computing algorithm iterates on the distribution
\( \phi(a, x) \) until convergence. Details are in the Appendix C. The numerical solution is fast, thus allowing to estimate the model through simulated method of moments, as I discuss next.

5.2 Estimation

I choose a parsimonious functional form: 
\[
f(x', x) = x'(1 + \lambda x).
\]
\( \lambda \in [0, 1] \) modulates the strength of the gap in skill-sensitivity between managers and workers: 
\[
\frac{\partial}{\partial x} f_t(x, y) < 0 \forall (x, y, z).
\]
\( \lambda \leq 1 \) guarantees that the Spence-Mirlees condition is satisfied. I use the cost of technology \( c(a; t, \bar{t}) \) defined in Section 3. The resulting model has 8 parameters for each country: the skill-sensitivity parameter \( \lambda \), the dispersion of individual-technology shocks \( \theta \), the within vintage cost elasticity of technology \( \eta \), the across vintage cost elasticity \( \varepsilon \), the level of the cost to import technology \( \chi_0 \), the improvement across vintages \( \gamma \), the level of development of a country \( t \), and the level of development of the frontier \( \bar{t} \). I fix \( \gamma = 1.02 \), a 2% growth rate each year. The value of \( \gamma \) does not affect the results, since it is not separately identified from \( t \) and \( \bar{t} \). I fix \( \bar{t} \) to match the level of GDP of the U.S. for a country with \( t = \bar{t} \). The value of \( \bar{t} \) does not matter as well for the results, since - as shown in Section 3 - the distance from the frontier is a sufficient statistic for the allocation of talent and technology, and the absolute level of development of a country is irrelevant. I estimate the five remaining parameters \( \lambda, \theta, \eta, \varepsilon, \chi_0 \) by simulated method of moments. I simulate, using the Metropolis–Hastings algorithm, a chain that converges to the vector of parameters that minimizes the distance between the model and the data.\(^{47}\)

Data Moments. I first describe how I construct the targeted moments in the data. I divide countries in four quartiles according to their Real GDP per capita relatively to the one of the United States.\(^{48}\) The first quartile has GDP per capita 78% of the one of the U.S., the second one 21%, the third one 8.7%, and the fourth one 2.5%. For each quartile, I compute the average concentration of talent, measured as described in Section 4. I then use the average agriculture to non-agriculture productivity gap by quartile reported in Gollin et al. (2014). The five targeted empirical moments are the concentrations of talents by income quartiles, and the agricultural productivity gap for the first quartile.

Model Moments. I next describe how I construct the same moments in the model. For a given vector \( i \) of parameters \( \{\theta_i, \lambda_i, \eta_i, \varepsilon_i, \chi_{0i}\} \), I solve the model for 100 countries, where each country corresponds to one level of development \( t \). I then calculate, using equation (3), the GDP per capita of each country relative to the one of the U.S., i.e. \( t = \bar{t} \). I pick the four countries \( (t_1, t_2, t_3, t_4) \) that most closely represent each quartile of the income distribution - i.e. 
\[
\frac{Y(t_1)}{Y(\bar{t})} \simeq 0.78, \frac{Y(t_2)}{Y(\bar{t})} \simeq 0.21, \frac{Y(t_3)}{Y(\bar{t})} \simeq 0.087, \frac{Y(t_4)}{Y(\bar{t})} \simeq 0.025.
\]
For each country, I calculate the model generated concentration of talent using the same procedure that I used on the micro data in

\(^{47}\)For details on the simulated method of moments, see McFadden (1989). The simulation method that I use has been developed first by Chernozhukov and Hong (2003).

\(^{48}\)GDP per capita is taken from the Penn World Table version 8.0. I build GDP per capita as rgdpna/pop for the year 2010. To compute income quartiles, I use the same set of countries as Gollin et al. (2014).
Section 4. I keep the assumption that an industry is a technology, and calculate average ability of individuals that use each technology $a$

$$\bar{x}(a) = \frac{\int_{0}^{1} x \phi(a,x) dx}{\int_{0}^{1} \phi(a,x) dx}.\$$

I then calculate the counterfactual average ability under perfect sorting

$$\bar{p}(a) = \sum_{\tilde{a} = a_{\text{min}}}^{a-1} \left[ \int_{0}^{1} \phi(\tilde{a},x) dx \right] + \frac{\int_{0}^{1} \phi(a,x) dx}{2},$$

where $a_{\text{min}}$ is the lowest technology in the set $\mathbb{A}$, and $a-1$ is the technology just smaller than $a$. The measure of concentration of talent is given by the coefficient $\beta_1$ in the regression $\bar{x}(a) = \beta_0 + \beta_1 \bar{p}(a) + \varepsilon$. Last, I calculate the agricultural productivity gap for country $t_1$. The model does not explicitly distinguish between agriculture and non-agriculture. However, for each technology $a$ (hence industry according to my interpretation), the model provides average labor productivity, call it $A(a)$,

$$A(a) = \frac{\int_{0}^{\bar{x}(a)} (a f(m(a,x),x) - c(a)) \phi(a,x) dx}{\int_{0}^{\bar{x}(a)} \phi(a,x) dx},$$

and the average ability of individuals working in it, $\bar{x}(a)$. I calculate in the data the average normalized ability $x$ in agriculture and non-agriculture, for each quartile of the income distribution, and then I let the model agricultural productivity be $A(\tilde{a})$, where $\tilde{a}$ is the technology such that $\bar{x}(\tilde{a})$ is equal to the average ability of individuals in agriculture. Similarly for non-agricultural productivity. I then compute the agricultural productivity gap simply as the ratio of the productivity in non-agriculture to productivity in agriculture.\(^{49}\) The agricultural productivity gap is positive in all countries, since, as known (e.g. Caselli and Coleman (2001)), agricultural workers have lower average education.

Identification. I discuss identification of the five parameters. More details are in Appendix D. The five moments are jointly determined by the five parameters. Nonetheless, I provide a discussion of the main links between moments and parameters. $\lambda$ decreases the skill-asymmetry, and thus, according to Proposition 2, a higher $\lambda$ implies a higher concentration of talent. $\eta$ changes the technology gap for teams that use the same technology vintage, since $\frac{\alpha(x,y)}{\alpha(z,y)} = \left( \frac{x(1+\lambda y)}{y(1+\lambda z)} \right)^{\frac{1}{\eta}}$. As a result, still by Proposition 2, a higher $\eta$ decreases the concentration of talent. Additionally, $\eta$ affects the agricultural productivity gap, since a higher $\eta$ implies a lower agricultural gap, for

\(^{49}\)The average ability of individuals in agriculture is 0.36, 0.37, 0.37, 0.38 respectively for the four quartiles form the richest to the poorest. Average ability in non-agriculture is 0.52, 0.56, 0.58, 0.62. Random assignment to individuals across industries would give average ability equal to 0.5, that is both the median and the mean ability individual. In all countries, individuals in agriculture are less able than the average, while individuals in non-agriculture are more able. Details are in Appendix D.1.
given team composition. In countries close to the frontier, \( \eta \) and \( \lambda \) are the main determinants of the concentration of talent, since most teams use same vintage. Therefore, \( \lambda \) and \( \eta \) together are mostly relevant in matching the concentration of talent and agricultural productivity gap close to the frontier. \( \varepsilon \) affects how much the concentration of talent increases moving away from the frontier. If two teams use different technology vintages, their technology gap is given by

\[
\frac{\alpha(x,y)}{\alpha(y,z)} = \left( \frac{x(1+\lambda y)}{y(1+\lambda z)} \right)^{\frac{1}{\eta_c}},
\]

where \( \eta_c = \frac{\eta \varepsilon - 1}{\varepsilon + 1} \) and is increasing in \( \varepsilon \). The higher \( \varepsilon \), the lower the increase in concentration of talent as we decrease the level of development \( t \). In fact, the higher is \( \varepsilon \), the costlier it is to upgrade to the next technology vintage, thus the more similar are the used technologies, hence the lower the concentration of talent. Next, \( \chi_0 \) affects the cost of importing technology vintages, hence the mass of people in each country that decides to not use the local technology. The higher \( \chi_0 \), the higher the decrease in GDP as we decrease the level of development \( t \). Therefore, \( \varepsilon \) and \( \chi_0 \) together match the relationship between the change in concentration of talent and the change in GDP per capita. Last, the dispersion of technology shocks \( \theta \). The lower is \( \theta \), the higher is the dispersion of technology shocks, and thus the lower the concentration of talent, since individuals allocate based not on their comparative advantage, but rather on their tastes for different technologies. The effect of \( \theta \) thus seems not separately identified from \( \lambda \). For this reason, I check the robustness of the results when I fix \( \theta \) to arbitrary values, rather than estimating it. Additionally, I am currently working to target additional moments that allow to distinguish between \( \theta \) and \( \lambda \). The dispersion of average ability within industry, conditional on the concentration of talent, should allow to separately identify \( \theta \). Finally, notice that the main goal of the quantitative exercise is not to find estimates of the vector of parameters, but rather to run counterfactual experiments, in this sense, identification concerns should be less relevant.

**Fit of The Model.** The model fits the data well. In Table 2, I display both the data targeted moments, and the model moments evaluated at the estimated parameter vector. To better visualize the results, I plot the model predicted concentration of talent together with the empirical one in Figure 9a.

The model also predicts the share of individuals in each country that use the frontier technology vintage. I construct, using the CHAT dataset (Comin and Hobijn (2009)), the number of internet user per capita for several countries around the world. I interpret internet as a frontier technology vintage, thus the relative fraction of internet users across the world is comparable to the model predicted relative fractions of individuals that use the frontier vintage.\(^{50}\) I plot the data along with the model predictions in Figure 9b. The model and the data alines very well.

\(^{50}\)Internet users is a very useful technology definition, because it only captures whether a technology is used and not its intensity. In the model notation, an internet user is coded as such as long as he uses \( \bar{t} \) vintage, independently on his choice of \( a \). I’ve computed similar results for other technologies in the CHAT dataset, such as number of computers, and number of cellphones. Results look similar, but are harder to interpret because they confound the number of users with the amount of technology per person.
Note: the left panel plots across countries the empirical measure of concentration of talent, calculated as described in Section 4, and the model simulated one for each income quartile (red crosses) as a function of the GDP per capita. The right panel plots the penetration rate of computers, calculated from the CHAT dataset, and the model simulated fraction of individuals that use the frontier technology (red crosses) as a function of the GDP per capita.

5.3 **Counterfactual Exercise [PRELIMINARY RESULTS]**

I use the model to compute the agricultural productivity gap for each income quartile. Recall that we targeted the productivity gap in the rich countries only. The results are shown in Table 2. The model, disciplined by cross-country differences in the concentration of talent, captures a sizable amount, approximately 40%, of the higher agricultural productivity gap in developing countries. What does this result tell us? Through the lens of the model, the cross-country empirical differences in concentration of talent are consistent with approximately 40% of the larger productivity gap in developing countries. As a result, a naive cross-country comparison that takes as null hypothesis that the agricultural productivity gap should be identical, would overstate the extent to which market frictions are larger in developing countries. At the same time, even after accounting for endogenous technology choice and team formation more than half of cross-country differences remain unexplained, thus still leaving a prominent role to larger market failures in developing countries, or to any other competitive explanations of course.
Table 2: Allocation of Talent and Agriculture to Non-Agriculture Productivity Gaps

<table>
<thead>
<tr>
<th></th>
<th>Quartiles of World Income Distribution</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GDP pc Relative to U.S.</td>
<td>$Q_1$</td>
<td>$Q_2$</td>
<td>$Q_3$</td>
<td>$Q_4$</td>
</tr>
<tr>
<td>Data</td>
<td>Concentration of Talent</td>
<td>38%</td>
<td>47%</td>
<td>53%</td>
<td>62%</td>
</tr>
<tr>
<td></td>
<td>Productivity Gap</td>
<td>2</td>
<td>3.2</td>
<td>3.4</td>
<td>5.6</td>
</tr>
<tr>
<td>Model</td>
<td>Concentration of Talent</td>
<td>36%</td>
<td>48%</td>
<td>51%</td>
<td>53%</td>
</tr>
<tr>
<td></td>
<td>Productivity Gap</td>
<td>1.8</td>
<td>2.4</td>
<td>2.9</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Notes:

Direct Contribution of Differences in Allocation of Talent [DECOMPOSITION OF ROLE OF $c(a)$ AND ROLE OF ENDOGENOUS TALENT ALLOCATION]

Robustness [RESULTS WITH FIXED $\theta$ + ALLOW $\theta$ TO VARY BY COUNTRIES].

6 Discussion and Extensions

[SELF-EMPLOYMENT + EMPIRICS, ENDOGENOUS EDUCATION + EMPIRICS, GROWTH. SEE JMP VERSION]

7 Conclusion

[TO BE COMPLETED]
References


Appendix

In this appendix I include all the proofs of the paper and additional figures.

[NOTE: Some proofs are still included only in Porzio (2016) and need to be rewritten with the new notation. Proofs of new results are included.]

A Additional Figures

Figure A.1: 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 unrecorded 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.2: 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.1. The slopes of the regression lines for United States in 1940 are, from left to right, 0.33; 0.34; 0.41.
Figure A.3: Concentration of Talent Across Sectors

(a) Cross-sectional Differences
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.

(b) Comparison with U.S. in the past

Figure A.4: Concentration of Talent Across Unrecoded Industries

(a) Cross-sectional Differences
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.5: 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.
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.7: 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.8: Dispersion of Average Education Across Firms

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

(b) Dispersion of Education, WBES 2014

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 it 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 ?? . 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).
data are from World Bank Enterprise Survey standardized waves 2002 to 2006. The 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 country is represented by its three digit country code and a circle, whose size depend on the population of the country. The dotted line represents 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 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 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”.

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 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 ???. 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) Average Size  
(b) Comparison Education in WBES and IPUMS  
(c) Dispersion of Size  
(d) Technology Embodied in Capital
B Proofs

B.1 Proofs of Section 2

Proposition 2: Assignment of Talent Across Teams.

In a competitive equilibrium, the for any worker \( i \in [0,1] \) the ability gap between him and his manager, \( m(i) - i \), is bounded above by \( \Upsilon(i) \) and below by \( \Lambda(i) \), where \( \Upsilon(i) \) and \( \Lambda(i) \) depends on \( f \) and \( \alpha \) as follows

1. consider two functions \( \alpha(x', x) \) and \( \alpha'(x', x) \), let \( \Upsilon(i), \Lambda(i) \) be the upper and lower bounds of the equilibrium with \( \alpha \) and similarly \( \Upsilon'(i) \) and \( \Lambda'(i) \) be the bounds for the equilibrium with \( \alpha' \), if \( \forall (x, y, z) \in [0,1] \times [0,1] \times [0,1] \) such that \( x \geq y \geq z \) \( \frac{\alpha(x, y)}{\alpha(y, z)} \geq \frac{\alpha'(x, y)}{\alpha'(y, z)} \) then \( \Upsilon(i) \geq \Upsilon'(i) \) and \( \Lambda(i) \geq \Lambda'(i) \) \( \forall i \in [0,1] \).

2. consider two functions \( f(x', x) \) and \( f'(x', x) \), let \( \Upsilon(i), \Lambda(i) \) be the upper and lower bounds of the equilibrium with \( f \) and similarly \( \Upsilon'(i) \) and \( \Lambda'(i) \) are the bounds for the equilibrium with \( f' \), if \( \forall (x, y, z) \in [0,1] \times [0,1] \times [0,1] \) such that \( x \geq y \geq z \) \( \frac{f_1(x, y)}{f_2(y, z)} \leq \frac{f'_1(x, y)}{f'_2(y, z)} \) then \( \Upsilon(i) \geq \Upsilon'(i) \) and \( \Lambda(i) \geq \Lambda'(i) \) \( \forall i \in [0,1] \).

Proof.

I first consider the upper bound. Let

\[
\hat{x} = \min_{x \leq z} z
\]

s.t. \( \omega(z) < 1 \)

then, it must be that \( \hat{x} \in [x, m(x)] \), since - by definition - \( \omega(m(x)) < 1 \). Also, by the necessary conditions for optimality, at \( \hat{x} \) the following inequality must hold

\[
\pi(\hat{x}) \geq w(\hat{x})
\]

that can be rewritten as

\[
\frac{f_1(\hat{x}, m^{-1}(\hat{x}))}{f_2(m(\hat{x}), \hat{x})} \geq \frac{\alpha(m(\hat{x}), \hat{x})}{\alpha(\hat{x}, m^{-1}(\hat{x}))}.
\]

Since \( m \) is increasing, \( m^{-1}(\hat{x}) \leq x \) and thus

\[
\frac{f_1(\hat{x}, x)}{f_2(m(\hat{x}), \hat{x})} \geq \frac{f_1(\hat{x}, m^{-1}(\hat{x}))}{f_2(m(\hat{x}), \hat{x})} \geq \frac{\alpha(m(\hat{x}), \hat{x})}{\alpha(\hat{x}, m^{-1}(\hat{x}))} \geq \frac{\alpha(m(\hat{x}), \hat{x})}{\alpha(\hat{x}, x)}.
\]

\( \frac{\alpha(m(\hat{x}), \hat{x})}{\alpha(\hat{x}, x)} \) is increasing in \( m(\hat{x}) \), while \( \frac{f_1(\hat{x}, x)}{f_2(m(\hat{x}), \hat{x})} \) is decreasing in \( m(\hat{x}) \), thus, for any \( \hat{x} \), \( \exists \hat{m}(\hat{x}) \) large enough such that \( \frac{\alpha(m(\hat{x}), \hat{x})}{\alpha(\hat{x}, x)} < \frac{f_1(\hat{x}, x)}{f_2(m(\hat{x}), \hat{x})} \) and so \( \pi(\hat{x}) < w(\hat{x}) \). Hence, we know that the manager worker gap at \( \hat{x} \) must satisfy:

\[
m(\hat{x}) - \hat{x} < \hat{m}(\hat{x}) - \hat{x}
\]

(4)

. However, we are interested in \( m(x) - x \). By market clearing and by the fact that \( m \) is increasing we get
that in equilibrium

$$m(\hat{x}) - m(x) \geq \int_x \omega(z) \, dz = \hat{x} - x$$

where the first inequality is an equality if $\omega(z) = 0 \ \forall x \in [m(x), m(\hat{x})]$, and the second equality holds by the definition of $\hat{x}$. Combining 4 and 5 we get that

$$m(x) - x \leq \hat{m}(\hat{x}) - \hat{x}.$$ 

However, $\hat{x}$ is still an unknown value. Therefore the upper bound is given by picking $\hat{x}$ that maximizes $\hat{m}(\hat{x}) - \hat{x}$, subject to the relevant constraints that $\hat{x} - x$ must be smaller than the upper bound itself and that $\hat{m}(\hat{x})$ must be smaller than 1. Last, the upper bound must be between $[0, \frac{1}{2}]$, since by the previous lemma, $m(x) \geq 0$ and the gap between a worker and a manager is bound by $\frac{1}{2}$, due to the fact that $m$ is increasing, $x \in [0, 1]$ and matches are one to one. The upper bound is thus given by

$$\hat{\Upsilon}(x) = \max_{\hat{x} \in [x, 1]} \left\{ \min \left\{ \hat{m}(\hat{x}) - \hat{x}, \frac{1}{2} \right\} \right\}$$

s.t. $\hat{x} - x \leq \hat{m}(\hat{x})$

$$\hat{m}(\hat{x}) \leq 1.$$ 

$$\Upsilon(x) = \max\left\{ 0, \min \left\{ \hat{\Upsilon}(x), \frac{1}{2} \right\} \right\}.$$ 

Notice, that so far we have proved that an upper bound exists, and we have described the implicit function that must satisfy. The comparative static with respect to the properties $\alpha$ holds immediately due to the fact that if $\forall (x, y, z) \in [0, 1] \times [0, 1] \times [0, 1]$ such that $x \geq y \geq z$, $\frac{\alpha(x, y)}{\alpha(y, z)} \geq \frac{\alpha'(x, y)}{\alpha'(y, z)}$ then $\hat{m}'(\hat{x}) \geq m(\hat{x}) \forall \hat{x}$. The same argument applies to $f$.

I next turn to the lower bound. A similar argument applies, even tough with slight modifications. First, let

$$\hat{x} = \max_{z \leq x} z$$

s.t. $\omega(z) < 1$

then, it must be that either there is no $x \leq x$ that satisfies the constraint, that is $\forall z \in [0, x]$, $\omega(z) = 1$; or $\hat{x} \in [x - (m(x) - x), x]$, since by the definition of $\hat{x}$, all individuals between $\hat{x}$ and $x$ are workers, and thus, even if all individuals between $x$ and $m(x)$ are managers, then if $\hat{x} < x - (m(x) - x)$, then there will be more workers than managers, since in equilibrium $m(x) > x \ \forall x$, hence market clearing would be violated. Next, notice that, by the necessary conditions and the definition of $\hat{x}$, $\pi'(\hat{x}) \leq w(\hat{x})$. This last inequality can be rewritten as

$$\frac{f_1(\hat{x}, m^{-1}(\hat{x}))}{f_2(m(\hat{x}), \hat{x})} \leq \frac{\alpha(m(\hat{x}), \hat{x})}{\alpha(\hat{x}, m^{-1}(\hat{x}))}$$

52
and further, exploiting the fact that \( f_{12} > 0 \) we get
\[
\frac{f_1 (\bar{x}, 0)}{f_2 (m (\bar{x}), \bar{x})} \leq \frac{f_1 (\bar{x}, m^{-1} (\bar{x}))}{f_2 (m (\bar{x}), \bar{x})} \leq \frac{\alpha (m (\bar{x}), \bar{x})}{\alpha (\bar{x}, m^{-1} (\bar{x}))} \leq \frac{\alpha (m (\bar{x}), \bar{x})}{\alpha (\bar{x}, 0)}.
\]
Next notice that \( \frac{\alpha (m (\bar{x}), \bar{x})}{\alpha (\bar{x}, 0)} \) is increasing in \( m (\bar{x}) \), while \( \frac{f_1 (\bar{x}, 0)}{f_2 (m (\bar{x}), \bar{x})} \) is decreasing in \( m (\bar{x}) \), as a result, there exist a value of \( \hat{m} (\bar{x}) \) small enough such that \( \frac{f_1 (\bar{x}, 0)}{f_2 (m (\bar{x}), \bar{x})} > \frac{\alpha (m (\bar{x}), \bar{x})}{\alpha (\bar{x}, 0)} \) and thus \( \bar{x}' (\bar{x}) > w (\bar{x}) \), which would violate the necessary conditions. Hence we get that
\[
m (\bar{x}) - \bar{x} \geq \hat{m} (\bar{x}) - \bar{x}.
\tag{6}
\]
By the by the definition of \( \bar{x} \), the fact that \( m \) is increasing, and market clearing, we have that
\[
\int_{\bar{x}}^{x} \omega (z) \, dz = (x - \bar{x}) = \int_{\hat{m} (\bar{x})}^{m (x)} (1 - \omega (z)) \, dz \leq m (x) - m (\bar{x}).
\tag{7}
\]
and combining equations 6 and 7 we have
\[
m (x) - x \geq \hat{m} (\bar{x}) - \bar{x}.
\tag{8}
\]
that gives a lower of \( m (x) - x \) for a given \( \bar{x} \). However, as previously, \( \bar{x} \) is unknown, thus the lower bound is going to the be the minimum value \( \hat{m} (\bar{x}) - \bar{x} \) such that \( \bar{x} \) satisfied the relevant constraints, and such that \( \Lambda (x) \in [0, \frac{1}{2}] \), as previously discussed
\[
\tilde{\Lambda} (x) = \min_{\bar{x} \in [0, x]} \{ \max \{ \hat{m} (\bar{x}) - \bar{x}, 0 \} \}
\tag{9}
\]
s.t.
\[
\bar{x} - x \leq \hat{m} (\bar{x})
\]
\[
\hat{m} (\bar{x}) \leq 1.
\]
\[
\Lambda (x) = \max \left\{ 0, \min \left\{ \tilde{\Lambda} (x), \frac{1}{2} \right\} \right\}.
\]
The discussion for the comparative statics of the lower bound is identical to the one for the upper bound.

**Corollary 1: Conditions for Segregation and Segmentation**

If \( \alpha (x', x) \) satisfies \( \frac{\alpha (x, y)}{\alpha (y, z)} \to \infty \) \( \forall (x, y, z) \in [0, 1] \times [0, 1] \times [0, 1] \), with \( x > y > z \), then talent is segregated by technology. If \( \alpha (x', x) \) and \( f (x', x) \) satisfy \( \frac{\alpha (x, y)}{\alpha (y, z)} < \frac{f_1 (x, y)}{f_2 (y, z)} \) \( \forall (x, y, z) \in [0, 1] \times [0, 1] \times [0, 1] \), with \( x > y > z \), then talent is segmented by occupation.

**B.2 Proofs of Section 3**

**Proposition 3: Technology Gap and Distance to the Technology Frontier**

The optimal technology function \( \alpha (x', x; t, \bar{t}) \) is such that the technology gap is higher further from the technology frontier: \( \forall (x, y, z) \in [0, 1] \times [0, 1] \times [0, 1] \) such that \( x \geq y \geq z \) \( \frac{\alpha (x, y; t', \bar{t})}{\alpha (y, z; t', \bar{t})} \geq \frac{\alpha (x, y; t, \bar{t})}{\alpha (y, z; t, \bar{t})} \) if and only if \( \bar{t} - t' \geq \bar{t} - t \).
Proof.

The first thing to notice is that there are 10 possible cases to consider, depending on the values of \( f(x, y) \) and \( f(y, z) \). I will consider each one of them below. To simplify notation, I call \( x' = f(x, y) \) and \( x = f(y, z) \). Also, I let \( d = \bar{t} - t \) and \( d' = \bar{t}' - \bar{t} \). Notice that, as shown, \( \chi_{1,d'} \geq \chi_{1,d} \) if and only if \( d' \geq d \).

I also define \( \eta_d (x', x) \equiv \alpha(\gamma_{y, x, \bar{t}}, \gamma_{y, x, \bar{t}'}, \gamma_{y, x, \bar{t}'}) \). Let’s now consider each case one by one:

1. If \( x' \leq \chi_0 \): then \( \eta_d (x', x) = \left( \frac{1 + \gamma x'}{1 + \gamma x} \right) \forall d \)

2. If \( x \leq \chi_0 \leq x' \leq \chi_{1,d} \): then \( \eta_d (x', x) = \left( \frac{1 + \gamma x'}{1 + \gamma x} \right) \forall d \)

3. If \( x \leq \chi_0 \leq \chi_{1,d} \leq x' \leq \chi_{1,d'} \): then

\[
\eta_d (x', x) = \left( \frac{1 + \gamma x'}{1 + \gamma x} \right)
\]

and thus

\[
\eta_d (x', x) \geq \eta_d (x', x)
\]

\[
\left( \frac{1 + \gamma x'}{1 + \gamma x} \right) \geq \left( \frac{1 + \gamma x'}{1 + \gamma x} \right)
\]

\[
\gamma x' \geq \gamma x'
\]

\[
\log x' \geq \log \gamma
\]

\[
\log x' \geq \chi_{1,d}
\]

and the last inequality holds by assumption.

4. If \( x \leq \chi_0 \leq \chi_{1,d} \leq \chi_{1,d'} \leq x' \):

\[
\eta_d (x', x) = \left( \frac{1 + \gamma x'}{1 + \gamma x} \right)
\]

\[
\eta_d (x', x) = \left( \frac{1 + \gamma x'}{1 + \gamma x} \right)
\]
and thus

\[ \eta_{d'}(x, x') \geq \eta_d(x, x) \]

\[ \left(1 + \gamma' x' \frac{1}{\eta} \right) \geq \left(1 + \gamma x \frac{1}{\eta} \right) \]

\[ \gamma' x' \frac{1}{\eta} \geq \gamma x \frac{1}{\eta} \]

\[ d' \geq d \]

and again, the last inequality holds by assumption.

5. If \( \chi_0 \leq x \leq x' \leq \chi_{1,d} \): then \( \eta_d(x, x) = \left( \frac{1 + \kappa x \frac{1}{\eta}}{1 + \kappa x' \frac{1}{\eta}} \right) \forall d \)

6. If \( \chi_0 \leq x \leq \chi_{1,d} \leq x' \leq \chi_{1,d'} \):

\[ \eta_{d'}(x, x) = \left( \frac{1 + \kappa' x' \frac{1}{\eta}}{1 + \kappa' x' \frac{1}{\eta}} \right) \]

\[ \eta_d(x, x) = \left( \frac{1 + \gamma x \frac{1}{\eta}}{1 + \kappa x \frac{1}{\eta}} \right) \]

and thus

\[ \eta_{d'}(x, x) \geq \eta_d(x, x) \]

\[ \left(1 + \kappa' x' \frac{1}{\eta} \right) \geq \left(1 + \gamma x \frac{1}{\eta} \right) \]

\[ \kappa' x' \frac{1}{\eta} \geq \gamma x \frac{1}{\eta} \]

\[ \log x' \geq \log \kappa_0 + \frac{\eta \kappa_0}{\eta - \gamma} d \log \gamma \]

\[ \log x' \geq \chi_{1,d} \]

which holds by assumption.

7. If \( \chi_0 \leq x \leq \chi_{1,d} \leq \chi_{1,d'} \leq x' \):

\[ \eta_{d'}(x, x) = \left( \frac{1 + \gamma' x' \frac{1}{\eta}}{1 + \kappa x \frac{1}{\eta}} \right) \]

\[ \eta_d(x, x) = \left( \frac{1 + \gamma x \frac{1}{\eta}}{1 + \kappa x \frac{1}{\eta}} \right) \]
and thus

\[ \eta_{d'}(x', x) \geq \eta_d(x', x) \]
\[ \left( \frac{1 + \gamma'^d x'^\frac{1}{\eta}}{1 + \kappa_2 x'^\frac{1}{\eta}} \right) \geq \left( \frac{1 + \gamma^d x^\frac{1}{\eta}}{1 + \kappa_2 x^\frac{1}{\eta}} \right) \]
\[ \gamma'^d x'^\frac{1}{\eta} \geq \gamma^d x^\frac{1}{\eta} \]
\[ d' \geq d \]

and again, the last inequality holds by assumption.

8. If \( \chi_0 \leq \chi_{1,d} \leq x \leq x' \leq \chi_{1,d'} \):

\[ \eta_{d'}(x', x) = \left( \frac{1 + \kappa_2 x'^\frac{1}{\eta}}{1 + \kappa_2 x^\frac{1}{\eta}} \right) \]
\[ \eta_d(x', x) = \left( \frac{1 + \gamma^d x^\frac{1}{\eta}}{1 + \gamma^d x^\frac{1}{\eta}} \right) \]

This requires a bit more steps.

\[ \eta_{d'}(x', x) \geq \eta_d(x', x) \]
\[ \left( \frac{1 + \kappa_2 x'^\frac{1}{\eta}}{1 + \kappa_2 x^\frac{1}{\eta}} \right) \geq \left( \frac{1 + \gamma^d x^\frac{1}{\eta}}{1 + \gamma^d x^\frac{1}{\eta}} \right) \]
\[ \left( 1 + \kappa_2 x'^\frac{1}{\eta} \right) \left( 1 + \gamma^d x^\frac{1}{\eta} \right) \geq \left( 1 + \gamma^d x^\frac{1}{\eta} \right) \left( 1 + \kappa_2 x^\frac{1}{\eta} \right) \]
\[ 1 + \gamma^d x^\frac{1}{\eta} + \kappa_2 x'^\frac{1}{\eta} + \kappa_2 x'^\frac{1}{\eta} \gamma^d x^\frac{1}{\eta} \geq 1 + \kappa_2 x^\frac{1}{\eta} + \gamma^d x^\frac{1}{\eta} + \gamma^d x^\frac{1}{\eta} \kappa_2 x^\frac{1}{\eta} \]
\[ \left( \kappa_2 x'^\frac{1}{\eta} - \gamma^d x^\frac{1}{\eta} \right) - \left( \kappa_2 x^\frac{1}{\eta} - \gamma^d x^\frac{1}{\eta} \right) \geq 0 \]

and then, using the fact that \( x' \geq x \) and \( \eta \geq \eta_e \) we get that \( \left( x'^\frac{1}{\eta} - x^\frac{1}{\eta} \right) \geq 0 \). Then using the fact that \( x' \geq x \geq \chi_{1,d} \) and \( \eta \geq \eta_e \) we get that \( \left( \kappa_2 x'^\frac{1}{\eta} - \gamma^d x^\frac{1}{\eta} \right) \geq \left( \kappa_2 x^\frac{1}{\eta} - \gamma^d x^\frac{1}{\eta} \right) \), since

\[ \frac{\partial}{\partial x} \left( \kappa_2 x'^\frac{1}{\eta} - \gamma^d x^\frac{1}{\eta} \right) = \left( \kappa_2 \frac{1}{\eta_e} x'^{\frac{1}{\eta_e} - 1} - \frac{1}{\eta} \gamma^d x^\frac{1}{\eta} \right) \geq 0 \]

9. If \( \chi_0 \leq \chi_{1,d} \leq x \leq \chi_{1,d'} \leq x' \):

\[ \eta_{d'}(x', x) = \left( \frac{1 + \gamma'^d x'^\frac{1}{\eta}}{1 + \kappa_2 x'^\frac{1}{\eta}} \right) \]
\[ \eta_d(x', x) = \left( \frac{1 + \gamma^d x^\frac{1}{\eta}}{1 + \gamma^d x^\frac{1}{\eta}} \right) \]
and thus

\[ \eta_{d'} (x', x) \geq \eta_d (x', x) \]
\[ \left( \frac{1 + \gamma^{d'} x^\frac{1}{\eta}}{1 + \kappa_2 x^\frac{1}{\eta}} \right) \geq \left( \frac{1 + \gamma^d x^\frac{1}{\eta}}{1 + \gamma^d x^\frac{1}{\eta}} \right) \]

and notice that, since \( x \leq \chi_{1,d'} \) we have that

\[ 1 + \kappa_2 x^\frac{1}{\eta} \leq 1 + \gamma^{d'} x^\frac{1}{\eta} \]

and thus

\[ \left( \frac{1 + \gamma^{d'} x^\frac{1}{\eta}}{1 + \kappa_2 x^\frac{1}{\eta}} \right) \geq \left( \frac{1 + \gamma^{d'} x^\frac{1}{\eta}}{1 + \gamma^{d'} x^\frac{1}{\eta}} \right) \]

and then (see below)

\[ \left( \frac{1 + \gamma^{d'} x^\frac{1}{\eta}}{1 + \gamma^{d'} x^\frac{1}{\eta}} \right) \geq \left( \frac{1 + \gamma^{d'} x^\frac{1}{\eta}}{1 + \gamma^{d'} x^\frac{1}{\eta}} \right) \]

since \( d' \geq d \) and \( x' \geq x \).

10. If \( \chi_{1,d'} \leq x \): then

\[ \eta_{d'} (x', x) = \left( \frac{1 + \gamma^{d'} x^\frac{1}{\eta}}{1 + \gamma^d x^\frac{1}{\eta}} \right) \]
\[ \eta_d (x', x) = \left( \frac{1 + \gamma^d x^\frac{1}{\eta}}{1 + \gamma^d x^\frac{1}{\eta}} \right) \]

and thus

\[ \eta_{d'} (x', x) \geq \eta_d (x', x) \]
\[ \left( \frac{1 + \gamma^{d'} x^\frac{1}{\eta}}{1 + \gamma^d x^\frac{1}{\eta}} \right) \geq \left( \frac{1 + \gamma^d x^\frac{1}{\eta}}{1 + \gamma^d x^\frac{1}{\eta}} \right) \]

that holds as long as \( d' \geq d \), since

\[ \frac{\partial}{\partial \kappa} \left( \frac{1 + \kappa x^\frac{1}{\eta}}{1 + \kappa x^\frac{1}{\eta}} \right) = \frac{x^\frac{1}{\eta} - x^\frac{1}{\eta}}{(1 + \kappa x^\frac{1}{\eta})^2} \geq 0 \]

for any \( x' > x \). This concludes the proof. □
B.3 Characterization and Proofs for the “Tractable Case”

I here completely characterize the optimal allocation for the tractable case described in section and provide the Proofs for Lemma 10 and Proposition 4.

The allocation can take one of five shapes, that are described in words in the propositions and are graphically represented in the Figures 4a, 4b, 4c, 4d, and 4e in the main text. Within the proof of this propositions I also provide the proofs of Lemmas 6, 7, and 8.

As a preliminary step, I define $\bar{a}$ to be the technology used with the frontier vintage, $\bar{a} = \gamma^t + \bar{t}$ and $\gamma$ be the technology used with the local vintage, $a = \gamma^t$. I call $\param$ the ratio between the two technologies: $\param = \frac{1 + \bar{t}}{\gamma}$, and $A(x)$ to be the average used technology of individuals of ability $x$, $A(x) = \omega(x) \alpha(m(x), x) + (1 - \omega(x)) \alpha(x, m^{-1}(x))$.

**Proposition A.1 (Characterization of Tractable Model).** The optimal allocation $\{\omega, \alpha, m\}$ takes one of five shapes depending on the value of $\bar{t} - t$. In each shape, the optimal technology $\alpha$ is given by the cutoff policy shown in Lemma 9. I include for completeness description of $A(x)$ as defined above. There exists four finite constants bigger than one, $\param_1 < \param_2 < \param_3 < \param_4$, such that

(i) If $\param \leq \param_1$. Skills are segmented by occupation: low skilled individuals are workers and high skilled ones are managers. Specifically,

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

$$A(x) = \begin{cases} \bar{a} & \text{if } x' \in \left[\frac{1}{2} - (1 - \param_0), \frac{1}{2}\right] \cup \left[\param_0, 1\right] \\ a & \text{if } x' \in \left[0, \frac{1}{2} - (1 - \param_0)\right] \cup \left[\frac{1}{2}, \param_0\right] \end{cases}$$

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

(ii) If $\param \in [\param_1, \param_2]$. 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 local one, $a$. Specifically there exists constants $\hat{x}_1, \hat{x}_2$ such that
\[\omega(x) = \begin{cases} 
1 & \text{if } x < \hat{x}_1 \\
\frac{\eta}{\eta + 1} & \text{if } x \in [\hat{x}_1, \hat{x}_2], \\
0 & \text{if } x > \hat{x}_2 
\end{cases}\]

\[A(x) = \begin{cases} 
\tilde{a} & \text{if } x \in [\hat{x}_1 - \left(1 - \chi_0 - \frac{\eta}{\eta + 1} (\hat{x}_2 - \hat{x}_1)\right), \hat{x}_1] \cup [\chi_0, 1] \\
\tilde{a} \frac{\eta}{\eta + 1} + \left(1 - \frac{\eta}{\eta + 1}\right) a & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\
a & \text{if } x \in [0, \hat{x}_1 - \left(1 - \chi_0 - \frac{\eta}{\eta + 1} (\hat{x}_2 - \hat{x}_1)\right)] \cup [\hat{x}_2, \chi_0] 
\end{cases}\]

\[m(x) = \begin{cases} 
\hat{x}_1 + \frac{1}{\eta + 1} x & \text{if } x \in [0, \hat{x}_1 - \left(1 - \chi_0 - \frac{\eta}{\eta + 1} (\hat{x}_2 - \hat{x}_1)\right)] \cup [\hat{x}_2, \chi_0] \\
\hat{x}_2 + x - \left(1 - \frac{\eta}{\eta + 1}\right) (\hat{x}_2 - \hat{x}_1) & \text{if } x \in \left(1 - \frac{\eta}{\eta + 1}\right) (\hat{x}_2 - \hat{x}_1), \hat{x}_1 \\
\hat{x}_2 + \hat{x}_1 - \left(1 - \frac{\eta}{\eta + 1}\right) (\hat{x}_2 - \hat{x}_1) + \lambda(x - \hat{x}_1) & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\
1 & \text{if } x \geq \hat{x}_2 
\end{cases}\]

(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 \(\tilde{a}\) 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 \(51\)

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

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

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

(iv) If \(\eta \in [\eta_3, \eta_4]\). Skills are almost perfectly segregated by technology: most of the individuals which use the traditional technology \(a\) are less skilled than those that use the frontier one \(\tilde{a}\), there is however a

\[51\text{In order to ease notation I am using again } \hat{x}_1, \text{ and } \hat{x}_2. \text{ However notice that these constant are not necessarily identical to the ones defined for the case } \frac{3}{2} \in [\eta_1, \eta_2]. \text{ In fact in general have a different value. This is true for all constants defined below.}\]
group of managers that use $\bar{a}$ which are more skilled than the lowest skilled workers among those that use $\bar{a}$. Specifically there exists constants $\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4$ such that

\[
\omega(x) = \begin{cases} 
1 & x \in [0, \hat{x}_1] \cup [\hat{x}_3, \hat{x}_4] \\
\frac{\eta}{\eta+1} & x \in (\hat{x}_2, \hat{x}_3) \\
0 & x \in (\hat{x}_1, \hat{x}_2) \cup (\hat{x}_4, 1) 
\end{cases}
\]

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

\[
m(x) = \begin{cases} 
\hat{x}_1 + x & \text{if } x \in [0, (\hat{x}_2 - \hat{x}_1)] \\
\hat{x}_2 + (\eta + 1)(x - (\hat{x}_2 - \hat{x}_1)) & \text{if } x \in [(\hat{x}_2 - \hat{x}_1), \hat{x}_1] \\
\hat{x}_3 & \text{if } x \in [\hat{x}_1, \hat{x}_2] \\
\hat{x}_4 + \frac{\eta}{\eta+1}(x - \hat{x}_2) & \text{if } x \in [\hat{x}_2, \hat{x}_3] \\
\hat{x}_4 + \frac{\eta}{\eta+1}(\hat{x}_3 - \hat{x}_2) + x - \hat{x}_3 & \text{if } x \in [\hat{x}_3, \hat{x}_4] \\
1 & \text{if } x \geq \hat{x}_4 
\end{cases}
\]

(v) If $\eta \geq \eta_4$. Skills are segregated by technology: low skilled individuals use the local technology $a$, and high skilled ones use the frontier technology $\bar{a}$. Among the individuals that use $a$, 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

\[
\omega(x) = \begin{cases} 
1 & x \in [0, \hat{x}_1] \cup [\hat{x}_2, \hat{x}_3] \\
0 & x \in (\hat{x}_1, \hat{x}_2) \cup (\hat{x}_3, 1) 
\end{cases}
\]

\[
A(x) = \begin{cases} 
\bar{a} & \text{if } x \in [\hat{x}_2, 1] \\
\bar{a} & \text{if } x \in [0, \hat{x}_2] 
\end{cases}
\]

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

**Proof.** [NOTE: OUTDATED NOTATION] 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 \( x' (1 + \lambda x) \), a team picks either \( \bar{a} \) or \( a \). Hence a team picks \( \bar{a} \) if and only if

\[
\bar{a} x' (1 + \lambda x) - \chi_0 (\bar{a} - a) (1 + \lambda x) > a x' (1 + \lambda x) \\
x' > \chi_0
\]

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

\[
v(x', x; \bar{a}, a) = \begin{cases} 
(\bar{a} (x' - \chi_0) + a \chi_0) (1 + \lambda x) & \text{if } x' \geq \chi_0 \\
\bar{a} x' (1 + \lambda x) & \text{if } x' < \chi_0
\end{cases}
\]

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

\[
Y = \int v(m(x), x; \bar{a}, a) \omega(x) \, dx = \int v(m(x), x; \eta, 1) \omega(x) \, dx
\]

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 \( v(x', x; \eta, 1) \) on \( \eta \).

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

\[
v_1(x, w(x)) = \begin{cases} 
\eta (1 + \gamma w(x)) & \text{if } x \geq \zeta \\
(1 + \gamma w(x)) & \text{if } x < \zeta
\end{cases}
\]

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

also notice that

\[
v_{12}(x, w(x)) = \begin{cases} 
\gamma \eta & \text{if } x \geq \zeta \\
\gamma & \text{if } x < \zeta
\end{cases}
\]

\[
v_{12}(m(x), x) = \begin{cases} 
\gamma \eta & \text{if } m(x) \geq \zeta \\
\gamma & \text{if } m(x) < \zeta
\end{cases}
\]
as a result we get that, since \( m(x), w(x), x \in [0,1] \)

\[
v_1(x, w(x)) > v_2(m(x), x) \quad \text{if} \quad x \geq \zeta
\]
\[
v_1(x, w(x)) > v_2(m(x), x) \quad \text{if} \quad m(x) < \zeta
\]
\[
v_1(x, w(x)) < v_2(m(x), x) \quad \text{if} \quad x < \zeta \text{and} \quad m(x) > \kappa(\eta, w(x))
\]

where

\[
\kappa(\eta, w(x)) = \zeta + \frac{1 + \gamma w(x) - \zeta \gamma}{\gamma \eta}
\]

so that \( \lim_{\eta \to \infty} \kappa(\eta, w(x)) = \zeta \) 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. \( \forall x. \text{s.t.} \omega^*(x) > 0 \) and \( \mu^*(x) > 0 \) then \( v_1(x, w^*(x)) = v_2(m^*(x), x) \);
2. \( \forall x. \text{s.t.} \mu^*(x - \epsilon) = 0 \) and \( \mu^*(x) > 0 \) then \( v_1(x, w^*(x)) \geq v_2(m^*(x), x) \);
3. \( \forall x. \text{s.t.} \mu^*(x - \epsilon) > 0 \) and \( \mu^*(x) = 0 \) then \( v_1(x, w^*(x)) \leq v_2(m^*(x), x) \).

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 \( \omega^*(x_1) = 1 \), \( \omega^*(x_2) = 1 \), 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 \( \zeta \). 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 \( \eta \) increases.

**Case 1:** \( \eta \in [1, \eta_1] \). Consider first \( \eta_1 = 1 \). 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 \( \eta_1 \) 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: $\eta \in (\eta_1, \eta_2]$, where $\eta_1$ is as previously defined and $\eta_2$ 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 $\eta$ the optimal allocation is given by

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

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

where $\hat{x}_1$, $\hat{x}_2$, and $\lambda$ 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

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

$\lambda$ 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 $\eta \leq \eta_2$.

Case 3: $\eta \in (\eta_2, \eta_3]$, where $\eta_3$ 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 $\zeta$

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

---

$\overset{52}{\text{I can also show how to solve explicitly for } \hat{x}_0 \text{ and } \eta_2, \text{ so that they are not expressed in terms of endogenous objects. We do this, by using market clearing, that gives}}$

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

and the definition of $\kappa$ itself.

$$
\zeta + \frac{1 + \gamma \hat{x}_0 - \zeta \gamma}{\gamma \eta_2} = 1.
$$
with \( \hat{x}_0 \) defined above. In fact the optimal allocation is given by

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

where \( \lambda \) as before is set to guarantee

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

hence \( \lambda = \frac{n}{\eta+1} \) 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: \( \eta \in (\eta_3, \eta_4] \), where \( \eta_4 \) solves \( \kappa (\eta_4, \hat{x}_1) = \hat{x}_3 \). Now consider a marginal increase in \( \eta \), 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

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

where as usual \( \lambda = \frac{n}{\eta+1} \). 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 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

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

This holds 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

\[
\hat{x}_2 - \hat{x}_1 = \hat{x}_1 \\
1 - \hat{x}_3 = \hat{x}_3 - \hat{x}_2
\]

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

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

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 4** [TO BE COMPLETED, very simple given the complete characterization].

C **Computing Algorithm**

[DETAILS ON COMPUTATION ALGORITHM];

D **Identification**

[ADD GRAPHS OF PARTIAL CHANGE OF PARAMETERS + CALIBRATION FOR CASE WITH \( \theta \) FIXED];

D.1 **Average Ability in Agriculture and Non-Agriculture**

[DESCRIPTION OF PROCEDURE TO CALCULATE AVERAGE ABILITY BY AGRICULTURE AND NON AGRICULTURE + FIGURE].
E Additional Details on the Data

E.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.
E.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.8a, 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.

F 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.10 and A.12. 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.11 and shows that the patterns are similar.

Last, in A.13 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.
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.

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.12: Concentration of Talent, Comparison between KLOSA and KLIPS dataset in 2005

Notes: this figure replicates figures A.1 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.13: Concentration of Talent Across Industries, South Korea Growth Path

Notes: I apply, across industries, the same procedure as described in Figure A.1. 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.