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

Tommaso Porzio†

UCSD

Click here for most current version

December 7, 2016

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 equilibrium 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, 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 cross-sectional productivity dispersion, 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 many insightful comments from seminar participants at Yale and several other universities.

†9500 Gilman Drive, La Jolla, CA 92093, email: tporzio@ucsd.edu.

1 Introduction

Low income 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 low income countries: most individuals see their labor produce very little output, but some others enjoy a fairly high productivity.[1] Often, this excess productivity dispersion is interpreted as evidence of market frictions, which may prevent mobility to more productive firms and sectors, or adoption of 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 propose entails two insights.

First, low income countries have the unique opportunity to 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. Heterogenous adoption thus leads naturally to dispersion of used technology and, as a result, 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 a person in the streets of Kolkata using a 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, when cellphones and the internet were still far from being invented.

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 using advanced technology, and so 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, the rest of the economy is left only with low skilled ones, that have low inherent productivity and few incentives to improve their technology. The concentration of similarly talented individuals within the same teams is thus both a consequence and a catalyst of productivity and technology dispersion.

The interaction of these two insights shapes the economic structure of poor countries. Some economic niches attract most of the high skilled individuals and adopt the advanced technology.

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

[2] Notable exceptions are Lagakos (2013), Lagakos and Waugh (2013), Keller et al. (2016), and Young (2013).

[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, and of Suri (2011) that provides micro evidence of heterogeneous technology adoption.

The rest of the economy is left without talent, and thus finds convenient to rely on backward, cheaper, technology. The possibility to adopt technology from the frontier endogenously generate cross-sectional dispersion and dual economies in poor countries.

The described 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 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. The economy is inhabited by a continuum of individuals of heterogenous ability. Production requires three inputs: a manager, a worker, and a technology. The production function satisfies two 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. The competitive equilibrium of this setting decentralizes the Pareto efficient allocation. I characterize it. The endogenous sets of managers and workers are the main objects of interest. Given these two sets, complementarity dictates positive assortative 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.[4] If every team would pick identical technology, the higher skill-sensitivity of output dictates that the most skilled individuals 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. The main theoretical result of the paper shows that the more important is the choice of an appropriate technology - i.e. the larger the elasticity to ability of the optimal technology choice relatively to the production's asymmetry in skill-sensitivity - the smaller is the ability gap between two individuals working together. The allocation of talent and technology (and thus productivity) are tied together: when technology

[4]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.

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 less advanced technology.

In Section 3, I explicitly model cross-country differences in the cost of technology. I distinguish between technology and technology vintage. I call technology the input of the production function, that is simply a productivity term that multiplies labor input.[5] I argue that this notion of technology depends on two elements: (i) the choice of which vintage of technology to use, for example to rely on animal or electric power; and (ii) the amount of capital of the used vintage, for example how many bullocks are purchased. More modern vintages have a lower variable cost of technology, but a larger fixed one to be operated. Countries differ in their local technology vintage. In each country, every production team can use the local vintage without paying any fixed cost. Teams need instead to pay a fixed cost to import more modern vintages from other countries. This cost increases in the distance between the chosen vintage and the local one. I derive from this environment a country specific cost of technology. I 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. This result is generated by the heteregeneity of returns to technology adoption: even in countries far from the frontier, the most able teams find optimal to pay the fixed cost to import modern vintages, while the less able ones prefer to rely on local ones.

Combining the insights of Sections 2 and 3, the theory provides sharp predictions on cross-country differences in their economic structure. First of all, the model predicts that the efficient competitive equilibrium generates larger technology and labor productivity dispersion in poor countries. Large productivity dispersion is often associated with misallocation. In this framework, instead, it is generated through differences in endowments: poor countries are endowed with less advanced local technology vintage and this leads them to an equilibrium with larger productivity dispersion. Second, the theory predicts cross-country differences in the way in which people form production teams. Specifically, talent should be more concentrated in countries far from the frontier: an empirical prediction, to the best of my knowledge, unique to this model.

In Section 4, I confront the model's predictions on the allocation of talent with the data.

In the main empirical exercise, I use large sample labor force surveys for 63 countries 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. 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

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

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 - but, critically, it was closer to the technology frontier. Middle income countries today have significantly higher concentration of talent than the U.S. did in the past. This result alleviates the concern that cross-sectional differences are merely capturing differences in levels of development. Last, I compare the growth paths of South Korea and the United States in the past seven decades. In the United States, the concentration of talent has remained largely unchanged, consistent with the fact that they have been growing steadily as world leaders (i.e. on the frontier). In contrast, South Korea has seen a rapid decline in the concentration of talent as it has approached the technology frontier.

I then provide further supporting evidence from occupation data: managers are on average less skilled in poor countries, while workers are more skilled, as predicted.[6] The definition of managers and workers in the data is not perfectly comparable across countries. For this reason, I also focus on one specific occupation, retail cashier. Retail cashier is a simple occupation (it would be a worker in the model), but one that uses an advanced technology, the cash register. The model predicts that in poor countries retail cashiers should be relatively more skilled. This prediction is confirmed in the data.

Last, I use the World Bank Enteprise Surveys to indirectly validate the main model assumptions - the tasks and skill-technology complementarities - and to confirm that the allocation of talent is different in poor countries even at the firm level.

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.[7] 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.

[6] As I show analytically, high concentration of talent implies low ability gap between managers and workers.

[7] Labor 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.

Related Literature. The theoretical contribution of the paper is to combine together two strands of the literature and show that they interact non-trivially. I consider an environment where individuals form production teams - following the seminal work of Lucas (1978) and Kremer (1993), and more recently of Garicano and Rossi-Hansberg (2006) - and at the same time choose an appropriate production technology - along the lines of the work of Atkinson and Stiglitz (1969), Basu and Weil (1998), and Acemoglu and Zilibotti (2001). I show that allowing teams to choose a production input, technology, that is complementary with the ability of team members changes the equilibrium assignment. To the best of my knowledge, this is a new insight for the literature on team formation. At the same time, taking into consideration endogenous team formation is relevant to understand the distribution of technology within a country. In particular, equilibrium team formation links the right and the left tails of the technology distribution: the possibility to adopt advanced technology generates a right tail of high productivity teams that gather all the talent, thus crowding out high skilled individuals from the rest of the economy, and creating a left tail of low productivity teams. In other words, modern technology in some economic niches within a country come at the cost of backward technology elsewhere. To the best of my knowledge, this is a new insight for the literature on appropriate technology and technology adoption. In studying the distribution of technology and productivity within a country as the result of behavior of maximizing agents, my work is similar to Perla and Tonetti (2014). In allowing countries to import more advanced technologies from abroad, my work is similar to Buera and Oberfield (2014). Both of these papers do not consider the endogenous allocation of talent.

In studying a setting where the pattern of matching is an endogenous outcome, my work mostly resembles Kremer and Maskin (1996).[8] While Kremer and Maskin (1996) studies how changes in skill distributions can impact the pattern of matching, I keep the skill distribution constant, and show instead that endogenous choice of technology shapes the pattern of matching. I also provide a new solution method and develop a quantitative version of the model.

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

The idea that distance to the frontier may impact organization of production is present in Acemoglu et al. (2006). They study selection into entrepreneurship with credit constraints. I thus see my work as complementary to theirs. Roys and Seshadri (2013) also studies differences across countries in the way in which production is organized. It uses a quantitative version of Garicano and Rossi-Hansberg (2006) in which human capital is endogenously accumulated,

[8]Notice the above cited literature on team formation relies instead on assumptions on production such that the pattern of matching is known-ex ante.

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.

Last, two other papers, Acemoglu (1999) and Caselli (1999), have already argued that the technological environment and the allocation of workers to jobs are connected. 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.

2 A Model of Technology Choice and Allocation of Talent

This section develops an assignment 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.[9] 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

[9] 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.

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

$$g\left(a, x', x\right) \quad = \quad af\left(x', x\right) - c(a).$$

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 managerial occupation.[10] 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 (1996)).[11] Further, the ability of the manager is more relevant to generate high returns from

[10] 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.

[11] 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.

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)).

Notice that the functional form assumption of g implicitly assumes a specific strength of complentarity between skills and technology. This restriction is useful for tractability, but is not necessary for the main results. In the Appendix section D I consider a more general g function.

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.[12]

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

$$\alpha\left(x', x\right) = \arg\max_{a \in \mathbb{R}} a f\left(x', x\right) - c(a).$$

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

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

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

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

[12] 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.

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 mass of workers less skilled than him must equal the mass of managers matched with them according to m

$$\int_0^x \omega(z)\, dz = \int_0^x (1 - \omega(m(z)))\, dz.$$

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

$$\begin{aligned} \pi'(x) &= \alpha\left(x, m^{-1}(x)\right) \quad f_1\left(x, m^{-1}(x)\right) \\ w'(x) &= \underbrace{\alpha(m(x), x)}_{\text{Appropriate Technology}} \quad \underbrace{f_2(m(x), x)}_{\text{Skill Sensitivity}} \end{aligned} \tag{1}$$

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. □

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:

$$\underbrace{\frac{f_1\left(x, m^{-1}(x)\right)}{f_2\left(m(x), x\right)}}_{\text{Skill-Sensitivity Gap}} \geq \underbrace{\frac{\alpha^*\left(m(x), x\right)}{\alpha^*\left(x, m^{-1}(x)\right)}}_{\text{Technology Gap}}.$$

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[13] 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.[14] 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 α.

Necessary Conditions. The occupational choice function $\omega(x)$ maximizes individual income:

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

$\omega(x)$ 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

[13] The one notable exception that I am aware of is Kremer and Maskin (1996).

[14] 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.

(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. $\forall x$ such that $\omega(x) \in (0,1)$: $\pi'(x) = w'(x)$
2. $\forall x$ such that $\omega(x-\epsilon) = 0$ and $\omega(x) > 0$: $\pi'(x) \leq w'(x)$
3. $\forall x$ such that $\omega(x-\epsilon) = 1$ 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 $\frac{\alpha(m(\hat{x}),\hat{x})}{\alpha(\hat{x},m^{-1}(\hat{x}))} \geq \frac{f_1(\hat{x},m^{-1}(\hat{x}))}{f_2(m(\hat{x}),\hat{x})}$. 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 $\frac{\alpha(m(\hat{x}),\hat{x})}{\alpha(\hat{x},m^{-1}(\hat{x}))} \leq 1$, since α is increasing in both his arguments. The Spence-Mirlees assumption (3) guarantees that $\frac{f_1(\hat{x},m^{-1}(\hat{x}))}{f_2(m(\hat{x}),\hat{x})} > 1$. 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 α 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 $\iota \in [0,1]$ *the ability gap between him and his manager,* $m(\iota) - \iota$*, is bounded above by* $\Upsilon(\iota)$ *and below by* $\Lambda(\iota)$*, where* $\Upsilon(\iota)$ *and* $\Lambda(\iota)$ *depend on* f *and* α *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* $x \geq y \geq z$ $\frac{f_1(x,y)}{f_2(y,z)} \geq \frac{\tilde{f}_1(x,y)}{\tilde{f}_2(y,z)}$ *then* $\Upsilon(\iota) \geq \tilde{\Upsilon}(\iota)$ *and* $\Lambda(\iota) \geq \tilde{\Lambda}(\iota)$ $\forall \iota \in [0,1]$;
2. *consider two functions* α *and* $\tilde{\alpha}$*, if* $\forall(x,y,z) \in [0,1]\times[0,1]\times[0,1]$ *such that* $x \geq y \geq z$ $\frac{\alpha(x,y)}{\alpha(y,z)} \geq \frac{\tilde{\alpha}(x,y)}{\tilde{\alpha}(y,z)}$ *then* $\Upsilon(\iota) \leq \tilde{\Upsilon}(\iota)$ *and* $\Lambda(\iota) \leq \tilde{\Lambda}(\iota)$ $\forall \iota \in [0,1]$.

Proof. See appendix. □

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 α. f and α 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 $1 - \int [m(x) - x]\,\omega(x)\,dx$ the concentration of talent.[15]

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 $\alpha\left(x', m^{-1}(x')\right) \geq \alpha(m(x), x)$ 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 $m(x) \to x$.[16] This second case is illustrated in Figure 1b.

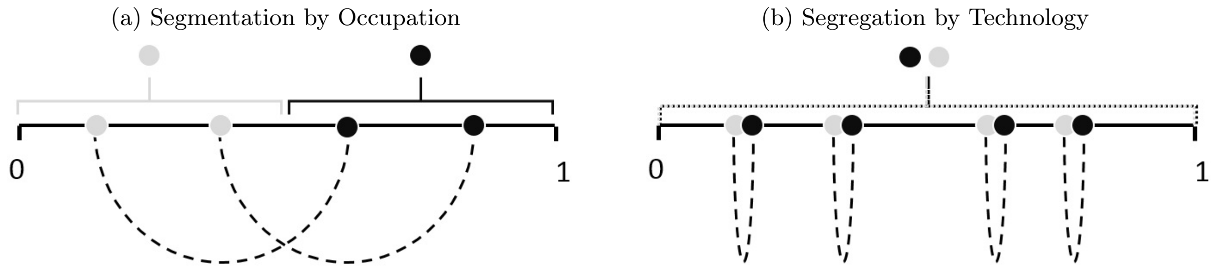
[15] Using the market clearing, we can notice that the concentration of talent is simply equal to 1 minus the difference between the ability of managers and workers, since $\int m(x)\,\omega(x) = \int x\,[1 - \omega(x)]\,dx$.

[16] To see why this is the case, consider that if $\exists \hat{x} \in (x, m(x))$, 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 $\alpha(m(x), x) > \alpha\left(\hat{x}, m^{-1}(\hat{x})\right)$, again violating the definition.

Corollary 1: Conditions for Segregation and Segmentation
If α *and* f *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 and* $m(x) = \frac{1}{2}$*. If* α *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 and* $m(x) \to x$*.*

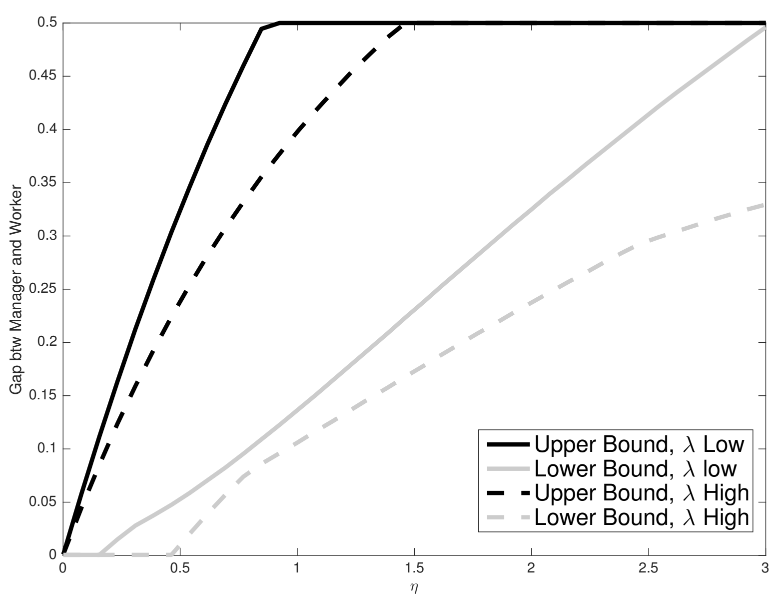
Proof. See appendix. □

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



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

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



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 $c(a) = \frac{a^{1+\eta}}{1+\eta}$, with $\eta \geq 0$, which implies $\alpha(x', x) = (x'(1 + \lambda x))^{\frac{1}{\eta}}$. This system of functional forms has two parameters: lower λ generates stronger skill asymmetry, and lower η 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 η and for two different values of λ. At least in this case, the bounds are informative. We can also note that they converge to the values provided in Corollary 1.

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 α, the technology gap, rather than in terms of the primitive $c(a)$.[17] 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.

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 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

[17] If the reader prefers to see the results in terms of primitives, he can let $c(a) = \frac{a^{1+\eta}}{1+\eta}$, and then Proposition 2 and Corollary 1 can be rewritten in terms of η: (i) both bounds are increasing in η; (ii) if $\eta \to 0$, $m(x) \to x$; (iii) if $\eta > \eta_1$, with η_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 η.

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

$$c_L\left(a;t\right) \quad = \quad \gamma^{-\eta t}\kappa_1^{-\eta}\frac{\left(a-a_t\right)^{1+\eta}}{1+\eta}$$

where γ captures the efficiency difference between each technology vintage, η 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 $c_L\left(a;t\right)$ as the variable cost component linked to the chosen capital intensity. κ_1 is a constant term that scales the cost of technology. I pick $\kappa_1\equiv\frac{\eta\varepsilon-1}{\eta\varepsilon}$ 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

$$c_I\left(a,\tilde{t};t\right) \quad = \quad \gamma^{-\eta\tilde{t}}\frac{\left(a-a_t\right)^{1+\eta}}{1+\eta}+\kappa_2^{-(\eta\varepsilon-1)}\frac{\gamma^t\gamma^{\varepsilon\eta\left(\tilde{t}-t\right)}}{\varepsilon\left(1+\eta\right)}$$

where $\kappa_2^{-(\eta\varepsilon-1)}\frac{\gamma^t\gamma^{\varepsilon\eta(\tilde{t}-t)}}{\varepsilon(1+\eta)}$ is the fixed cost associated with learning technology vintage $\tilde{t}$, ε captures the elasticity of the cost of technology across vintages, $\kappa_2\equiv\chi_0^{-\frac{\eta+1}{\eta}\frac{1}{\eta\varepsilon-1}}$ is defined so that the marginal team that decides to import technology rather than using the local one is given by χ_0. I assume that $\chi_0<f\left(\frac{1}{2},0\right)$, so that not all teams decide to import, even when talent is segmented. 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

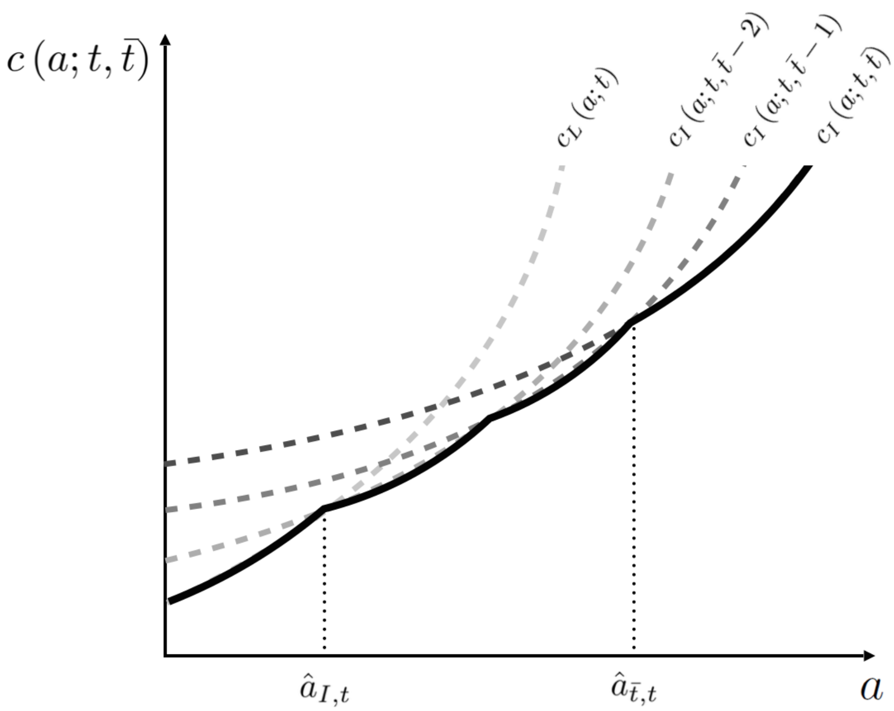
learnings, using and extracting proper returns from more advanced technologies.[18]

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

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

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

Figure 3: Cost of Technology



Optimal Technology. I next show that this system of functional forms provides simple solutions for the optimal technology α 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.

[18]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).

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*

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

where $\eta_\varepsilon \equiv \frac{\eta\varepsilon-1}{\varepsilon+1} < \eta$, $\hat{a}_{I,t} = a_t + \gamma^t\kappa_1\chi_0^{\frac{1}{\eta}}$ *and* $\hat{a}_{\bar{t},t} = a_t + \gamma^t\gamma^{\left(\frac{1+\varepsilon}{1+\eta}\right)\eta(\bar{t}-t)}\chi_0^{\frac{1}{\eta}}$.

Proof. See appendix. □

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*

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

where $\chi_1(t,\bar{t}) = \chi_0\gamma^{\frac{\eta(\eta\varepsilon-1)}{\eta+1}(\bar{t}-t)}$. *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. □

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:* $\forall (x',x,)\ \alpha(x',x;t',\bar{t}) \geq \alpha(x',x;t,\bar{t})$ *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:* $\frac{\alpha(x,y;t,\bar{t})}{\alpha(y,z;t,\bar{t})} = \frac{\tilde{\alpha}(x,y;\bar{t}-t)}{\tilde{\alpha}(y,z;\bar{t}-t)}$, *where* $\tilde{\alpha}(x',x;\bar{t}-t) = \alpha(x',x;t,\bar{t})\gamma^{-t}$;
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.[19] 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 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. 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 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. □

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.

[19] A more refined, and empirically documented, version of this argument is present in the work of Comin and Mestieri (2013).

Lemma 8: GDP per capita and Distance to the Frontier

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

$$Y(t,\bar{t}) = \gamma^{\bar{t}} \tilde{Y}(\bar{t}-t)$$

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

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 to completely characterize 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 $\varepsilon \to 0$.[20] The within vintage cost elasticity of technology (η) goes to infinite: conditional on a given technology vintage, all teams would pick identical productivity a. The vintage cost elasticity (ε) 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. Last, I assume that $\chi_0 \geq 0.75$.[21] 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).

Characterization. Lemmas 1-4 and Proposition 1 holds for this functional form. The main difference of this tractable case is the discrete nature of the optimal choice of technology.

[20]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.

[21]This assumption implies that, in equilibrium, at most half of the individuals use the best one among the two technologies.

Lemma 9: Optimal Technology for the Limit Case
The optimal technology for a team (x', x) *is given by*

$$\alpha(x', x) = \begin{cases} \gamma^t & x' < \chi_0 \\ \gamma^t + \gamma^{\bar{t}} & x' \geq \chi_0 \end{cases}.$$

Proof. See appendix. □

As a result, the technology gap takes only one of two values: $\frac{\alpha(m(x),x)}{\alpha(x,m^{-1}(x))} \in \left\{1; 1 + \gamma^{\bar{t}-t}\right\}$. It is equal to 1 if an individual would use the same technology if a manager or a worker. It is equal to $1 + \gamma^{\bar{t}-t}$ if an individual is itself less skilled than χ_0, but by being a worker is matched with a manager $m(x) \geq \chi_0$, thus getting access to the frontier technology vintage.

I define a notion of equilibrium shape, that captures the structure of equilibrium with respect to the occupational choice sets of managers and workers.

Definition 4: Equilibrium Shape
Let $\mathbb{X}_\omega$ *be the set of* $x \in (0,1)$ *at which the individual occupational choice changes, that is* $\mathbb{X}_\omega = \{x \in (0,1) : \lim_{\varepsilon \to 0} \omega(x - \varepsilon) \neq \lim_{\varepsilon \to 0} \omega(x + \varepsilon)\}$. *Let* $\mathbb{I}_\omega$ *be the cardinality of such set,* $\mathbb{I}_\omega = |\mathbb{X}_\omega|$. *An equilibrium shape is a sequence* $\{i_0, i_1, .., i_{\mathbb{I}_\omega}\}$ *where* $i_0 = \omega(0)$ *and for each component* $x_i \in \mathbb{X}_\omega$*, ordered such that* $x_1 < x_2 < .. < x_{\mathbb{I}_\omega}$*,* $i_i = 1$ *if* $\lim_{\varepsilon \to 0} \omega(x_i + \varepsilon) = 1$*,* $i_i = 0$ *if* $\lim_{\varepsilon \to 0} \omega(x_i + \varepsilon) = 0$*, and* $i_i = \frac{1}{2}$ *if* $\lim_{\varepsilon \to 0} \omega(x_i + \varepsilon) \in (0,1)$.

With general functional forms, the equilibrium can possibly take any one of infinite many shapes. The functional forms assumed in this section - by restricting the choice of technology to be discrete, and the marginal value of skills to not depend on individual's ability - limit considerably the number of attainable shapes.

Lemma 10: Possible Equilibrium Shapes
The equilibrium depicts, depending on parameter values, one of the following five shapes: $S_1 = (1,0)$*,* $S_2 = \left(1, \frac{1}{2}, 0\right)$*,* $S_3 = \left(1, \frac{1}{2}, 1, 0\right)$*,* $S_4 = \left(1, 0, \frac{1}{2}, 1, 0\right)$*,* $S_5 = (1,0,1,0)$.
Proof. See appendix. □

The five equilibrium shapes are pictured in Figure 4. Keeping all other parameters constant, the shape of the equilibrium allocation, and the resulting concentration of talent, is pinned down by the distance of a country from the frontier.

Proposition 4: Equilibrium Shapes and Distance to the Frontier
Let $S(d)$ *be the equilibrium shape for a country at distance* $d = \bar{t} - t$ *from the frontier. There exists four positve constants* $d_1 < d_2 < d_3 < d_4$ *such that if* $\bar{t} - t \leq d_1$*,* $S(d) = S_1$*; if* $\bar{t} - t \in \left(d_1, d_2\right]$*,* $S(d) = S_2$*; if* $\bar{t} - t \in \left(d_2, d_3\right]$*,* $S(d) = S_3$*; if* $\bar{t} - t \in \left(d_3, d_4\right]$*,* $S(d) = S_4$*; and if* $\bar{t} - t > d_4$*,*

$S(d) = S_5$.

Proof. See appendix. □

Corollary 3: Concentration of Talent and Distance to the Frontier

The concentration of talent increases in the distance from the technology frontier.

Proof. See appendix. □

Moreover, talent is segmented by occupation in countries very close to the frontier, while is segregated by technology in countries very far from it.[22]

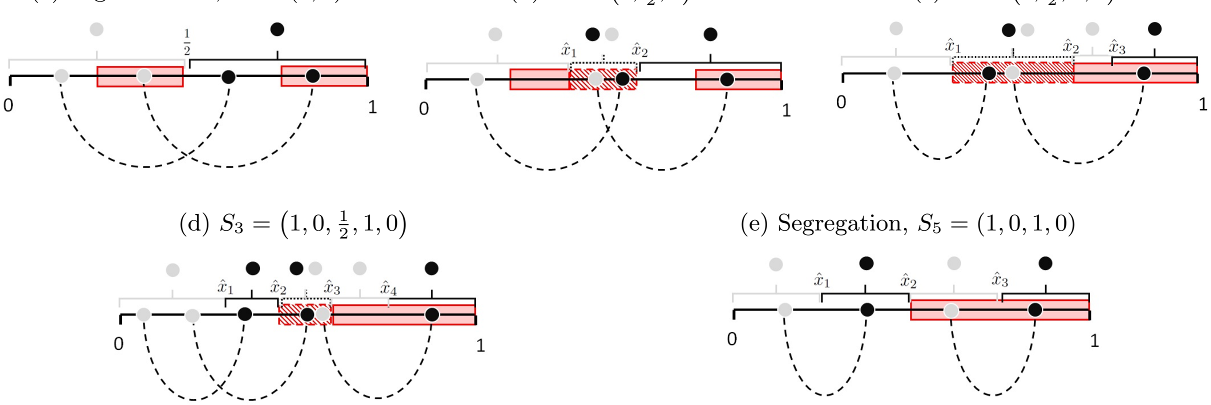
Corollary 4: Conditions for Segregation and Segmentation in the Tractable Case

If $\bar{t} - t \leq d_1$, talent is segmented by occupation; (ii) if $\bar{t} - t \geq d_4$, talent is segregated by technology.

Proof. See appendix. □

Figure 4: Five Equilibrium Shapes

(a) Segmentation, $S_1 = (1, 0)$ (b) $S_2 = \left(1, \frac{1}{2}, 0\right)$ (c) $S_3 = \left(1, \frac{1}{2}, 1, 0\right)$ (d) $S_3 = \left(1, 0, \frac{1}{2}, 1, 0\right)$ (e) Segregation, $S_5 = (1, 0, 1, 0)$

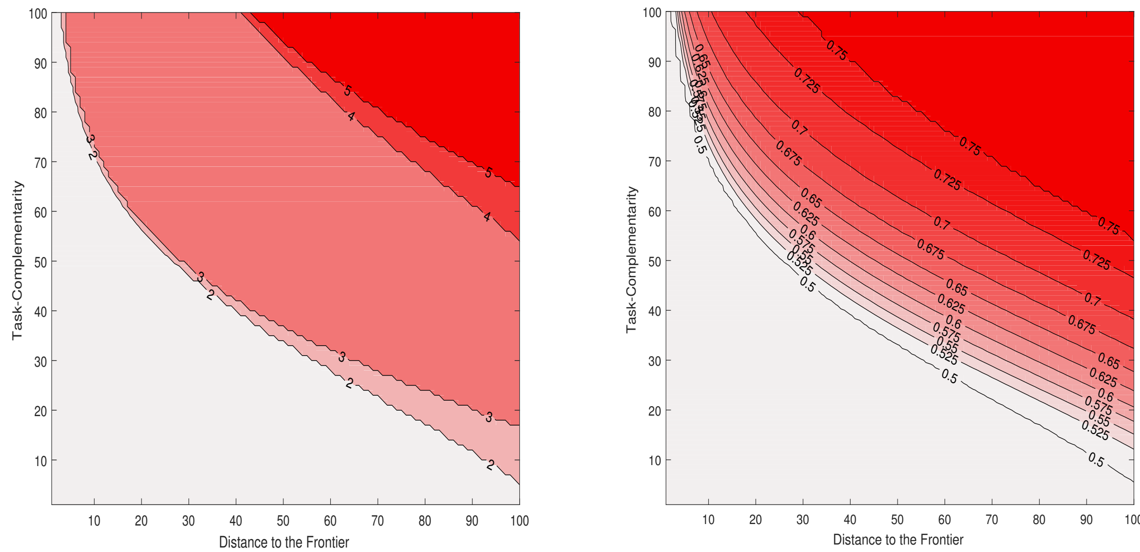


Notes: the squared brackets put together individuals with the same occupation. Workers are highlighted with light grey square brackets, and managers with black ones. Dotted brackets signal mixing: some are workers and some managers. The red regions covers the set of individuals using the frontier technology 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.

[22] Notice in fact that S_1 has talent segmentation, while S_5 has talent segregation. There is one main difference with respect to Corollary 1 in Section 2: here, 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.

Figure 5: Roles of Task-Complementarity and Distance to the Frontier

(a) Equilibrium Shape (b) Concentration of Talent



Notes: the left figure shows the contour plots of the subsets of parameter values in the space of λ (Task-Complementarity) and d (Distance to the Frontier) for which each equilibrium shape attains. Specifically, the gray area highlights the set of pairs (λ, d) for which talent is segmented (S_1). The line "2" then separates the region of S_2 to the one of S_1, and so on until the last shape, S_5. The right figure shows the contour plots, in the same space, for the concentration of talent. When talent is segmented, concentration of talent is equal to 0.5, its minimum value. When, instead, either d or λ are sufficiently high for talent not to be segmented, the concentration of talent increases smoothly in both parameters. Last, notice that both axis for $\lambda$and d are normalized on a scale from 1 to 100, even tought λ only takes values in $[0, 1]$ while I use values of d between 0 and 50.

Few features are worth discussing, keeping an eye on Figures 4 and 5.

First, the equilibrium may dictate that some types x are indifferrent between being managers or workers, that is $w(x) = \pi(x)$ and $w'(x) = \pi'(x)$ (applying Lemma 4). How can $\pi'(x) = w'(x)$ be satisfied? The slope of the matching function $m(x)$ depends on the fraction of workers at x, $\omega(x)$, and fraction of managers at $m(x)$, $1 - \omega(m(x))$. The restriction $w'(x) = \pi'(x)$ therefore pins down the unique value of $\omega(x)$ in the indifference regions.[23] *Second*, just as in the general case, the skill-asymmetry parameter λ plays an important role.[24] The higher is λ, the stronger task-complementarity is and therefore the lower the skill-asymmetry. As a result, concentration of talent increases in λ. To visualize the role of λ and d together, in Figure 5 I plot the countor plots of parameter values for which each equilibrium shape attains, and the correponding values of the concentration of talent. *Third*, as the distance from the frontier (or similarly task-complementarity) increases, the economic structure changes smoothly, in contrast with most frictionless matching models that feature discrete jump between two polar cases.[25]

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

[24] With this functional forms, $\frac{f_1(x, m^{-1}(x))}{f_2(m(x), x)} = \frac{1+\lambda m^{-1}(x)}{\lambda m(x)}$ that is decreasing in λ.

[25] 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)).

How does the equilibrium evolves as $\bar{t} - t$ increases? 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 $\bar{t} - t$ 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 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.[26] In some cases, it is even possible that some individuals would

[26] 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*

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 consistent 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.[27] 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 $x \sim U\left[0,1\right]$. 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:

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

these high capital–labour ratios, leaving even lower capital– labour ratios for the rest of the economy."

[27] 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.

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.[28]

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.[29]

Therefore, in the main empirical exercise, I take an indirect approach that exploits one of the assumption of the model, the complementarity between skills and technology. This assumption 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

[28] 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 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.

[29] 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.

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.

Last, at the end of this section I explore two alternatives empirical strategies. I first use occupation data, and compare average ability of managers and workers across countries. I then use the available firm level data to compare the distributions of workers to firms across countries. Both these alternative strategies support the model predictions. However, as I discuss in furhter details, they are both plagued by cross-country comparability concerns. For this reason, my main empirical exercises leverage the indirect, but transparent strategy that I next describe in details.

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 in the data. In order to have each country have similar sample size and avoid comparability concerns, I extract for each country year pair a random sample of 500,000 individuals.[30] 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 own-account 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 I.

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

[30] For some countries (13), I have fewer than 500,000 individuals. For all countries, however, I have at least 10,000 individuals that satisfy the sample selection criteria described below.

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 J, 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.[31] Second, I compute the average skill in each industry j: $\bar{\hat{x}}_j = E[\hat{x}_i | I_{ij} = 1]$, where I_{ij} is an indicator function equal to 1 if individual i works in industry j. The ranking of industries according to their average skill level provides the measure of the technology rank. 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 the highest skilled ones among the remaining workforce are then assigned into the second highest and so on.[32] Fourth, I compute the average skill in each industry under the perfect sorting counterfactual: $\bar{\hat{p}}_j = E\left[\hat{x}_i | I_{ij}^C = 1\right]$, where I_{ij}^C is the constructed indicator function. Fifth and last, I regress $\bar{\hat{x}}_j = B_0 + B_1 \bar{\hat{p}}_j + \varepsilon$.[33] The measure of concentration of talent is $\hat{\pi} = \hat{B}_1$. By the definition of the least squares estimator, $\hat{\pi} = \frac{E\left[\bar{\hat{x}}_j - \bar{\hat{x}}_{j'}\right]}{\bar{\hat{p}}_j - \bar{\hat{p}}_{j'}}$. This measure 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.

In Figure 6, I show two examples, Brazil in 2010, and United States in 1940, to illustrate how $\hat{\pi}$ is constructed. I plot the average skill in an industry, $\bar{\hat{x}}_j$, as a function of the skill in the perfect sorting counterfactual, $\bar{\hat{p}}_j$. Each dot in the figure corresponds to an industry and its size increases in the number of individuals there employed. A linear regression $\bar{\hat{x}}_j = \alpha + \pi \bar{\hat{p}}_j + \varepsilon$ fits the data well.[34] 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

[31]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.

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

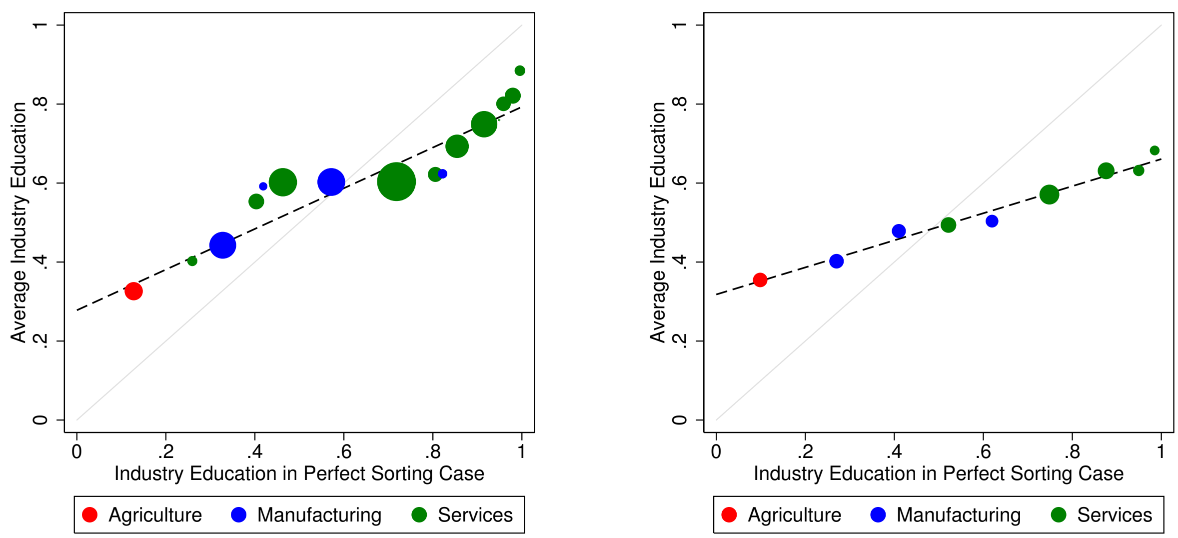
[33]I weight the regression by the number of individuals in each industry j. Unweighted results are similar and available upon request.

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

United States was in 1940. I build this measure of concentration of talent for each country-year pair in my sample and document systematic differences between countries depending on their distance from the frontier.[35]

Figure 6: Construction of the Measure of Concentration of Talent

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



Link to the 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 β. Let's solve for β.

[35]For brevity, I do not include Figure 6 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.

Market clearing implies that

$$\begin{aligned}
\frac{1}{2}\int_0^1 x_w(\tilde{a})\,d\tilde{a} + \frac{1}{2}\int_0^1 x_m(\tilde{a})\,d\tilde{a} &= \int_0^1 x dx \\
\frac{1}{4}\beta + \frac{1}{4}\beta + \frac{1}{2}\int_0^1 \left[x_m(\tilde{a}) - x_w(\tilde{a})\right] d\tilde{a} &= \frac{1}{2} \\
\beta &= 1 - \int_0^1 (m(x) - x)\,\omega(x)\,dx
\end{aligned}$$

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.[36] There are few concerns. First, education might measure skill with white noise.[37] 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

[36] 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.

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

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.[38] 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.[39] 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.

1^{st} **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 per capita relative to one of the United States in 2010.[40] The result is shown in Figure 7: poor countries, i.e. those farther from the technology frontier, have larger concentration of talent.[41] In order to interpret the magnitude of cross-country differences, it is useful to conduct the following thought experiment. Consider a country with two industries and two types of workers, high and low-skilled. Each industry is of equal size, and half of the population is high-skilled and half is low-skilled. If $\hat{\pi}$ in this economy is equal to 0 it means that half of the high skilled individuals are in each industry. If $\hat{\pi} = 1$ it means that all high skilled individuals are in one industry, which I call the modern one. If $\hat{\pi} = \frac{1}{2}$, instead, 75% of the high skilled are in the modern industry, and hence a high-skilled individual is three times more likely to work in the modern industry. Using this thought experiment, the estimates imply that high-skilled individuals in poor countries would be approximately five times as likely to work in the modern industry as the traditional one, while in rich ones, they would be only twice as likely. In Section 5 I further explore the quantitative properties of the model and show that these differences are sizable.

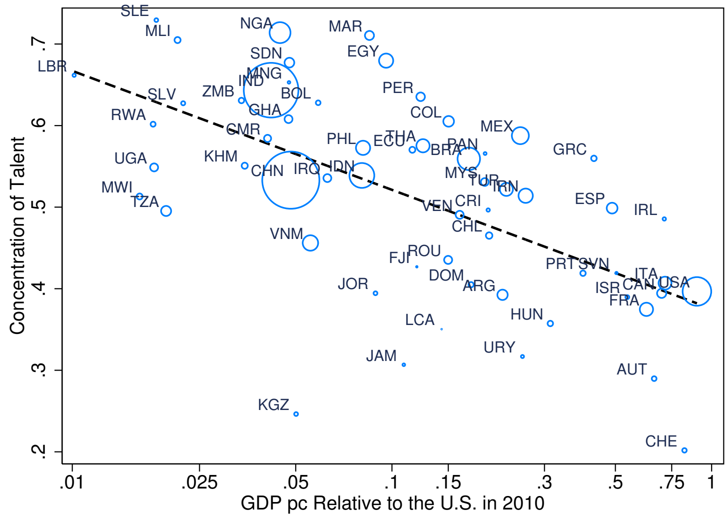
[38] See for example Caselli (2005) and Gollin et al. (2014) on agriculture productivity gaps, and Acemoglu and Zilibotti (2001) for a more disaggregated study.

[39] Recall: Lemma 8 shows that the GDP per capita of a country can be decomposed into a term that depend on the level of the frontier and one that is decreasing in the distance to the frontier: $Y(t, \bar{t}) = \gamma^t \tilde{Y}(\bar{t} - t)$ where, $\frac{\partial \tilde{Y}(\bar{t}-t)}{\partial(\bar{t}-t)} < 0$.

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

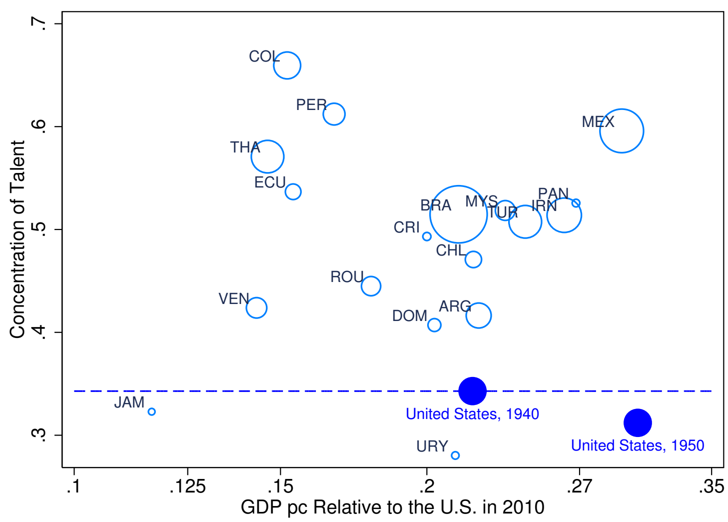
[41] The regression line has a positive slope that is significant at 1% level.

Figure 7: 1^{st} Comparison: Cross-county Differences in Concentration of Talent



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

Figure 8: 2^{nd} Comparison: Developing Countries today and U.S. in the past



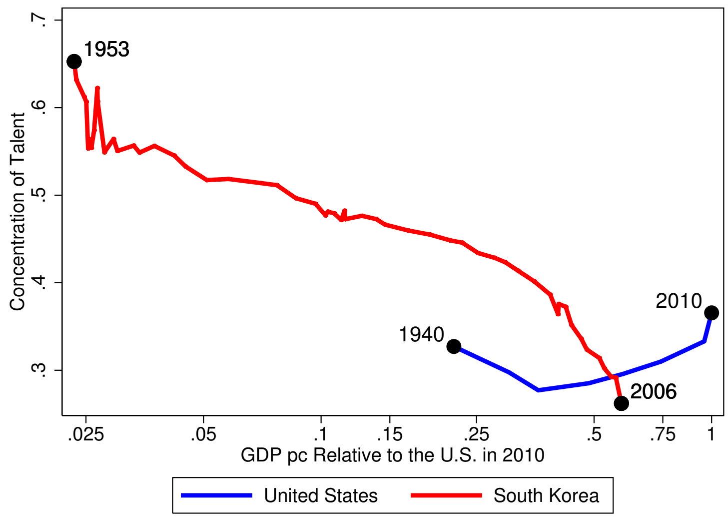
Notes: light blue circles show how populated each country is. The blue dotted line is at the level of concentration of talent of United States in 1940.

2^{nd} 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 frontier. Specifically, I have comparable census data for the United States every ten years from 1940 to 2010.[42] GDP per capita in the United States in 1940 is comparable to the one of many

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

middle income countries - such as Brazil, Mexico, Turkey, and Argentina - that I observed in my sample between 2000 and 2010. [43] I observe in fact 18 such countries, and among them, 16 have a higher concentration of talent than the U.S. used to have, as shown in Figure 8.[44]

Figure 9: 3^{rd} Comparison: South Korea as it Approaches the Frontier



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%.[45] In Figure 9, I plot the growth path for concentration of talent across sectors for both countries.[46] U.S. concentration of talent remained fairly constant along the growth path, consistent with fact that U.S. has been growing over this period constantly as a world leader, i.e. on the technology frontier. South Korea's concentration of talent instead decreased steeply. This last comparison alleviates the concern that cross-country differences might be driven by time invariant country characteristics that are correlated with GDP per capita.

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.

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

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

[45] These facts can be appreciated in Figure A.5.

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

Table 1: Robustness Table

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

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

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.[47] 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.[48]

Next, I explore robustness to alternative sample selections or measures of concentration of

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

[48] 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.

talent. All results are reported in Table 1, and I refer below to its rows. For brevity, I focus on the cross-sectional comparison.

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).[49] 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.[50] 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. $\bar{\hat{x}}_j = B_0 + B_1\bar{\hat{p}}_j + \varepsilon$. 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 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

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

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

4.5 Evidence from Occupation Data

IPUMS data report information on the individual occupation, coded following the International Standard Classification of Occupations (ISCO). ISCO occupations do *not* correspond to the notion of managers and workers in my model. In fact, ISCO definition of occupations seems

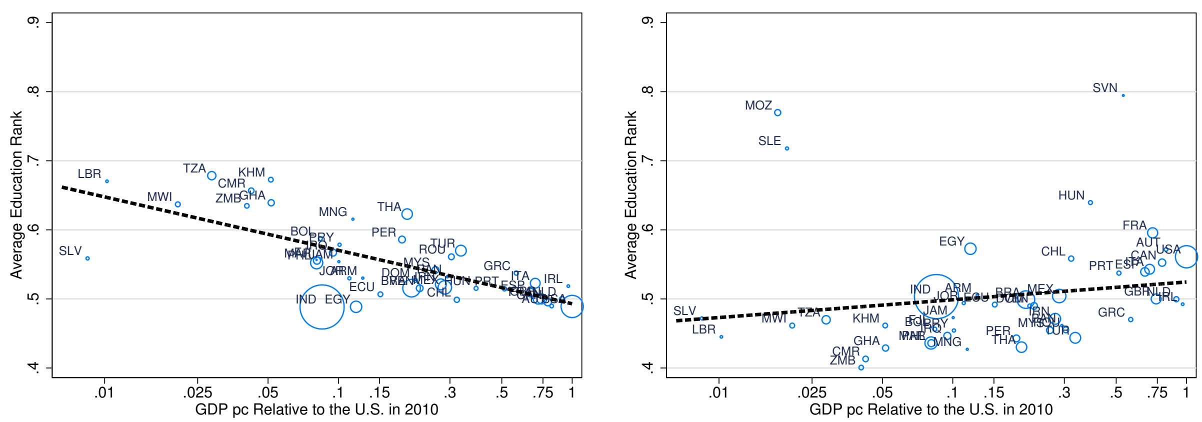
[49] Examples of Brazil 2010 and United States 1940 are in Figures A.1 and A.2 in the Section A. We can appreciate that the nice fit of the measure of concentration of talent is present at any level of aggregation.

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

to depend on the technology used: a manager that uses a backward technology would likely not be coded as a manager, but rather as an "elementary occupation". In fact, a manager according to ISCO is an occupation with skill level 4. Citing from their report available at ilo.org: "Occupations at this skill level (4) generally require extended levels of literacy and numeracy, sometimes at very high level..and 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..". Clearly, a manager of low technology firm in developing countries - that should be coded as a manager according to my model - does not fit well into this definition. As a result, I cannot directly use ISCO definition of managers to provide support on the model. Instead, I follow two alternative approaches.

Figure 10: Education Ranks by Occupation

(a) Workers (b) Managers



First, I build from the data two variables that most closely resemble the model definition of managers and workers. I code as managers individuals that are either managers according to the ISCO definition or report being self-employed. Managers of high technology firms are likely coded as ISCO-managers, while those at low technology, small, and informal firms are most likely captured in the data by self-employed. I code as workers all other employed individuals: they are wage-workers and not managers according to ISCO. I then calculate, in each country, the average rank in the education distribution (my empirical measure of ability x) for managers and workers. Recall that the measure of inverse concentration of talent in a country is given by the average ability distance between a worker and his manager, $\int (m(x) - x)\, \omega(x)\, dx$, where $\omega(x)$ is the probability that individual x is a worker. Using the market clearing we get that

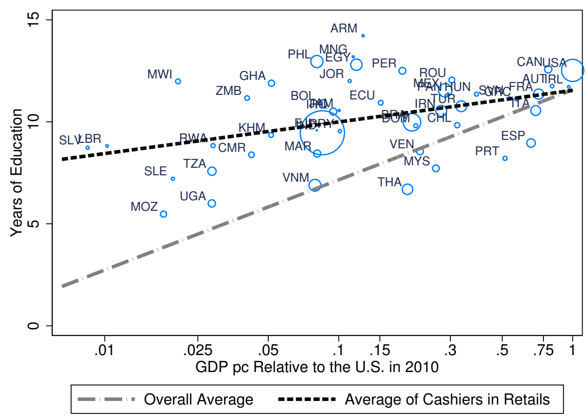
$$\underbrace{\int (m(x) - x)\, \omega(x)\, dx}_{\text{Concentration of Talent}} = \underbrace{\int x\,(1 - \omega(x))\, dx}_{\text{Average Manager Ability}} - \underbrace{\int x\omega(x)\, dx}_{\text{Average Worker Ability}} .$$

The model predicts that in poor countries the concentration of talent is higher - $\int (m(x) - x)\, \omega(x)\, dx$

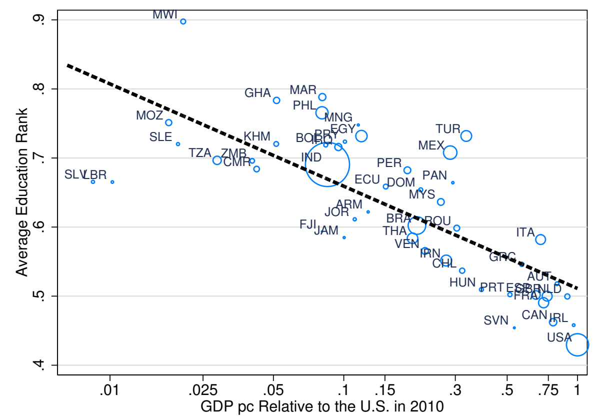
is lower. Therefore we should observe that the average ability of workers is higher in poor countries, and that the one of managers is lower. This prediction is supported in the data, as shown in Figure 10.[51]

Figure 11: Occupations in Retailing

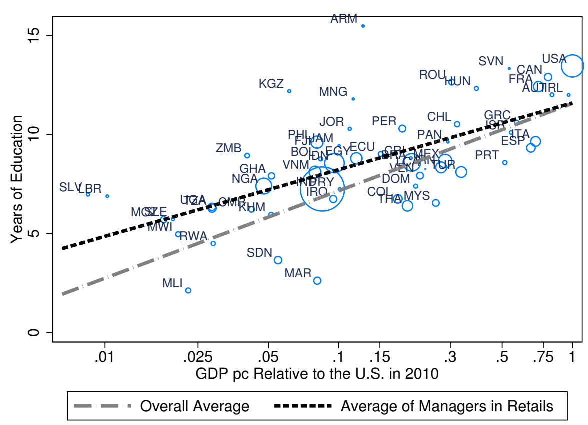
(a) Average Education of Retail Cashiers



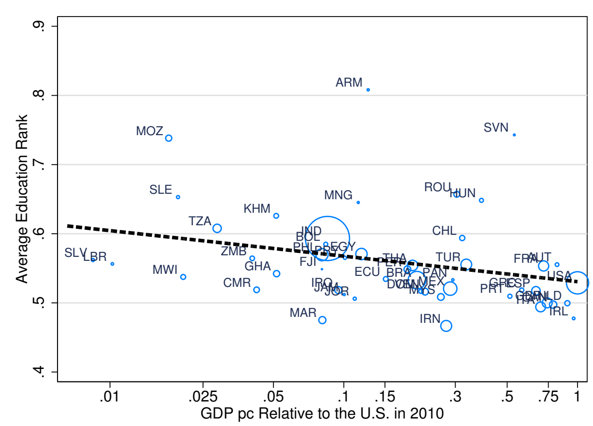
(b) Education Rank of Retail Cashiers



(c) Average Education of Retail Managers



(d) Education Rank of Retail Managers



This first approach has one main concern: the definition of workers and managers may be not comparable across countries. I thus take a second, specular, approach in which I attempt to code only one occupation that is highly comparable across countries, and show that it is performed by differently ranked individuals. I choose retail cashiers. In the language of the model, retail cashiers are "workers" that use an advanced technology: they perform a relatively simple task within their team, working below a store or chain manager, but are using an advanced technology, the cash

[51]The average ability of managers is only slightly lower in poor countries. However, this is driven by two counteracting forces. The ISCO-managers are significantly more skilled in poor countries, while the self-employed are significantly less-skilled. On net, managers are slightly less skilled in poor countries. This additional piece of evidence shows that there is larger dispersion of skills among managers in poor countries, consistent with the model predictions.

register. Retail cashier are comparable across countries. In almost all countries, the codebooks for the unharmonized occupation variable encompass the category cashier. The industry variable then allow to identify retail cashiers. [52] I compute their average years of schooling and normalized skills. The main mechanism of the model - the occupation-technology tradeoff - suggests that individuals that are in simple occupations, like a cashier, but use a modern technology should be relatively more skilled in poor countries than in rich ones. Figures 11a and 11b show that the data support this prediction. In poor countries retail cashier have on average fewer years of education than in rich ones, but are much more skilled relatively to the average education level of their country. In poor countries relatively high skilled individuals find their skills more useful in a simple occupation that pairs them with an advanced technology, a cash register. Last, we can confirm that the results are not driven by selection into retail by looking at the education of retail managers, as previously defined. Figures 11c and 11d shows that retail managers are only mildly more skilled in poor countries.

4.6 Firm Level Evidence

The World Bank Enterprise Surveys provide firm level information, from several countries around the world, on average education of workers and managers in the firm, and measures of technology and labor productivity. Unfortunately, these data suffers from major cross-country comparability concerns.[53] Most importantly, they cover only registered firms, that constitute a much larger fraction of the labor share in rich countries. Despite these concerns, this data set can be used to provide further validation for the model. I do so, in Appendix F. In interpret a team in the model as firm in the data. The three main findings are as follows. First, I validate the main prediction of the assumptions of skill-technology and manager-worker complementarity. I show that, similarly in rich and poor countries, more educated workers are employed in firms with higher technology (measured by computer usage), higher labor productivity, and more educated managers. Second, I show that - consistent with Proposition 3 - dispersion of used technology across firms is larger in less developed countries. Third, I show that talent - measured as years of education - is more concentrated within firms in less developed countries, as suggested by Corollary 2.

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

[52] Specifically, I call an individual retail cashier if he has the variable INDGEN = 60, and he is coded as cashier in the unharmonized occupation variable (OCC in IPUMS. Cashiers are identifiable from the codebooks available for each country). I managed to code retail cashiers for 54 countries. Beware that in some countries there are few retail cashiers. In fact, the median country has 30 observations, and the minimum number of observations is 5.

[53] This dataset has been recently used by Asker et al. (2014), who also discuss some of the cross-country comparability concerns.

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.[54] 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

$$\max_{a \in \mathbb{A}} \upsilon_a y\left(a,x\right)$$

where $\mathbb{A}$ is a discrete approximation[55] of the set of available technologies, υ_a is distributed according to a Frechet with dispersion parameter θ (dispersion of shocks increases as θ decreases)

$$\upsilon_a \sim e^{-\upsilon_a^{-\theta}},$$

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

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

[54] 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 $\frac{1}{2}\frac{(2N)!}{N!N!}$ possible partitions in two groups, and then we need to evaluate each possible matching for each partition. There is thus a total of $\frac{(2N)!}{2N!}$ 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.

[55] I describe in details in the Appendix G how this set is constructed for the computation.

Managers choose the optimal worker among the set Ω_a that gathers the individuals that are using technology a:

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

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

$$\int_0^{\hat{x}(a)} \phi(a,x)\,dx = \int_{\hat{x}(a)}^1 \phi(a,x)\,dx.$$

The income of an individual x is then given by

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

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 $\forall(x,a)$

$$\int_{\hat{x}(a)}^{m(a,x)} \phi(a,z)\,dz = \int_0^x \phi(a,z)\,dz.$$

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

$$\begin{aligned} w(a,x) &= \kappa + \int_0^x w_2(a,z)\,dz \\ \pi(a,x) &= w(a,\hat{x}(a)) + \int_{\hat{x}(a)}^x \pi_2(a,z)\,dz \end{aligned}$$

where κ satisfies

$$\int_0^{\hat{x}(a)} w(a,x)\,\phi(a,x)\,dx + \int_{\hat{x}(a)}^1 \pi(a,x)\,\phi(a,x)\,dx = \int_0^{\hat{x}(a)} (af(m(a,x),x) - c(a))\,\phi(a,x)\,dx.$$

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, $\frac{\phi(a,x)}{\sum_{\mathbb{A}}\phi(a,x)}$ - i.e. probability that an individual x selects into technology a - is given by

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

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

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

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 G. 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: $\frac{\partial}{\partial\lambda}\frac{f_1(x,y)}{f_2(y,z)} < 0\ \forall(x,y,z)$. $\lambda \leq 1$ 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 λ, the dispersion of individual-technology shocks θ, the within vintage cost elasticity of technology η, the across vintage cost elasticity ε, the level of the cost to import technology χ_0, the improvement across vintages γ, 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 γ 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 λ, θ, η, ε, and χ_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.[56]

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.[57] The first quartile has GDP per capita 78% of the one of the U.S., the second

[56] 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).

[57] 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).

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 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

$$\bar{p}(a) = \sum_{\tilde{a}=a_{min}}^{a_{-1}} \left[\int_0^1 \phi(\tilde{a},x)\,dx\right] + \frac{\int_0^1 \phi(a,x)\,dx}{2}$$

where a_{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 $\hat{\beta}_1$ 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 $A(a)$,

$$A(a) = \frac{\int_0^{\hat{x}(a)} (af(m(a,x),x) - c(a))\,\phi(a,x)\,dx}{\int_0^{\hat{x}(a)} \phi(a,x)\,dx},$$

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.[58] 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 H. The five moments are jointly determined by the five parameters. Nonetheless, I provide a discussion of the main links between moments and paramaters. λ decreases the skill-asymmetry, and thus, according to Proposition 2, a higher λ implies a higher concentration of talent. η changes the technology gap for teams that use the same technology vintage, since $\frac{\alpha(x,y)}{\alpha(x,y)} = \left(\frac{x(1+\lambda y)}{y(1+\lambda z)}\right)^{\frac{1}{\eta}}$. As a result, still by Proposition 2, a higher η decreases the concentration of talent. Additionally, η affects the agricultural productivity gap, since a higher η implies a lower agricultural gap, for given team composition. In countries close to the frontier, η and λ are the main determinants of the concentration of talent, since most teams use same vintage. Therefore, λ and η together are mostly relevant in matching the concentration of talent and agricultural productivity gap close to the frontier. ε 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 $\frac{\alpha(x,y)}{\alpha(y,z)} = \left(\frac{x(1+\lambda y)}{y(1+\lambda z)}\right)^{\frac{1}{\eta_\varepsilon}}$, where $\eta_\varepsilon = \frac{\eta\varepsilon-1}{\varepsilon+1}$ and is increasing in ε. The higher ε, the lower the increase in concentration of talent as we decrease the level of development t. In fact, the higher is ε, 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, χ_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 χ_0, the higher the decrease in GDP as we decrease the level of development t. Therefore, ε and χ_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 θ. The lower is θ, 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 θ thus seems not separately identified from λ. For this reason, I check the robustness of the results when I fix θ to arbitrary values, rather than estimating it. Additionally, I am currently working to target additional moments that allow to distinguish between θ and λ. The dispersion of average ability within industry, conditional on the concentration of talent, should allow to separately identify θ. 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.

[58] 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 abele than the average, while individuals in non-agriculture are more able. Details are in Appendix H.1.

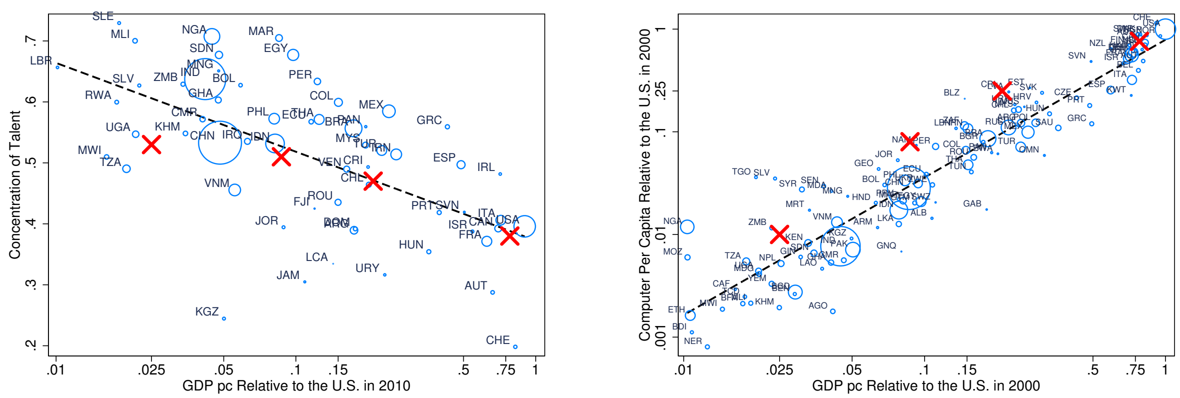
Fit of The Model. The model fits the data well. In Table 2, I display both the data targeted moments, and the model moments evaluated at the estimated parameter vector. To better visualize the results, I plot the model predicted concentration of talent together with the empirical one in Figure 12a.

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.[59] I plot the data along with the model predictions in Figure 12b. The model and the data alines very well.

Figure 12: Model Estimation

(a) Targeted Moments: Concentration of Talent

(b) Validation: Penetration Rate of Computers



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]

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

[59] Internet users is a very useful technology definition, because it only captures whether a technology is used an 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.

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

		Quartiles of World Income Distribution			
		Q1	*Q2*	*Q3*	*Q4*
	GDP pc Relative to U.S.	78%	21%	8.7%	2.5%
Data	Concentration of Talent	38%	47%	53%	62%
	Productivity Gap	2	3.2	3.4	5.6
Model	Concentration of Talent	36%	48%	51%	53%
	Productivity Gap	1.8	2.4	2.9	3.6

6 Few Further Discussions [PRELIMINARY]

In this section, I describe few extensions of the analytical model of Section 2.

Self-Employment. I consider the tractable framework of Section 3.2 and allow individuals to produce alone, rather than necessarily in teams. Details and empirical resutls are in the appendix Section C. I model self-employment along the lines of Garicano and Rossi-Hansberg (2004), as an occupation that is more skill intense than being a worker, but less so than being a manager. Therefore, when everyone uses similar technology, as in countries close to the technology frontier, self-employed are just as skilled as the average individual. This result, however, does not hold in countries far from the technology frontier. In fact, I show that the further a country is from the technology frontier the more negatively selected self-employed are. Intuitively, in countries far from the frontier, low skilled individuals become self-employed because those types that would be their managers in a country close to the frontier choose to be workers themselves to get access to the advanced technology. In this sense, the model provide an equilibrium explanation for the prevalence of subsistence self-employment in poor countries, a phenomenon widely documented in the literature (e.g. Schoar (2010), Ardagna and Lusardi (2008) and Banerjee and Duflo (2011)). Last, I use the same dataset described in Section 4 to document that, consistently with model predictions, own-account self-employed workers are relatively less skilled in countries further from the frontier.

Growth and the Dual Economy Trap.

[TO BE COMPLETED]

Firm Size. In the main model, a manager can only leverage his talent by hiring a more skilled worker, since he cannot hire more than one worker.In Section E, I show that this restriction can be relaxed while still preserving the main theoretical insights. I show the marginal values of skills for manages and workers when managers are allowed to choose firm size (i.e. the number of workers). Endogenous firm size does play a relevant role, but - as long as we assume a strnoger degree of complementarity between skills and technology, specifically log-supermodularity rather than supermodularity - the key technology-occupation tradeoff is still present and so is a role of the shape of the cost of technology in driving the assignment.[60]

7 Conclusion

This paper develops a theoretical framework to study the link between the allocation of talent into production teams and the cross-sectional distribution of technology. The theory highlights that the main purpose of team production, and thus the equilibrium assignment of talent, depends on the distance of a country from the technology frontier. In countries close to the frontier, all teams use similar technology, and the main purpose of team production is to gather together high and low skilled individuals to allow for task-specialization. In contrast, in countries far from the frontier, the main purpose of team production is to match individuals with their appropriate production technology. As a result, high skilled individuals gather together in teams that adopt most advanced tehcnology, and low skilled ones are left to match with one another and use backward technology. The theoretical predictions are supported by cross-country micro data on the allocation of talent. Moreover, a quantitative analysis demostrates that the observed cross-country differences in the allocation of talent can account for a sizable fraction of larger productivity dispersion in poor countries. The model thus suggests that we should be cautious to assign all the osberved cross-country differences to larger frictions in developing countries, as often done by the extensive literature that studies the (mis)allocation of resources in developing countries. Through the lens of the theory here proposed, those type of exercises are mispecified, because they do not take into account the endogenous team formation and technology choice.

More broadly, this paper has brought to the attention a new feature of economic development that has been previously overlooked. Individuals in poor countries form production teams differently than those in rich countries. This paper has focused on the implications of this phenomenon for cross-sectional productivity dispersion. However, the cross-country differences on "who matches with whom" may matter for several other core questions of economic development.

[60] These results are in line with previous work (e.g. Grossman et al. (2013)) that shows that log-supermularity is necessary to have positive assortative matching when firm size is endogenous.

For example, the different patterns of matching may influence the formation of social networks, or the degree through which information and knowledge spread out within an economy, thus ultimately affecting the growth rate of a country.[61]

[61] A recent literature (e.g. Lucas and Moll (2014), Perla and Tonetti (2014), and Buera and Oberfield (2014)) develops theoretical models in which growth is generated through the diffusion of ideas. If embedded within these models, the different patterns of assignment in developing countries would affect the overall speed of learning and possibly generate a trade-off between the allocation of talent that maximizes output and the one that maximizes technology diffusion. Similarly, Luttmer (2014) builds an assignment model in which student match with teachers for both production and learning. Evidence of slow accumulation of human capital over the life-cycle, that might be driven by slow learning due to lack of interaction of individuals of different skills, is documented in Lagakos et al. (2016).

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

_ **and Wilbur John Coleman**, "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, 2001, *109* (3), 584–616.

Chernozhukov, Victor and Han Hong, "An MCMC approach to classical estimation," *Journal of Econometrics*, 2003, *115* (2), 293–346.

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

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

Comin, Diego A and Bart Hobijn, "The CHAT dataset," Technical Report, National Bureau of Economic Research 2009.

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

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

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

Foster, Andrew D and Mark R Rosenzweig, "Technical change and human-capital returns and investments: evidence from the green revolution," *The American economic review*, 1996, pp. 931–953.

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

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

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

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

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

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

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

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

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

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

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

Keller, Elisa, Julieta Caunedo et al., "Capital Obsolescence and Agricultural Productivity," in "2016 Meeting Papers" number 686 Society for Economic Dynamics 2016.

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

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

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

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

_ **, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman**, "Life-Cycle Wage Growth Across Countries," 2016.

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

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

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

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

McFadden, Daniel, "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica: Journal of the Econometric Society*, 1989, pp. 995–1026.

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

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

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

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

Porzio, Tommaso, "Distance to the Technology Frontier and the Allocation of Talent," *Unpublished Manuscript*, 2016.

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

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

Suri, Tavneet, "Selection and Comparative Advantage in Technology Adoption," *Econometrica*, 2011, pp. 159–209.

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

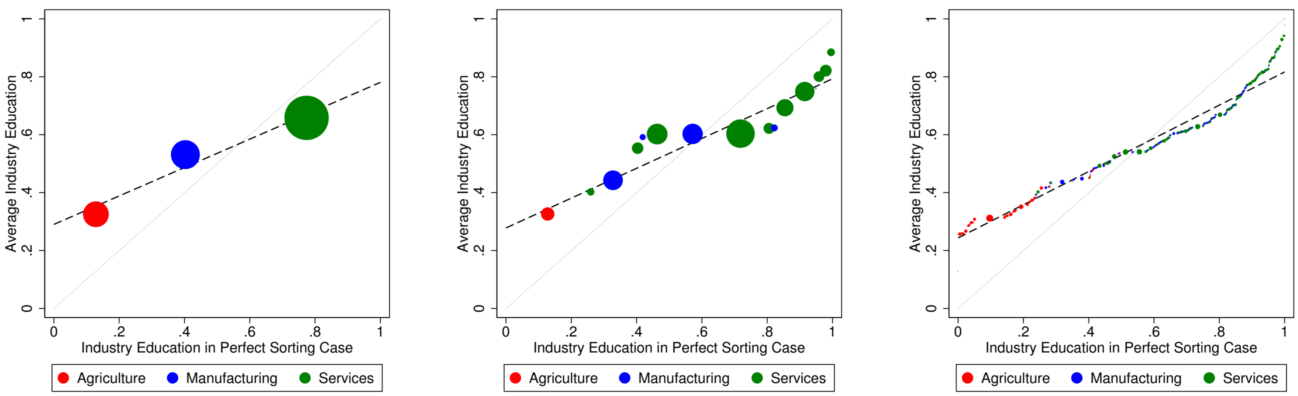
Appendix

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 all new results are included.]

A Additional Figures

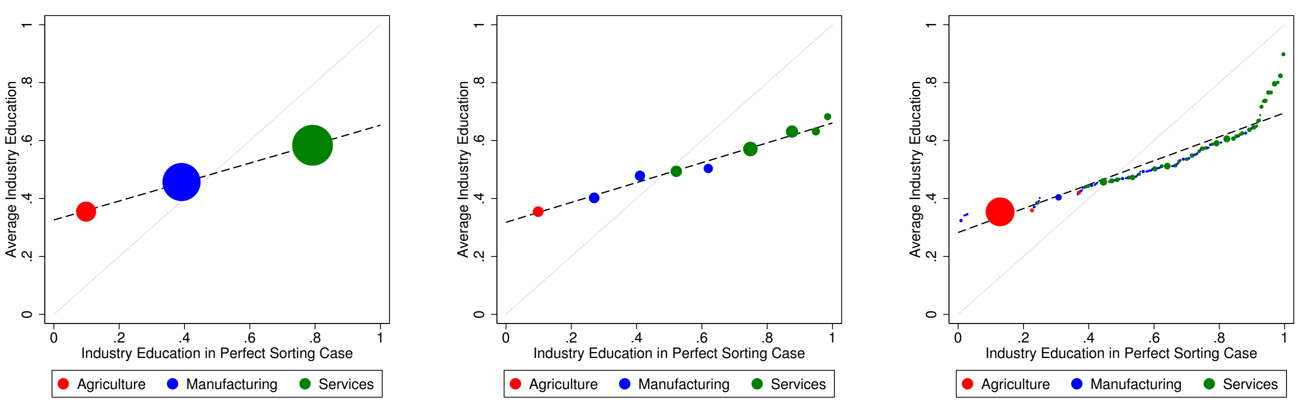
Figure A.1: Construction of Measure of Concentration of Talent, Brazil in 2010



(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 unrecoded industry, which in the case of Brazil is at the 3-digits level. In each figure, I plot the average normalized education in an industry as a function of the average education in a counterfactual scenario in which there is perfect sorting of individuals across industries. Each dot correspond to an industry, as defined, and the size of the dot is increasing in the number of individuals employed in that industry. The dotted lines are the prediction from a linear regression weighted by the number of individuals in each industry. The slopes of the regression lines are the measures of the concentration of talent, which for Brazil in 2010 are, from left to right, 0.49, 0.51, 0.57.

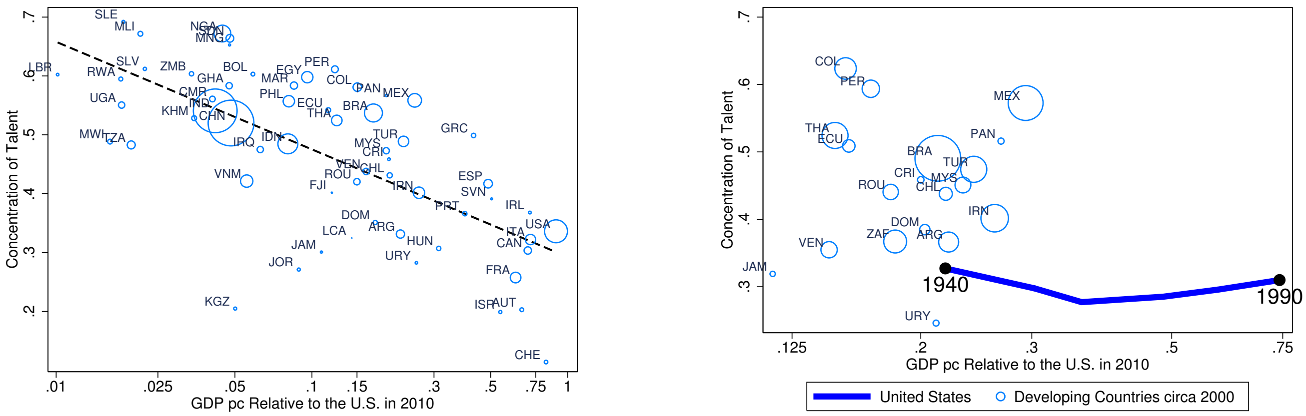
Figure A.2: Construction of Measure of Concentration of Talent, United States in 1940



(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.

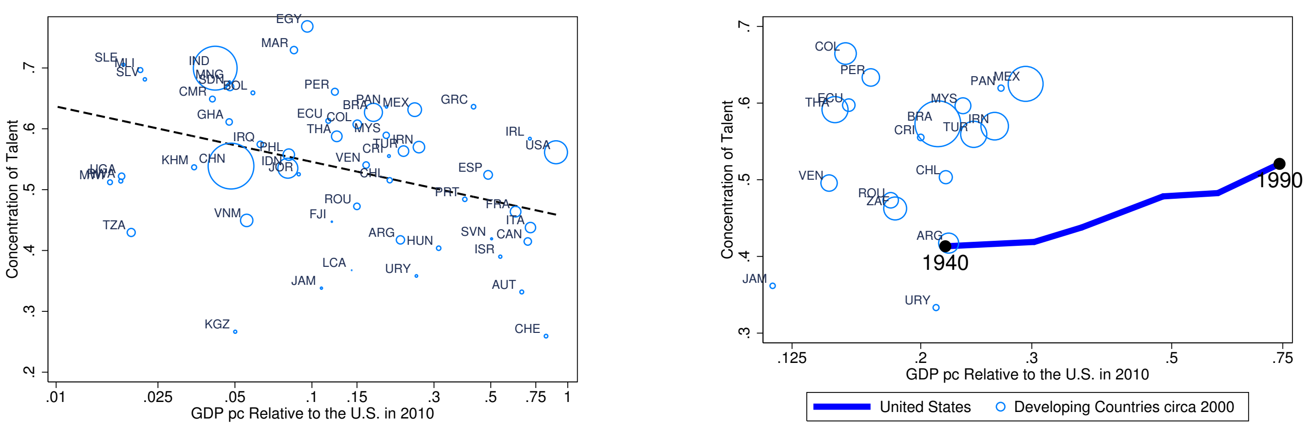
Figure A.3: Concentration of Talent Across Sectors



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

Notes: see Figures 7 and 8. The difference with respect to those figures is that here I plot the concentration of talent across the three sectors - agriculture, manufacturing, services - rather than industries. Sectors are recoded from the harmonized variable industry.

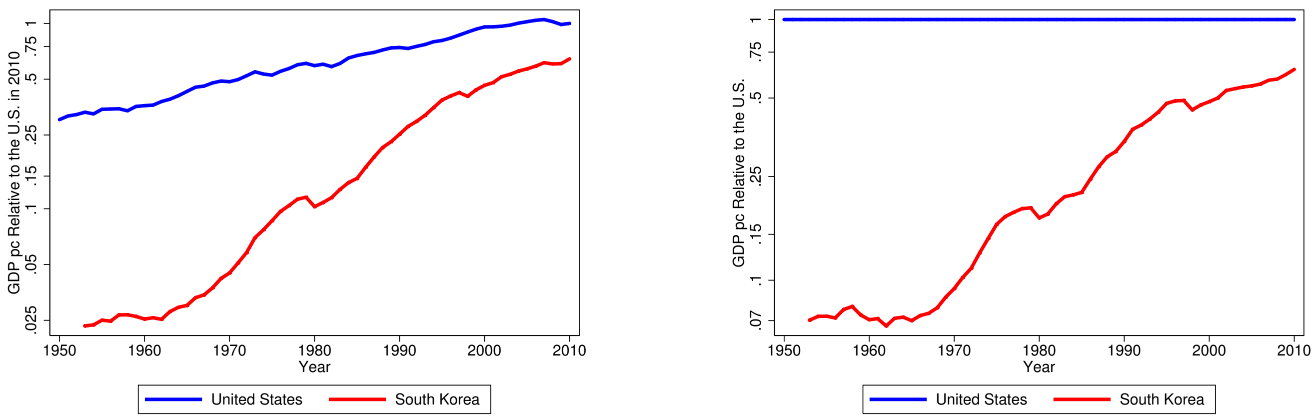
Figure A.4: Concentration of Talent Across Unrecoded Industries



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

Notes: see Figures 7 and 8. 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.

B Proofs

B.1 Proofs of Section 2

Proposition 1: Existence and Pareto Efficiency.

A competitive equilibrium exists and is Pareto Efficient.

Proof. The described competitive equilibrium is called in the matching literature a roomate problem (one-sided matching). A general theorem of an equilibrium existence in roomate problems when the number of types is even is shown in Chiappori et al. (2014). The proof is obtained by showing that the equilibrium is isomorphic to a two-sided matching problem, and thus to an optimal transportation problem, for which we have known existence results.[62] To prove Pareto efficiency, I proceed by contraddiction. Let's assume that the equilibrium is not Pareto efficient. Hence, there must be a possible deviation for an individual to change either partner, or occupation, or both and increase his income, without reducing the one of other individuals. However, this would violate the individual optimality conditions. Since if there is such a possibility an individual could take it and offer an ε fraction of his gain to his production partner to make him accept the deviation as well. □

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. The appropriate technology of a team (x', x) solves $f(x', x) = c'(\alpha(x', x))$. If we differentiate this first order condition with respect to x' we get

$$\alpha_1(x', x) = \frac{f_1(x', x)}{c''(f(x', x))}$$

with respect to x we get instead

$$\alpha_2(x', x) = \frac{f_2(x', x)}{c''(f(x', x))}.$$

Since, by assumption, $f_1 > f_2 \geq 0$, $\alpha_1(x', x) > \alpha_2(x', x)$, and the result is proved. □

Lemma 2: Matching Function

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

Proof. I prove the lemma by contraddiction. Let $x' > x$ and $m(x') < m(x)$. Optimality requires that

$$g\left(\alpha\left(m\left(x'\right), x'\right), m\left(x'\right), x'\right) + g\left(\alpha\left(m\left(x\right), x\right), m\left(x\right), x\right) \geq g\left(\alpha\left(m\left(x\right), x'\right), m\left(x\right), x'\right) + g\left(\alpha\left(m\left(x'\right), x\right), m\left(x'\right), x\right)$$

that can be rewritten, rearranging the inequality, using the fundamental theorem of calculus and the

[62] For details see also Chiappori et al. (2010). In Porzio (2016) I explicitly construct the isomorphic two-matching problem.

enveloped condition for α, as

$$\int_{m(x')}^{m(x)} \alpha\left(z, x'\right) f_1\left(z, x\right) dz \geq \int_{m(x')}^{m(x)} \alpha\left(z, x\right) f_1\left(z, x'\right) dz$$

$$\int_{m(x')}^{m(x)} \alpha\left(z, x\right)\left[f_1\left(z, x\right) - f_1\left(z, x'\right)\right] dz \geq 0$$

However, since $f_{12} > 0$, then $f_1\left(z, x\right) < f\left(z, x'\right)\ \forall z$, and thus this inequality does not hold, thus reaching a contraddiction. We still allow in theory for the possibility, that $m\left(x'\right) = m\left(x\right)$. If $\omega\left(x'\right) > 0$ and $\omega\left(x\right) > 0$, $m\left(x'\right) = m\left(x\right)$ would violate market clearing. This concludes the proof. □

Lemma 3: Occupational Choice within A Team

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

Proof. Let $x > x'$, with $m\left(x\right) = x'$. Then let $\hat{x}$ be the lowest type larger than x' with $\omega\left(\hat{x}\right) > 0$. Since $\omega\left(x\right) > 0$ and $\omega\left(x'\right) < 1$, $\hat{x} \in [x', x]$. By the necessary conditions, $w'\left(\hat{x}\right) \geq \pi'\left(\hat{x}\right)$. Substituting in Equation (1), $w'\left(\hat{x}\right) \geq \pi'\left(\hat{x}\right)$ becomes $\frac{\alpha(m(\hat{x}),\hat{x})}{\alpha(\hat{x},m^{-1}(\hat{x}))} \geq \frac{f_1\left(\hat{x},m^{-1}(\hat{x})\right)}{f_2(m(\hat{x}),\hat{x})}$. Lemma 1 showed that m and m^{-1} are increasing. As a result, since $x \geq \hat{x}$, then $m\left(\hat{x}\right) \leq x' \leq \hat{x}$, and also $m^{-1}\left(\hat{x}\right) \geq x \geq \hat{x}$. Therefore $\frac{\alpha(m(\hat{x}),\hat{x})}{\alpha(\hat{x},m^{-1}(\hat{x}))} \leq 1$, since α is increasing in both his arguments. The Spence-Mirlees assumption (3) guarantees that $\frac{f_1\left(\hat{x},m^{-1}(\hat{x})\right)}{f_2(m(\hat{x}),\hat{x})} > 1$. This lead to a contradiction, and thus to $m\left(x\right) \geq x$. □

Lemma 4: Necessary Conditions of Occupational Choice Function

The occupational choice function satisfies, for $\epsilon \to 0$ *:*

1. $\forall x$ such that $\omega\left(x - \epsilon\right) \in (0,1)$ and $\omega\left(x\right) \in (0,1)$: $\pi'\left(x\right) = w'\left(x\right)$

2. $\forall x$ such that $\omega\left(x - \epsilon\right) = 0$ and $\omega\left(x\right) > 0$: $\pi'\left(x\right) \leq w'\left(x\right)$

3. $\forall x$ such that $\omega\left(x - \epsilon\right) = 1$ and $\omega\left(x\right) < 1$: $\pi'\left(x\right) \geq w'\left(x\right)$.

Proof. If any one of the three conditions is not satisfied, I show that some individual is not choosing his occupation optimally. First, let (1) be not satisfied, for example $\pi'\left(x\right) < w'\left(x\right)$. Since $\omega\left(x\right) \in (0,1)$, by the occupational choice problem, $\pi\left(x\right) = w\left(x\right)$. Since $\pi'\left(x\right) < w'\left(x\right)$, $\pi\left(x - \epsilon\right) > w\left(x - \epsilon\right)$, and thus optimality requires that $\omega\left(x - \epsilon\right) = 0$, but this violates $\omega\left(x - \epsilon\right) \in (0,1)$. A similar argument applies if $\pi'\left(x\right) > w'\left(x\right)$. Next, let (2) be not satisfied, that is $\pi'\left(x\right) > w'\left(x\right)$. Since $\omega\left(x\right) > 0$, it must be that $w\left(x\right) \geq \pi\left(x\right)$, but then since $\pi'\left(x\right) > w'\left(x\right)$, $w\left(x - \epsilon\right) > \pi\left(x - \epsilon\right)$, and thus optimality requires $\omega\left(x - \epsilon\right) = 1$, which violates $\omega\left(x - \epsilon\right) = 0$. Last, let (3) be not satisfied, that is $\pi'\left(x\right) < w'\left(x\right)$. Since $\omega\left(x\right) < 1$, it must be that $\pi\left(x\right) \geq w\left(x\right)$. Since $\pi'\left(x\right) < w'\left(x\right)$, we get that $\pi\left(x - \epsilon\right) > w\left(x - \epsilon\right)$, leading to $\omega\left(x - \epsilon\right) = 0$, which violates $\omega\left(x - \epsilon\right) = 1$. □

Proposition 2: Assignment of Talent Across Teams.

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

1. *consider two functions* $\alpha(x',x)$ *and* $\alpha^{'}(x',x)$, *let* $\Upsilon(\iota)$, $\Lambda(\iota)$ *be the upper and lower bounds of the equilibrium with* α *and similarly* $\Upsilon^{'}(\iota)$ *and* $\Lambda^{'}(\iota)$ *be the bounds for the equilibrium with* α', *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* $\Upsilon(\iota)\geq\Upsilon'(\iota)$ *and* $\Lambda(\iota)\geq\Lambda'(\iota)$ $\forall\iota\in[0,1]$.

2. *consider two functions* $f(x',x)$ *and* $f^{'}(x',x)$, *let* $\Upsilon(\iota)$, $\Lambda(\iota)$ *be the upper and lower bounds of the equilibrium with* f *and similarly* $\Upsilon^{'}(\iota)$ *and* $\Lambda^{'}(\iota)$ *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* $\Upsilon(\iota)\geq\Upsilon'(\iota)$ *and* $\Lambda(\iota)\geq\Lambda'(\iota)$ $\forall\iota\in[0,1]$

Proof. I first consider the upper bound. Let

$$\begin{aligned}\hat{x} \quad &= \quad \min_{x\leq z} z\\ \text{s.t.} \quad &\omega(z)<1\end{aligned}$$

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

$$\frac{f_1\left(\hat{x},m^{-1}(\hat{x})\right)}{f_2(m(\hat{x}),\hat{x})}\geq\frac{\alpha(m(\hat{x}),\hat{x})}{\alpha(\hat{x},m^{-1}(\hat{x}))}.$$

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

$$\frac{f_1(\hat{x},x)}{f_2(m(\hat{x}),\hat{x})}\geq\frac{f_1\left(\hat{x},m^{-1}(\hat{x})\right)}{f_2(m(\hat{x}),\hat{x})}\geq\frac{\alpha(m(\hat{x}),\hat{x})}{\alpha(\hat{x},m^{-1}(\hat{x}))}\geq\frac{\alpha(m(\hat{x}),\hat{x})}{\alpha(\hat{x},x)}.$$

$\frac{\alpha(m(\hat{x}),\hat{x})}{\alpha(\hat{x},x)}$ 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:

$$m(\hat{x})-\hat{x} \quad < \quad \hat{m}(\hat{x})-\hat{x} \tag{4}$$

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

$$m(\hat{x})-m(x)\geq\int_x^{\hat{x}}\omega(z)\,dz=\hat{x}-x \tag{5}$$

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 $\left[0,\frac{1}{2}\right]$, 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

$$\begin{aligned}\tilde{\Upsilon}(x) &= \max_{\hat{x}\in[x,1]}\left\{\min\left\{\hat{m}(\hat{x})-\hat{x},\frac{1}{2}\right\}\right\}\\ &\text{s.t.}\quad \hat{x}-x\leq\hat{m}(\hat{x})\\ &\qquad\ \ \hat{m}(\hat{x})\leq 1.\end{aligned}$$

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

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 α 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

$$\begin{aligned}\tilde{x} &= \max_{z\leq x} z\\ &\text{s.t.}\quad \omega(z)<1\end{aligned}$$

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 $\tilde{x}\in[x-(m(x)-x),x]$, since by the definition of $\tilde{x}$, all individuals between $\tilde{x}$ and x are workers, and thus, even if all individuals between x and $m(x)$ are managers, then if $\tilde{x}<x-(m(x)-x)$, 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 $\tilde{x}$, $\pi'(\tilde{x})\leq w(\tilde{x})$. This last inequality can be rewritten as

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

and further, exploiting the fact that $f_{12}>0$ we get

$$\frac{f_1(\tilde{x},0)}{f_2(m(\tilde{x}),\tilde{x})}\leq\frac{f_1\left(\tilde{x},m^{-1}(\tilde{x})\right)}{f_2(m(\tilde{x}),\tilde{x})}\leq\frac{\alpha(m(\tilde{x}),\tilde{x})}{\alpha(\tilde{x},m^{-1}(\tilde{x}))}\leq\frac{\alpha(m(\tilde{x}),\tilde{x})}{\alpha(\tilde{x},0)}.$$

Next notice that $\frac{\alpha(m(\tilde{x}),\tilde{x})}{\alpha(\tilde{x},0)}$ is increasing in $m(\tilde{x})$, while $\frac{f_1(\tilde{x},0)}{f_2(m(\tilde{x}),\tilde{x})}$ is decreasing in $m(\tilde{x})$, as a result, there exist a value of $\tilde{m}(\tilde{x})$ small enough such that $\frac{f_1(\tilde{x},0)}{f_2(m(\tilde{x}),\tilde{x})}>\frac{\alpha(m(\tilde{x}),\tilde{x})}{\alpha(\tilde{x},0)}$ and thus $\pi'(\tilde{x})>w(\tilde{x})$, which would violate the necessary conditions. Hence we get that

$$m(\tilde{x}) - \tilde{x} \quad \geq \quad \tilde{m}(\tilde{x}) - \tilde{x}. \tag{6}$$

By the by the definition of $\tilde{x}$, the fact that m is increasing, and market clearing, we have that

$$\int_{\tilde{x}}^{x} \omega(z)\, dz = (x - \tilde{x}) = \int_{m(\tilde{x})}^{m(x)} (1 - \omega(z))\, dz \leq m(x) - m(\tilde{x}). \tag{7}$$

and combining equations 6 and 7 we have

$$m(x) - x \quad \geq \quad \tilde{m}(\tilde{x}) - \tilde{x}. \tag{8}$$

that gives a lower of $m(x) - x$ for a given $\tilde{x}$. However, as previously, $\tilde{x}$ is unknown, thus the lower bound is going to the be the minimum value $\tilde{m}(\tilde{x}) - \tilde{x}$ such that $\tilde{x}$ satisfied the relevant constraints, and such that $\Lambda(x) \in \left[0, \frac{1}{2}\right]$, as previously discussed

$$\begin{aligned} \tilde{\Lambda}(x) \quad &= \quad \min_{\tilde{x} \in [0,x]} \left\{ \max \left\{ \tilde{m}(\tilde{x}) - \tilde{x}, 0 \right\} \right\} \\ \text{s.t.} \quad & \hat{x} - x \leq \hat{m}(\hat{x}) \\ & \hat{m}(\hat{x}) \leq 1. \end{aligned}$$

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

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 $\alpha(x', x)$ 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. 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.

Proof. I first prove segmentation by occupation. If $\frac{\alpha(x,y)}{\alpha(y,z)} < \frac{f_1(x,y)}{f_2(y,z)} \forall (x, y, z) \in [0,1] \times [0,1] \times [0,1]$, with $x > y > z$. Then for any possible matching function m, the profit function is stepeer than the wage function - $\forall x \; \pi'(x) > w'(x)$. As a result they cross only once, which implies that the set of managers and workers are connected. Thus talent is segmented by occupation. More formally, we can notice that - by the necessary conditions of Lemma 4 - if $\pi'(x) > w'(x)$, there cannot be any manager less able than a worker, since that would read as $\omega(x - \epsilon) < 1$ and $\omega(x) > 0$, which requires that $\pi'(x) \leq w'(x)$.

Next, I provide segregation. I proceed by contraddiction. Let's assume that there is no segregation by technology. Hence, there exist a type $x' > x$, such that $\alpha(m(x), x) > \alpha\left(x', m^{-1}(x')\right)$ and $\omega(x') > 0$ and $\omega(x) < 0$. First, notice that for the inequality $\alpha(m(x), x) > \alpha\left(x', m^{-1}(x')\right)$ to satisfied, $m(x) > x'$. Next, I use the same type of argument made in the proof of Proposition 2. Let $\hat{x}$ be the smallest type bigger than x, such that $\omega(x) < 1$. We know that $\hat{x} \in [x, x']$ and that - by the necessary conditions - the following must hold $\alpha\left(\hat{x}, m^{-1}(\hat{x})\right) f_1\left(\hat{x}, m^{-1}(\hat{x})\right) \geq \alpha(m(\hat{x}), \hat{x}) f_2(m(\hat{x}), \hat{x})$. Since, $\frac{\alpha(x,y)}{\alpha(y,z)} \to \infty$, for this inequality to hold, $m(\hat{x}) \to \hat{x}$. Since by market clearing, by the definition of $\hat{x}$, and by the monotonicity of m we have that $m(x) - x \leq m(\hat{x}) - \hat{x}$, we get that $m(x) \to x$. This contraddicts $m(x) > x'$, since $x' > x$, and proves that talent is segregated by technology. □

B.2 Proofs of Section 3

Lemmas 5, 6, and 7

In order to prove these Lemmas. I solve the maximization problem for a team (x', x) in a country t, when the frontier is $\bar{t}$.

I first consider the problem of a team that choose optimal technology, using the local vintage t:

$$\max_a af(x', x) - \gamma^{-\eta t}\kappa_1^{-\eta}\frac{(a - a_t)^{1+\eta}}{1+\eta}.$$

Taking the first order condition we get

$$\alpha_L(x', x; t, \bar{t}) = a_t + \gamma^t \kappa_1 f(x', x)^{\frac{1}{\eta}}.$$

The net output $g_L(x', x; t, \bar{t}) = \max_a af(x', x) - \gamma^{-\eta t}\kappa_1^{-\eta}\frac{(a-a_t)^{1+\eta}}{1+\eta}$ is

$$g_L(x', x; t, \bar{t}) = a_t f(x', x) + \frac{\eta}{1+\eta}\gamma^t \kappa_1 f(x', x)^{\frac{1+\eta}{\eta}}.$$

Next, consider the problem of choosing the optimal vintage $\tilde{t}$ to minimize the cost of obtaining a technology a:

$$\min_{\tilde{t} \leq \bar{t}} \gamma^{-\eta\tilde{t}}\frac{(a - a_t)^{1+\eta}}{1+\eta} + \kappa_2^{-(\eta\varepsilon-1)}\frac{\gamma^t\gamma^{\varepsilon\eta(\tilde{t}-t)}}{\varepsilon(1+\eta)}.$$

The first order condition, if the constraint does not bind, gives

$$\begin{aligned}
\eta\log(\gamma)\gamma^{-\eta\tilde{t}^*}\frac{(a - a_t)^{1+\eta}}{1+\eta} &= \varepsilon\eta\log(\gamma)\gamma^{-[\eta\varepsilon-1]t}\kappa_2^{-(\eta\varepsilon-1)}\frac{\gamma^{\varepsilon\eta\tilde{t}}}{\varepsilon(1+\eta)} \\
\gamma^{-\eta\tilde{t}^*} &= \gamma^{-\eta_\varepsilon t}\kappa_2^{-\eta_\varepsilon}(a - a_t)^{-\frac{1+\eta}{1+\varepsilon}}.
\end{aligned}$$

The equation shows that $\tilde{t}^*$ is increasing in a. We can solve for the value of a such that the constraint $\tilde{t} \leq \bar{t}$ binds:

$$\begin{aligned}
\gamma^{-\eta\bar{t}} &= \gamma^{-\eta_\varepsilon t}\kappa_2^{-\eta_\varepsilon}\left(\hat{a}_{\bar{t},t} - a_t\right)^{-\frac{1+\eta}{1+\varepsilon}}. \\
\hat{a}_{\bar{t},t} &= a_t + \gamma^t\gamma^{\left(\frac{1+\varepsilon}{1+\eta}\right)\eta(\bar{t}-t)}\chi_0^{\frac{1}{\eta}}
\end{aligned}$$

where I also substituted $\kappa_2 = \chi_0^{-\frac{\eta+1}{\eta}\frac{1}{\eta\varepsilon-1}}$. Hence, if $a \leq \hat{a}_{\bar{t},t}$ we get that

$$\min_{\tilde{t} \leq \bar{t}} c_I(a, \tilde{t}; t) = \gamma^{-\eta\tilde{t}^*}\frac{(a - a_t)^{1+\eta}}{1+\eta} + \kappa_2^{-(\eta\varepsilon-1)}\frac{\gamma^t\gamma^{\varepsilon\eta(\tilde{t}^*-t)}}{\varepsilon(1+\eta)} = \gamma^{-\eta_\varepsilon t}\kappa_2^{-\eta_\varepsilon}\frac{(a - a_t)^{1+\eta_\varepsilon}}{1+\eta_\varepsilon},$$

while if $a \geq \hat{a}_{\bar{t},t}$ we get

$$\min_{\tilde{t}\leq\bar{t}} c_I\left(a,\tilde{t};t\right) = c_I\left(a,\bar{t};t\right).$$

Let's define $\tilde{c}_I\left(a;t,\bar{t}\right) = \min_{\tilde{t}\leq\bar{t}} c_I\left(a,\tilde{t};t\right)$. We can then solve the problem of optimal technology choice of a team that imports technology

$$\max_a af\left(x',x\right) - \tilde{c}_I\left(a;t,\bar{t}\right).$$

The first order condition gives

$$\alpha_I\left(x',x;t,\bar{t}\right) = \begin{cases} a_t + \gamma^t\kappa_2 x^{\frac{1}{\eta_\varepsilon}} & \text{if } f\left(x',x\right) \leq \chi_1\left(t,\bar{t}\right) \\ a_t + \gamma^{\bar{t}} x^{\frac{1}{\eta}} & \text{if } f\left(x',x\right) \geq \chi_1\left(t,\bar{t}\right) \end{cases}$$

where $\chi_1\left(t,\bar{t}\right)$ is the lowest ability team for which the constraint $\tilde{t} \leq \bar{t}$ binds, that is $\chi_1\left(t,\bar{t}\right)$ solves

$$\max_a a\chi_1\left(t,\bar{t}\right) - \gamma^{-\eta_\varepsilon t}\kappa_2^{-\eta_\varepsilon}\frac{\left(a-a_t\right)^{1+\eta_\varepsilon}}{1+\eta_\varepsilon}$$

$$a^*\left(\chi_1\left(t,\bar{t}\right)\right) = \hat{a}_{\bar{t},t}$$

where

$$a^*\left(\chi_1\left(t,\bar{t}\right)\right) = \arg\max_a a\chi_1\left(t,\bar{t}\right) - \gamma^{-\eta_\varepsilon t}\kappa_2^{-\eta_\varepsilon}\frac{\left(a-a_t\right)^{1+\eta_\varepsilon}}{1+\eta_\varepsilon},$$

hence we get

$$\begin{aligned} a_t + \gamma^t\kappa_2\chi_1\left(t,\bar{t}\right)^{\frac{1}{\eta_\varepsilon}} &= \hat{a}_{\bar{t},t} \\ \chi_1\left(t,\bar{t}\right) &= \chi_0\gamma^{\frac{\eta(\eta\varepsilon-1)}{\eta+1}\left(\bar{t}-t\right)}. \end{aligned}$$

Next, we need to calculate which teams decide to import, and which other teams decide instead to rely on local technology. The net output for a team that imports technology and is not bound by the frontier constraint is given by

$$g_I\left(x',x;t,\bar{t}\right) = a_t f\left(x',x\right) + \frac{\eta_\varepsilon}{1+\eta_\varepsilon}\gamma^t\kappa_2 f\left(x',x\right)^{\frac{1+\eta_\varepsilon}{\eta_\varepsilon}}.$$

A team (x',x) decides to import foreign technology as long as

$$\begin{aligned} g_I\left(x',x;t,\bar{t}\right) &> g_L\left(x',x;t,\bar{t}\right) \\ a_t f\left(x',x\right) + \frac{\eta_\varepsilon}{1+\eta_\varepsilon}\gamma^t\kappa_2 f\left(x',x\right)^{\frac{1+\eta_\varepsilon}{\eta_\varepsilon}} &> a_t f\left(x',x\right) + \frac{\eta}{1+\eta}\gamma^t\kappa_1 f\left(x',x\right)^{\frac{1+\eta}{\eta}} \\ \frac{\eta_\varepsilon}{1+\eta_\varepsilon}\kappa_2 f\left(x',x\right)^{\frac{1+\eta_\varepsilon}{\eta_\varepsilon}} &> \frac{\eta}{1+\eta}\kappa_1 f\left(x',x\right)^{\frac{1+\eta}{\eta}} \\ f\left(x',x\right)^{\frac{1+\eta_\varepsilon}{\eta_\varepsilon}-\frac{1+\eta}{\eta}} &> \frac{\kappa_1}{\kappa_2}\frac{\left(1+\eta_\varepsilon\right)\eta}{\left(1+\eta\right)\eta_\varepsilon} \\ f\left(x',x\right) &> \chi_0 \end{aligned}$$

where in the last line I substituted the values of κ_1, κ_2, and η_ε. We can then solve for $\hat{a}_{I,t}$ that is the technology chosen by the marginal type χ_0. This is given by

$$\begin{aligned}\hat{a}_{I,t} &= \arg\max_a a\chi_0 - \gamma^{-\eta t}\kappa_1^{-\eta}\frac{(a-a_t)^{1+\eta}}{1+\eta} \\ \hat{a}_{I,t} &= a_t + \gamma^t\kappa_1 f(x',x)^{\frac{1}{\eta}}.\end{aligned}$$

Therefore the optimal technology is given by

$$\alpha(x',x;t,\bar{t}) = \begin{cases}\alpha_L(x',x;t,\bar{t}) & \text{if } a \leq \hat{a}_{I,t} \\ \alpha_I(x',x;t,\bar{t}) & \text{if } a > \hat{a}_{I,t}\end{cases}$$

So far I have shown the shape of the cost function $c(a;t,\bar{t})$ and of the optimal technology $\alpha(x',x;t,\bar{t})$. Hence I've provided the proofs of Lemmas 5 and 6. Next, I need to show that the four properties of Lemma 7 are satisfied.

First notice that we can rewrite $\alpha(x',x;t,\bar{t})$ as

$$\alpha(x',x;t,\bar{t}) = \underbrace{\gamma^t\begin{cases}\nu + \kappa_1 f(x',x)^{\frac{1}{\eta}} & \text{if } f(x',x) \leq \chi_0 \\ \nu + \kappa_2 f(x',x)^{\frac{1}{\eta\varepsilon}} & \text{if } f(x',x) \in \left(\chi_0, \chi_0\gamma^{\frac{\eta(\eta\varepsilon-1)}{\eta+1}(\bar{t}-t)}\right) \\ \nu + \gamma^{\bar{t}-t} f(x',x)^{\frac{1}{\eta}} & \text{if } f(x',x) \geq \chi_0\gamma^{\frac{\eta(\eta\varepsilon-1)}{\eta+1}(\bar{t}-t)}\end{cases}}_{\equiv\tilde{\alpha}(x',x;\bar{t}-t)}$$

where I have substituted $\chi_1(t,\bar{t}) = \chi_0\gamma^{\frac{\eta(\eta\varepsilon-1)}{\eta+1}(\bar{t}-t)}$ solved above. $\chi(t,\bar{t})$ increases in $\bar{t}-t$, since by assumption $\eta\varepsilon > 1$. We thus have proved properties 3. and 4. of the Lemma 7. Next, we need to shows properties 1. and 2., namely that $\alpha(x',x;t,\bar{t})$ is increasing in both t and $\bar{t}$. Let's first show that is increasing in t. Consider any two $t' > t$. As shown, $\chi_1(t',\bar{t}) < \chi_1(t,\bar{t})$. For any (x',x) such that $f(x',x) < \chi_0$, $\alpha(x',x;t',\bar{t}) - \alpha(x',x;t,\bar{t}) = \left(\gamma^{t'} - \gamma^t\right)\left(\nu + \kappa_1 f(x',x)^{\frac{1}{\eta}}\right) > 0$. For any (x',x) such that $f(x',x) \in [\chi_0, \chi_1(t',\bar{t})]$, then $\alpha(x',x;t',\bar{t}) - \alpha(x',x;t,\bar{t}) = \left(\gamma^{t'} - \gamma^t\right)\left(\nu + \kappa_2 f(x',x)^{\frac{1}{\eta\varepsilon}}\right) > 0$. For any (x',x) such that $f(x',x) \geq \chi_1(t,\bar{t})$, $\alpha(x',x;t',\bar{t}) - \alpha(x',x;t,\bar{t}) = \left(\gamma^{t'} - \gamma^t\right)v + \gamma^{\bar{t}} f(x',x)^{\frac{1}{\eta}} \geq 0$. Last, we need to consider the case when (x',x) is such that $f(x',x) \in [\chi_1(t',\bar{t}), \chi_1(t,\bar{t})]$, then $\alpha(x',x;t',\bar{t}) - \alpha(x',x;t,\bar{t}) = \left(\gamma^{t'} - \gamma^t\right)v + \gamma^{\bar{t}} f(x',x)^{\frac{1}{\eta}} - \gamma^t\kappa_2 f(x',x)^{\frac{1}{\eta\varepsilon}} \geq \gamma^{\bar{t}} f(x',x)^{\frac{1}{\eta}} - \gamma^t\kappa_2 f(x',x)^{\frac{1}{\eta\varepsilon}}$, and $\gamma^{\bar{t}} f(x',x)^{\frac{1}{\eta}} - \gamma^t\kappa_2 f(x',x)^{\frac{1}{\eta\varepsilon}} \geq 0$ if and only if $f(x',x) \leq \kappa_2^{\frac{\eta\eta\varepsilon}{\eta\varepsilon-\eta}}\gamma^{\frac{\eta\eta\varepsilon}{\eta-\eta\varepsilon}(\bar{t}-t)} = \chi_1(t,\bar{t})$, which holds since we are considering $f(x',x) \in [\chi_1(t',\bar{t}), \chi_1(t,\bar{t})]$. Let's next show that is increasing in $\bar{t}$. Consider any two $\bar{t}' > \bar{t}$. As shown, $\chi_1(t,\bar{t}') > \chi_1(t,\bar{t})$. For any (x',x) such that $f(x',x) \leq \chi_1(t,\bar{t})$, $\alpha(x',x;t,\bar{t}') - \alpha(x',x;t,\bar{t}) = 0$. For any (x',x) such that $f(x',x) \geq \chi_1(t,\bar{t}')$, $\alpha(x',x;t,\bar{t}') - \alpha(x',x;t,\bar{t}) = \left(\gamma^{\bar{t}'} - \gamma^{\bar{t}}\right) f(x',x)^{\frac{1}{\eta}} > 0$. Last, consider any (x',x) such that $f(x',x) \in (\chi_1(t,\bar{t}), \chi_1(t,\bar{t}'))$, then $\alpha(x',x;t,\bar{t}') - \alpha(x',x;t,\bar{t}) = \gamma^{\bar{t}'-t} f(x',x)^{\frac{1}{\eta}} - \kappa_2 f(x',x)^{\frac{1}{\eta\varepsilon}}$ and $\gamma^{\bar{t}'-t} f(x',x)^{\frac{1}{\eta}} - \kappa_2 f(x',x)^{\frac{1}{\eta\varepsilon}} \geq 0$ if and only if $f(x',x) \leq \kappa_2^{\frac{\eta\eta\varepsilon}{\eta\varepsilon-\eta}}\gamma^{\frac{\eta\eta\varepsilon}{\eta-\eta\varepsilon}(\bar{t}'-t)} = \chi_1(t,\bar{t}')$, which holds since we are considering $f(x',x) \in [\chi_1(t,\bar{t}), \chi_1(t,\bar{t}')]$. This concludes the proof.

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 $x = f(y, z)$, so that $x' > x$. 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 $\eta_d(x', x) \equiv \frac{\alpha(x,y;t,\bar{t})}{\alpha(y,z;t,\bar{t})}$. Let's now consider each case one by one:

1. If $x' \leq \chi_0$: then $\eta_d(x', x) = \left(\frac{\nu + \kappa_1 x'^{\frac{1}{\eta}}}{\nu + \kappa_1 x^{\frac{1}{\eta}}}\right) \forall d$

2. If $x \leq \chi_0 \leq x' \leq \chi_{1,d}$: then $\eta_d(x', x) = \left(\frac{\nu + \kappa_2 x'^{\frac{1}{\eta_\varepsilon}}}{\nu + \kappa_1 x^{\frac{1}{\eta}}}\right) \forall d$

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

$$\begin{aligned}
\eta_{d'}(x', x) &= \left(\frac{\nu + \kappa_2 x'^{\frac{1}{\eta_\varepsilon}}}{\nu + \kappa_1 x^{\frac{1}{\eta}}}\right) \\
\eta_d(x', x) &= \left(\frac{\nu + \gamma^d x'^{\frac{1}{\eta}}}{\nu + \kappa_1 x^{\frac{1}{\eta}}}\right)
\end{aligned}$$

   and thus

$$\begin{aligned}
\eta_{d'}(x', x) &\geq \eta_d(x', x) \\
\left(\frac{\nu + \kappa_2 x'^{\frac{1}{\eta_\varepsilon}}}{\nu + \kappa_1 x^{\frac{1}{\eta}}}\right) &\geq \left(\frac{\nu + \gamma^d x'^{\frac{1}{\eta}}}{\nu + \kappa_1 x^{\frac{1}{\eta}}}\right) \\
\kappa_2 x'^{\frac{1}{\eta_\varepsilon}} &\geq \gamma^d x'^{\frac{1}{\eta}} \\
\log x' &\geq \log \kappa_0 + \frac{\eta\eta_\varepsilon}{\eta - \eta_\varepsilon} d \log \gamma \\
\log x' &\geq \chi_{1,d}
\end{aligned}$$

   and the last inequality holds by assumption.

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

$$\begin{aligned}
\eta_{d'}(x', x) &= \left(\frac{\nu + \gamma^{d'} x'^{\frac{1}{\eta}}}{\nu + \kappa_1 x^{\frac{1}{\eta}}}\right) \\
\eta_d(x', x) &= \left(\frac{\nu + \gamma^d x'^{\frac{1}{\eta}}}{\nu + \kappa_1 x^{\frac{1}{\eta}}}\right)
\end{aligned}$$

and thus

$$\begin{aligned} \eta_{d'}\left(x',x\right) &\geq \eta_d\left(x',x\right) \\ \left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\kappa_1 x^{\frac{1}{\eta}}}\right) &\geq \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\kappa_1 x^{\frac{1}{\eta}}}\right) \\ \gamma^{d'}x'^{\frac{1}{\eta}} &\geq \gamma^{d}x'^{\frac{1}{\eta}} \\ d' &\geq d \end{aligned}$$

and again, the last inequality holds by assumption.

5. If $\chi_0 \leq x \leq x' \leq \chi_{1,d}$: then $\eta_d\left(x',x\right) = \left(\frac{\nu+\kappa_2 x'^{\frac{1}{\eta_\varepsilon}}}{\nu+\kappa_2 x^{\frac{1}{\eta_\varepsilon}}}\right) \forall d$

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

$$\begin{aligned} \eta_{d'}\left(x',x\right) &= \left(\frac{\nu+\kappa_2 x'^{\frac{1}{\eta_\varepsilon}}}{\nu+\kappa_2 x^{\frac{1}{\eta_\varepsilon}}}\right) \\ \eta_d\left(x',x\right) &= \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\kappa_2 x^{\frac{1}{\eta_\varepsilon}}}\right) \end{aligned}$$

and thus

$$\begin{aligned} \eta_{d'}\left(x',x\right) &\geq \eta_d\left(x',x\right) \\ \left(\frac{\nu+\kappa_2 x'^{\frac{1}{\eta_\varepsilon}}}{\nu+\kappa_2 x^{\frac{1}{\eta_\varepsilon}}}\right) &\geq \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\kappa_2 x^{\frac{1}{\eta_\varepsilon}}}\right) \\ \kappa_2 x'^{\frac{1}{\eta_\varepsilon}} &\geq \gamma^{d}x'^{\frac{1}{\eta}} \\ \log x' &\geq \log\kappa_0 + \frac{\eta\eta_\varepsilon}{\eta-\eta_\varepsilon} d\log\gamma \\ \log x' &\geq \chi_{1,d} \end{aligned}$$

which holds by assumption.

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

$$\begin{aligned} \eta_{d'}\left(x',x\right) &= \left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\kappa_2 x^{\frac{1}{\eta_\varepsilon}}}\right) \\ \eta_d\left(x',x\right) &= \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\kappa_2 x^{\frac{1}{\eta_\varepsilon}}}\right) \end{aligned}$$

and thus

$$\begin{aligned}
\eta_{d'}\left(x',x\right) &\geq \eta_d\left(x',x\right)\\
\left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\kappa_2 x^{\frac{1}{\eta\varepsilon}}}\right) &\geq \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\kappa_2 x^{\frac{1}{\eta\varepsilon}}}\right)\\
\gamma^{d'}x'^{\frac{1}{\eta}} &\geq \gamma^{d}x'^{\frac{1}{\eta}}\\
d' &\geq d
\end{aligned}$$

and again, the last inequality holds by assumption.

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

$$\begin{aligned}
\eta_{d'}\left(x',x\right) &= \left(\frac{\nu+\kappa_2 x'^{\frac{1}{\eta\varepsilon}}}{\nu+\kappa_2 x^{\frac{1}{\eta\varepsilon}}}\right)\\
\eta_d\left(x',x\right) &= \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\gamma^{d}x^{\frac{1}{\eta}}}\right)
\end{aligned}$$

This requires a bit more steps.

$$\begin{aligned}
\eta_{d'}\left(x',x\right) &\geq \eta_d\left(x',x\right)\\
\left(\frac{\nu+\kappa_2 x'^{\frac{1}{\eta\varepsilon}}}{\nu+\kappa_2 x^{\frac{1}{\eta\varepsilon}}}\right) &\geq \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\gamma^{d}x^{\frac{1}{\eta}}}\right)\\
\left(\nu+\kappa_2 x'^{\frac{1}{\eta\varepsilon}}\right)\left(\nu+\gamma^{d}x^{\frac{1}{\eta}}\right) &\geq \left(\nu+\gamma^{d}x'^{\frac{1}{\eta}}\right)\left(\nu+\kappa_2 x^{\frac{1}{\eta\varepsilon}}\right)\\
\nu^2+\nu\gamma^d x^{\frac{1}{\eta}}+\nu\kappa_2 x'^{\frac{1}{\eta\varepsilon}}+\kappa_2 x'^{\frac{1}{\eta\varepsilon}}\gamma^d x^{\frac{1}{\eta}} &\geq \nu^2+\nu\kappa_2 x^{\frac{1}{\eta\varepsilon}}+\nu\gamma^d x'^{\frac{1}{\eta}}+\gamma^d x'^{\frac{1}{\eta}}\kappa_2 x^{\frac{1}{\eta\varepsilon}}\\
\nu\left[\left(\kappa_2 x'^{\frac{1}{\eta\varepsilon}}-\gamma^d x'^{\frac{1}{\eta}}\right)-\left(\kappa_2 x^{\frac{1}{\eta\varepsilon}}-\gamma^d x^{\frac{1}{\eta}}\right)\right]+\kappa_2\gamma^d\left(x'^{\frac{1}{\eta\varepsilon}}x^{\frac{1}{\eta}}-x'^{\frac{1}{\eta}}x^{\frac{1}{\eta\varepsilon}}\right) &\geq 0
\end{aligned}$$

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

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

$$\begin{aligned}
\eta_{d'}\left(x',x\right) &= \left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\kappa_2 x^{\frac{1}{\eta\varepsilon}}}\right)\\
\eta_d\left(x',x\right) &= \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\gamma^{d}x^{\frac{1}{\eta}}}\right)
\end{aligned}$$

and thus

$$\begin{aligned}\eta_{d'}\left(x',x\right) &\geq \eta_d\left(x',x\right)\\ \left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\kappa_2 x^{\frac{1}{\eta\varepsilon}}}\right) &\geq \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\gamma^d x^{\frac{1}{\eta}}}\right)\end{aligned}$$

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

$$\nu+\kappa_2 x^{\frac{1}{\eta\varepsilon}} \leq \nu+\gamma^{d'}x^{\frac{1}{\eta}}$$

and thus

$$\left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\kappa_2 x^{\frac{1}{\eta\varepsilon}}}\right) \geq \left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\gamma^{d'}x^{\frac{1}{\eta\varepsilon}}}\right)$$

and then (see below)

$$\left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\gamma^{d'}x^{\frac{1}{\eta\varepsilon}}}\right) \geq \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\gamma^d x^{\frac{1}{\eta}}}\right)$$

since $d' \geq d$ and $x' \geq x$.

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

$$\begin{aligned}\eta_{d'}\left(x',x\right) &= \left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\gamma^{d'}x^{\frac{1}{\eta}}}\right)\\ \eta_d\left(x',x\right) &= \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\gamma^d x^{\frac{1}{\eta}}}\right)\end{aligned}$$

and thus

$$\begin{aligned}\eta_{d'}\left(x',x\right) &\geq \eta_d\left(x',x\right)\\ \left(\frac{\nu+\gamma^{d'}x'^{\frac{1}{\eta}}}{\nu+\gamma^{d'}x^{\frac{1}{\eta}}}\right) &\geq \left(\frac{\nu+\gamma^{d}x'^{\frac{1}{\eta}}}{\nu+\gamma^d x^{\frac{1}{\eta}}}\right)\end{aligned}$$

that holds as long as $d' \geq d$, since

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

for any $x' > x$. This concludes the proof. □

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. This result holds immediately by combining Proposition 3 and 2. □

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

$$Y(t,\bar{t}) = \gamma^{\bar{t}}\tilde{Y}(\bar{t}-t)$$

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

Proof.

We can use the previous Proofs of Lemma 5, 6 and 7 to construct $G(x', x; t, \bar{t}) \equiv g(\alpha(x', x; t, \bar{t}), x', x)$, that is given by

$$G(x', x; t, \bar{t}) = \gamma^{\bar{t}} \underbrace{\begin{cases} \gamma^{-(\bar{t}-t)}\left[\nu f(x',x) + \frac{\eta}{\eta+1}\kappa_1 f(x',x)^{\frac{\eta+1}{\eta}}\right] & \text{if } f(x',x) \leq \chi_0 \\ \gamma^{-(\bar{t}-t)}\left[\nu f(x',x) + \frac{\eta\varepsilon}{\eta\varepsilon+1}\kappa_2 f(x',x)^{\frac{\eta\varepsilon+1}{\eta\varepsilon}}\right] & \text{if } f(x',x) \in (\chi_0, \chi_1(t,\bar{t})) \\ \gamma^{-(\bar{t}-t)}\nu f(x',x) + \frac{\eta}{\eta+1} f(x',x)^{\frac{1}{\eta}} - \kappa_2^{-(\eta\varepsilon-1)}\frac{\gamma^{(\varepsilon\eta-1)(\bar{t}-t)}}{\varepsilon(1+\eta)} & \text{if } f(x',x) \geq \chi_1(t,\bar{t}) \end{cases}}_{\equiv \tilde{G}(x',x;\bar{t}-t)}.$$

where $\chi_1(t,\bar{t}) = \chi_0\gamma^{\frac{\eta(\eta\varepsilon-1)}{\eta+1}(\bar{t}-t)}$. The GDP can thus be rewritten as

$$Y(t,\bar{t}) = \int_0^1 G(m(x), x; t, \bar{t})\,\omega(x)\,dz = \gamma^{\bar{t}}\int_0^1 \tilde{G}(m(x), x; \bar{t}-t)\,\omega(x)\,dz$$

where the first equality holds by the definition of G and the second one by the definition of $\tilde{G}$. Last, we need to show that $\tilde{G}(x', x, \bar{t}-t)$ is decreasing in $(\bar{t}-t)$ $\forall (x', x)$. Consider any $\bar{t}'-t' \geq \bar{t}-t$. As shown, $\chi_1(t', \bar{t}') \geq \chi_1(t, \bar{t})$. First, take (x', x) such that $f(x', x) \leq \chi_0$ or $f(x', x) \geq \chi_1(t, \bar{t})$. Here it is immediate to see, since $\eta\varepsilon > 1$, that $\tilde{G}(x', x; \bar{t}-t) \geq \tilde{G}(x', x; \bar{t}'-t')$. Next, consider instead (x', x) such that $f(x', x) \in (\chi_1(t, \bar{t}), \chi_1(t', \bar{t}'))$. In this case, first notice that - since $f(x', x) > \chi_1(t, \bar{t})$

$$\gamma^{-(\bar{t}'-t')}\nu f(x',x) + \frac{\eta}{\eta+1} f(x',x)^{\frac{1}{\eta}} - \kappa_2^{-(\eta\varepsilon-1)}\frac{\gamma^{(\varepsilon\eta-1)(\bar{t}'-t')}}{\varepsilon(1+\eta)} > \gamma^{-(\bar{t}'-t')}\left[\nu f(x',x) + \frac{\eta\varepsilon}{\eta\varepsilon+1}\kappa_2 f(x',x)^{\frac{\eta\varepsilon+1}{\eta\varepsilon}}\right],$$

then notice that $\gamma^{-(\bar{t}'-t')}\left[\nu f(x',x) + \frac{\eta\varepsilon}{\eta\varepsilon+1}\kappa_2 f(x',x)^{\frac{\eta\varepsilon+1}{\eta\varepsilon}}\right] > \gamma^{-(\bar{t}-t)}\left[\nu f(x',x) + \frac{\eta\varepsilon}{\eta\varepsilon+1}\kappa_2 f(x',x)^{\frac{\eta\varepsilon+1}{\eta\varepsilon}}\right]$. This completes the proof. □

B.3 Characterization and Proofs for the "Tractable Case"

[TO BE COMPLETED]

C Introducing Self-Employment

I extend the tractable version of the model - the one of Section 3.2 - and introduce self-employment. This extension provides a new empirical prediction: in countries further from the technology frontier self-employed are more negatively selected. Additionally, under ceratin parametrization, the model generates a larger number of self-employed far from the frontier. In this sense, the model provides an equilibrium explanation for the empirically observed large number of low-skilled self-employed in developing countries. In the model, these individuals become self-employed because those types that would be their managers in a developed country choose to be workers themselves to get access to the advanced technology. This type of self-employment resembles the notion of subsistence entrepreneurship in developing countries, as emphasized by Schoar (2010), Ardagna and Lusardi (2008) and Banerjee and Duflo (2011).

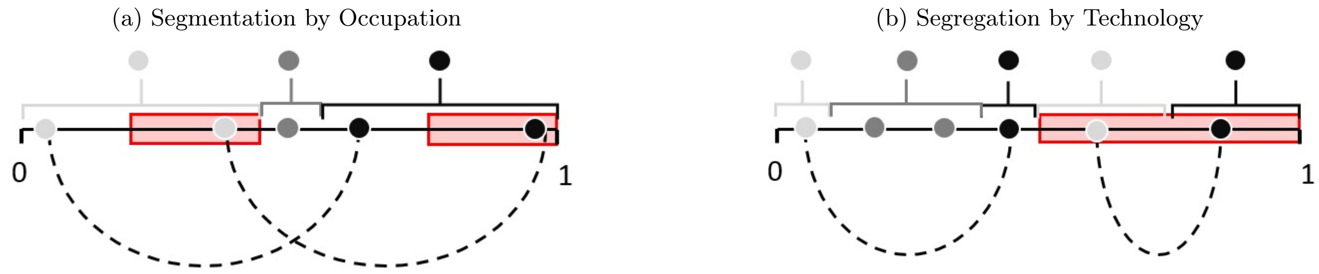
Specifically, I model being self-employed, relative to being a manager, along the lines of Garicano and Rossi-Hansberg (2004). Under this setup, self-employment is a more skill intense occupation than being a worker, but less so than being a manager. The production function is given by

$$\begin{aligned} g\left(a, x', x\right) &= ax'\left(2-\lambda+\lambda x\right) - c\left(a; t, \bar{t}\right)\left(2-\lambda+\lambda x\right) \\ g^s\left(a, x\right) &= ax' - c\left(a; t, \bar{t}\right) \end{aligned}$$

where $s(a, x)$ is the output of a self-employed individual, and the cost function is as defined in 3.2. I assume that that $\lambda < \frac{1}{2}$ and $\chi_0 \in \left(\frac{1}{2} + \frac{\lambda+\sqrt{\lambda^2+\lambda(2-\lambda)}}{4-2\lambda}, 1\right)$. These parametric assumptions guarantee that: (i) a Spence-Mirrlees conditions holds, that is $g_2(a, x, y) \geq g_2^s(a, x) \geq g_3(a, z, x)\ \forall (x, y, z)$, so that - conditional on technology used - types are separated into occupations; (ii) in equilibrium not everyone is self-employed, but at least someone is. The definition of the competitive equilibrium mirrors the one for the case without self-employment, with the only addition that an individual is allowed to choose to become self-employed and earn $s(x) = \max_a g^s(a, x)$. Labor market clearing need to take into accont the possibility of being self-employed, and thus is given by $\omega(x) = 1 - \sigma(m(x)) - \omega(m(x))\ \forall x$, where $\sigma(x)$ is the new equilibrium object that defines the probability that a worker of ability x is self-employed.

Characterization. As a result of the functional form assumptions, in equilibrium, among individuals who use the same technology there is going to be talent segmentation by occupation: self-employed are more skilled than workers, but less so than managers. If everyone uses a similar technology, as when the distance from the frontier $d = \bar{t} - t$ is small, the self-employed are thus going to be as skilled as the average. In Figure A.6a, I depict this allocation. Instead, if the gap between the frontier and local technology, d is large, then the complementarity between labor and technology implies that a worker using the frontier technology has a higher marginal value of his talent than a self-employed using the local technology. Additionally, self-employed must be less skilled than managers, due to the fact that they have lower marginal value of skills. Since - due to the functional form assumptions - at least some manager chooses the local technology, this implies that no self-employed individual uses the frontier technology. This second case, which displays talent segregation by technology, is shown in Figure A.6b. Self-employed are now less skilled than the average.

Figure A.6: Allocation of Talent with Self-Employment



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

I have also completely characterized intermediate cases, however I omit for brevity such characterization, that follows similar steps as the ones shown in Appendix B.3. It is available upon requests. Nonetheless I here present the main theoretical results that links the distance of a country from the frontier and the ability of self-employed relative to managers and workers.

Proposition C.1 (Selection of Self-Employed). *The selection of self-employed, defined as* $\vartheta(d) = \int x\sigma(x)\,dx - \int x(1-\sigma(x))\,dx$*, decreases in* d.

Proof.

I first define the marginal values of individuals skills as a manager, a worker, and a self-employed, taking into accout optimal technology choice

$$\pi'(x) = \gamma^t \begin{cases} (1+\gamma^d)\left(2-\gamma+\gamma m^{-1}(x)\right) & \text{if } x \geq \chi_0 \\ \left(2-\gamma+\gamma m^{-1}(x)\right) & \text{if } x < \chi_0 \end{cases}$$

$$w'(x) = \gamma^t \begin{cases} \lambda\left((1+\gamma^d)(m(x)-\chi_0)+\chi_0\right) & \text{if } m(x) \geq \chi_0 \\ \lambda m(x) & \text{if } m(x) < \chi_0 \end{cases}$$

$$s'(x) = \gamma^t \begin{cases} (1+\gamma^d) & \text{if } x \geq \chi_0 \\ 1 & \text{if } x < \chi_0 \end{cases}$$

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

1. $\forall x$ s.t. $\omega(x) > 0$ and $\mu(x) > 0$ then $\pi'(x) = w'(x)$;

2. $\forall x$ s.t. $\omega(x) > 0$ and $\sigma(x) > 0$ then $s'(x) = w'(x)$;

3. $\forall x$ s.t. $\mu(x) > 0$ and $\sigma(x) > 0$ then $s'(x) = \pi'(x)$;

4. $\forall x$ s.t. $\omega(x-\epsilon) = 0$; $\mu(x-\epsilon) > 0$; and $\omega(x) > 0$ then $\pi'(x) \leq w'(x)$;

5. $\forall x$ s.t. $\omega(x-\epsilon) = 0$; $\sigma(x-\epsilon) > 0$; and $\omega(x) > 0$ then $s'(x) \leq w'(x)$;

6. $\forall x$ s.t. $\mu(x-\epsilon) = 0$; $\omega(x-\epsilon) > 0$; and $\mu(x) > 0$ then $w'(x) \leq \pi'(x)$;

7. $\forall x$ s.t. $\mu(x-\epsilon) = 0$; $\sigma(x-\epsilon) > 0$; and $\mu(x) > 0$ then $s'(x) \leq \pi'(x)$;

8. $\forall x$ s.t. $\sigma(x-\epsilon) = 0$; $\omega(x-\epsilon) > 0$; and $\sigma(x) > 0$ then $\pi'(x) \leq s'(x)$;

9. $\forall x$ s.t. $\sigma(x-\epsilon) = 0$; $\mu(x-\epsilon) > 0$; and $\sigma(x) > 0$ then $w'(x) \leq s'(x)$;

I omit the proof of the necessary conditions. It follows closely the one for the case without self-employment, and simply requires to show an optimal deviation for each one of the 9 conditions.

Using condition (9) and the fact that, given the chosen functional form, $s'(x) < \pi'(x)$ $\forall x$, we get that in the optimal allocation the most skilled self-employed must be less skilled than the least skilled manager. As a result, since we work with parameter values (high enough χ_0) such that at least some managers are less skilled than χ_0 , then it must be that the highest skilled self-employed is less skilled than χ_0, and therefore no self-employed uses the advanced technology. As a result, all self-employed have identical marginal value of skills, equal to $s'(x) = 1$. This in turn, using again the necessary conditions, implies that the set of self-employed is connected, and that $\sigma(x) \in \{0, 1\}$ almost everywhere. Then, in order to calculate the expected ability of self-employed is sufficient to calculate the difference between the mass of individuals which are more skilled than self-employed and those that are less skilled than them. In fact, let $\sigma_{max} = \max_x \{x : \ \sigma(x) = 1\}$ and $\sigma_{min} = \min_x \{x : \ \sigma(x) = 1\}$, then the average ability of self-employed is given by $\vartheta(d) = \frac{\sigma_{max}+\sigma_{min}}{2}$. Where the set $[\sigma_{max}, 1]$ is the set of individuals more skilled than self-employed and $[0, \sigma_{min}]$ the set of types less skilled. They are respectively of masses (defined h and l) $h = (1-\sigma_{max})$ and $l = \sigma_{min}$. We can then rewrite $\vartheta(d) = \frac{1-(h-l)}{2}$, that shows that $\vartheta(d)$ is decreasing in $(h-l)$. Next, let's notice that since $w'(x)$ is increasing in x, while $s'(x)$ is constant, it must be that

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

and then by market clearing we get

$$
h - l = 2 \int_{x:\,w'(x)=s'(x)}^1 \omega(x)\,dx.
$$

Last, we need to show that $2\int_{x:\,w'(x)=s'(x)}^{1}\omega(x)\,dx$ increases in d. Notice that this is the set of workers that is matched with managers more skilled than χ_0 and such that $w'(x)\geq s'(x)$. This last inequality can be rewritten, using the functional forms shown above, as

$$m(x)\geq\chi_0+\frac{1-\lambda\chi_0}{\lambda(1+\gamma^d)},$$

we know that the set of manager more skilled than χ_0 is connected as well (again using the first order conditions), as a result

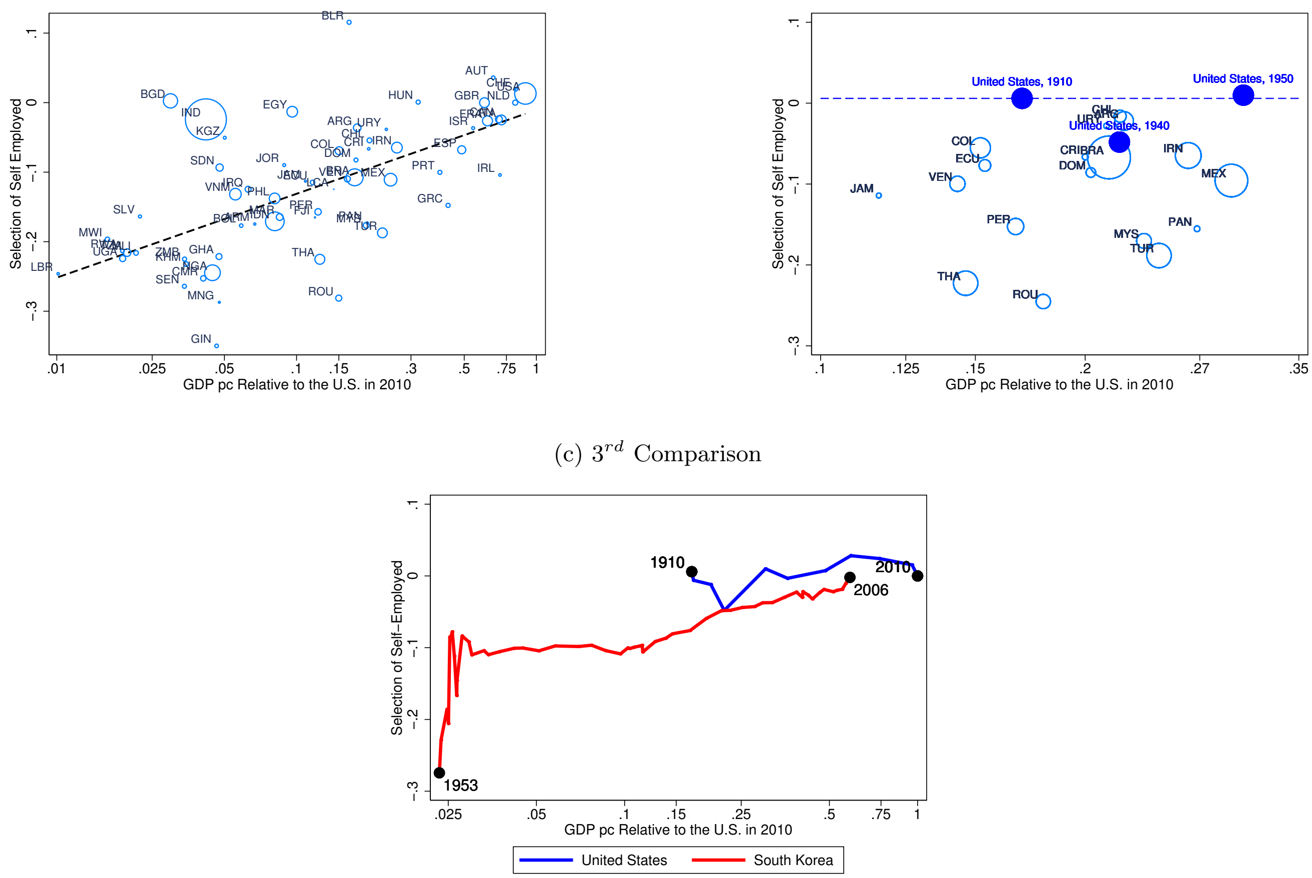
$$\int\limits_{x:\,w'(x)=s'(x)}^{1}\omega(x)\,dx=\left[\min\left\{1,\chi_0+\frac{1-\lambda\chi_0}{\lambda(1+\gamma^d)}\right\},1\right]$$

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

Figure A.7: Selection of Self-Employed

(a) 1^{st} Comparison

(b) 2^{nd} Comparison



(c) 3^{rd} Comparison

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

Empirical Evidence. I next bring empirical evidence to support the prediction of Proposition C.1. I use the same data source described in Section 4. The IPUMS data report whether an individual is self-employed or a wage/salary worker. In order to build the empirical measure of self-employment, I use my measure of skill, namely the normalized number of years of education, and compute the difference between the average skill of self-employed and the average skill of wage-workers. Further details are in the Section I.

I follow the same three empirical comparisons as for the measure of concentration of talent. Results are shown in Figure A.7. Poor countries - farther from the frontier - have more negatively selected self-employed, both than rich ones today and than United States when the latter were at similar level of development.[63,64] Last, as South Korea approached the frontier, the selection of its self-employed decreased.[65]

Number of Self-Employed It may also be interesting to ask what are the determinants of the number of individuals that decides to become self-employment, and how it changes as a function of distance from the frontier. Let's consider two individuals x and x', with $x' > x$. The two of them generates a higher added value as a team rather than self-employed if and only if

$$g\left(\alpha\left(x'\right), x', x\right) > g_s\left(\alpha\left(x'\right), x'\right) + g_s\left(\alpha\left(x\right), x\right)$$

where I've used the fact that in the tractable setting the optimal technology depends only on the type of the manager, thus $\alpha\left(x', x\right) = \alpha\left(x'\right)$. From the expression above we can notice that forming a pair allows to leverage the technology of the most skilled individuals, since both workers then use $\alpha\left(x'\right)$. However there is also a cost to form a pair, since - due to the chosen functional forms - $g\left(a, x, x\right) < g_s\left(a, x\right) + g_s\left(a, x\right)$. The result is that it is never optimal to form a pair with two identical individuals. They would rather both work as self-employed. Intuitively, a firm is optimal only to the extent that the cost in terms of transaction costs within the firms is compensated by the fact that the manager can leverage his higher talent. With this trade-off in mind we can explore the role of distance to the frontier. When a country is farther from the frontier, there is more scope to leverage the talent, since the gap between the frontier and the traditional technology is higher, this force alone tends to reduce the number of self-employed. However, if we focus within the traditional sector, then we are left with individuals which use the same technology, and which are more similar between each other, since all the high skilled ones are allocated to the modern sector. The compression of the skill distribution in the traditional sector increases the number of self-employed. The overall effect depends on parameter values, but the model can replicate, as in the

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

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

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

example shown in figures A.6a and A.6b, the fact that in countries far from the frontier we observe a very large number of low skilled self-employed.

D A More General g and the Role of Skill-Technology Complementarity

I here consider a more general production function for $g(a, x', x)$ to illustrate the role of the complementarity between technology and labor input. Recall that in the main text I used $g(a, x', x) = af(x', x) - c(a)$. I now consider instead $g(a, x', x) = \tilde{f}(a, x', x) - c(a)$. As a result, the marginal values of skills of x as a manager and worker becomes

$$\begin{aligned} \pi'(x) &= \tilde{f}_2\left(\alpha\left(x, m^{-1}(x)\right), x, m^{-1}(x)\right) \\ w'(x) &= \tilde{f}_3\left(\alpha\left(m(x), x\right), m(x), x\right), \end{aligned}$$

and - using a first order Taylor expansion with respect to a around $\tilde{f}_3\left(\alpha\left(x, m^{-1}(x)\right), m(x), x\right)$ - we get that $\pi'(x) > w'(x)$ if an only if

$$\begin{aligned} &\underbrace{\tilde{f}_2\left(\alpha\left(x, m^{-1}(x)\right), x, m^{-1}(x)\right) - \tilde{f}_3\left(\alpha\left(x, m^{-1}(x)\right), m(x), x\right)}_{\text{Skill-Sensitivity Gap}} \\ > \; &\underbrace{\tilde{f}_{13}\left(\alpha\left(x, m^{-1}(x)\right), m(x), x\right)}_{\text{Skill-Technology Complementarity}} \underbrace{\left(\alpha\left(m(x), x\right) - \alpha\left(x, m^{-1}(x)\right)\right)}_{\text{Technology Gap}} \end{aligned}$$

This equation illustrates the role of skill-technology complementarity. If $\tilde{g}_{13} = 0$, then $\pi'(x) > w'(x)$ $\forall x$. In fact, the trade-off occupation-technology exists because there is strictly positive complementarity between the used technology and the ability of workers. The stronger is this complementarity, the smaller the technology gap has to be to make $w'(x) > \pi'(x)$.

E Discussion of Endogenous Team Size

I discuss how the model can be extended to allow for the possibility that a manager chooses the size of the firm - i.e. how many workers to hire. I show that under a stricter notion of complementarity the main results of the paper are likely to hold for this case as well.

Let's assume the following production function

$$g\left(a, x', x, l\right) = \tilde{f}\left(a, x', x\right) l^{\gamma} - c(a),$$

where l is the mass of workers of type x. The main difference with the previous setting is that now the manager can choose l - while previously it was fixed and equal to 1. The manager problem thus reads as

$$\pi(x) = \max_{z,l} \tilde{f}\left(\alpha(x, z, l), x, z\right) l^{\gamma} - w(z) l - c\left(\alpha\left(x', x, l\right)\right)$$

where - using the usual notation - $\alpha(x', x, l) = \arg\max_a g(a, x', x, l)$. The first order conditions with respect to z and l give

$$\begin{aligned}\tilde{f}_3\left(\alpha\left(m\left(x\right),x,l\right),m\left(x\right),x\right)l^*\left(m\left(x\right)\right)^{\gamma} &= w'\left(x\right)l^*\left(m\left(x\right)\right)\\ \gamma\tilde{f}\left(\alpha\left(m\left(x\right),x,l\right),m\left(x\right),x\right)l^*\left(m\left(x\right)\right)^{\gamma-1} &= w\left(x\right)\end{aligned}$$

where $l^*\left(m\left(x\right)\right)=\arg\max g\left(a,m\left(x\right),x,l\right)$. Combining the two first order conditions together gives

$$\frac{w'\left(x\right)x}{w\left(x\right)} = \frac{1}{\gamma}\frac{\tilde{f}_3\left(\alpha\left(m\left(x\right),x,l\right),m\left(x\right),x\right)x}{\tilde{f}\left(\alpha\left(m\left(x\right),x,l\right),m\left(x\right),x\right)}.$$

This last equation highlights the implications of allowing firm size to be endogenous. Now, a manager can substitute high skilled workers for more low skilled and thus cheaper workers. Hence depending - once again - on the strength of the technology-skill complementarity we have different pattern of matching. In particular it is easier to see that under the functional form assumption used in the paper, hence with $\tilde{f}\left(a,x',x\right)=af\left(x',x\right)$, we would have that

$$\frac{w'\left(x\right)x}{w\left(x\right)}=\frac{f_3\left(m\left(x\right),x\right)x}{f\left(m\left(x\right),x\right)}$$

that shows that the marginal value of an individual does not depend on the technology used. Hence, in order for the results of the paper to hold, we need a stronger notion of technology-skill complementarity. In particular, we need $\tilde{f}$ to display log-supermodularity, that is $\frac{\partial}{\partial a}\log\tilde{f}_3>0$. It is then immediate to see that under the assumption of log-supermodularity $w'\left(x\right)$ is increasing in $\alpha\left(m\left(x\right),x,l\right)$. As a result, the technology-occupation trade-off illustrated in the main text hold as long as $\tilde{f}$ is log-supermodular. It is known in the literature that when firm size is endogenous supermodularity is not sufficient to generate positive assortative matching between managers and workers, and log-supermodularity is required instead. (See for example Grossman et al. (2013) and Eeckhout and Kircher (2012)). That same insight holds in my setting as well.

F Firm Level Evidence

In this section, I provide further evidence in support of the theory using firm level data. Firm level data allow to directly observe who works with whom, and thus to get closer to the model's notion of a team. In the model, a team is characterized by two key features: (i) it is made of individuals that interact with each other, to the extent that there is complementarity in production; and (ii) the members of a team use identical technology. A firm is the empirical object that better represents these two features. Individuals within a firm mostly share the same production technology, task-specialize based on their ability, and interact with each other in a manner that may generate complementarities.[66]

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

F.1 Data (Firm Level)

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

As a measure of technology, in the WBES 2014 I use labor productivity, defined as sales divided by total number of employed workers. This is consistent with the prediction of the model that labor productivity and technology are positively correlated. In the WBES 2006 instead, I can use instead three different measures: (i) labor productivity; (ii) fraction of labor force that uses computers; (iii) self-assessed level of technology relative to main competitors.

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

The sampling methodology targets formal (registered) companies with 5 or more employees. These data therefore are representative only of a selected fraction of the population, which is smaller for poor countries, where a large part of the population is employed in the informal sector. For this reason we should be cautious in interpreting the results. In Section F.3, I farther discuss this concern. Additional details on the data are included in Section I.

F.2 Empirical Results

I use the data to provide support to three empirical predictions of the model. Results are summarized in Figures A.8, A.9, and A.10.

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

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

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

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

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

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

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

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

In the wave WBES 2006, I observe the distribution of individual education years for each firm. I use this to compute the overall variance of education in each country and decompose it between the within-firm and across-firm variance. The fraction of overall variance that is explained across firms decreases significantly in GDP per capita, that is, poor countries indeed have a larger concentration of talent across firms.[67]

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

F.3 Robustness and Alternative Interpretations

The data cover only registered firms with more than five employees. We thus miss a large fraction of the population in poor countries, where average firm size is smaller and informality is widespread. We should thus interpret the evidence presented as suggestive. It can nonetheless be useful to discuss and

[67] I compute a similar variance decomposition using the normalized measure of education between 0 and 1, as in Section 4. Results are similar - although slightly less significant - and are available upon request. I decided not to use the normalized version for the benchmark result because in the WBES we do not observe the whole population, but only a selected subsample of it.

address the most obvious concerns. Results for this robustness check are shown in Figure A.11.

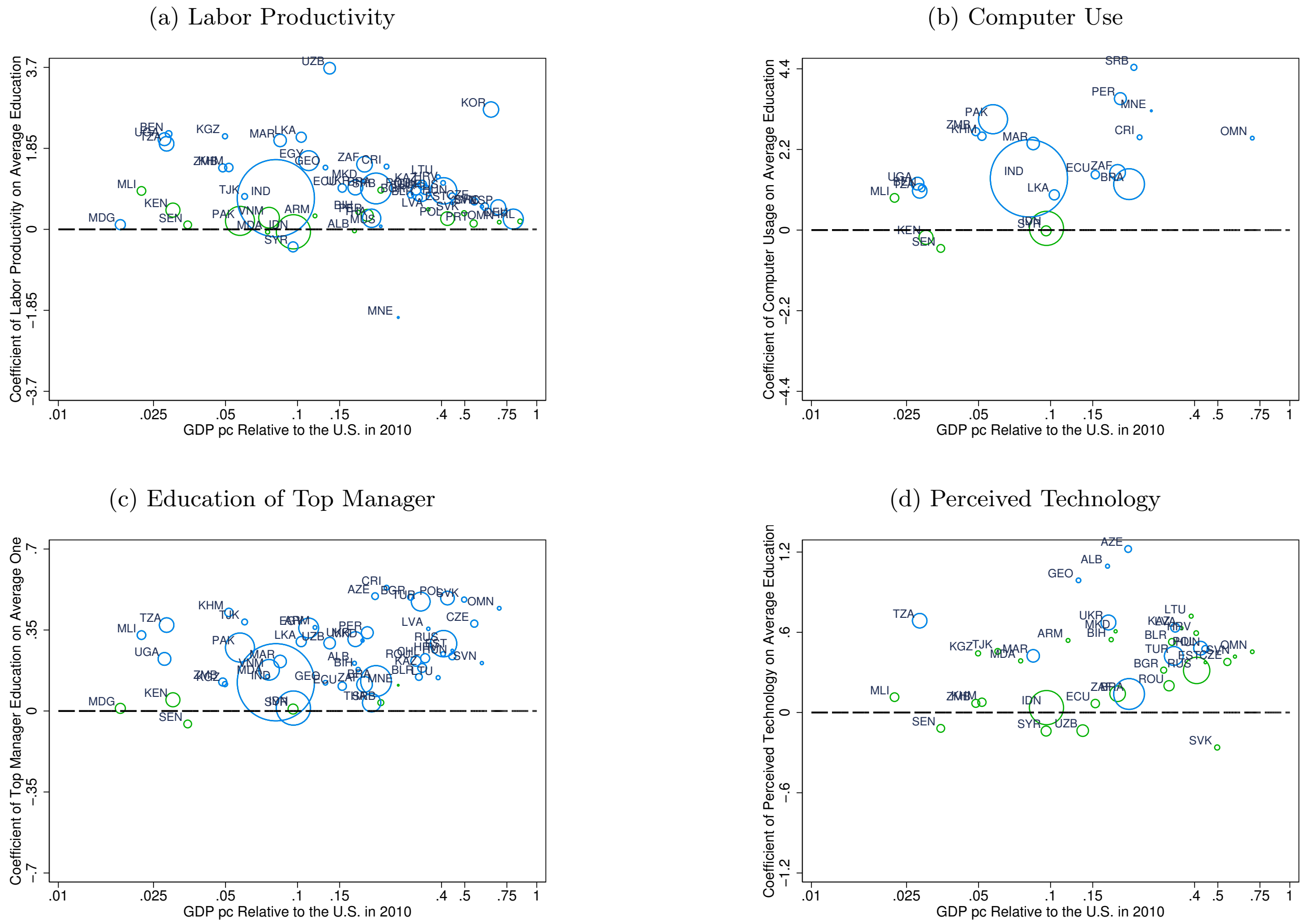
The empirical results highlight that there is more dispersion of economic activity in poor countries. One concern would be if the set of firms included in the sample are, in poor countries, drawn more frequently from the tails of the overall distribution of firms. I show that in both rich and poor countries firms have similar size, with 100 employees on average. In developing countries, these firms are likely to be overly representative of the high productivity portion of the economy. Consistent with this observation, the workforce there employed is more educated than the country average, and especially so in the poorest countries. For this reason, we are likely observing a right truncation of the firm distribution in poor countries. This fact is suggestive that the results might even understate the amount of dispersion in poor countries, since we are not capturing the possibly large gap between firms included and those not included in the sample; that is between the formal and informal side of the economy.[68]

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

Last, we might be concerned that the notion of technology in the data captures a different concept than the one in the model. The key assumption in the model is that technology can be improved subject to a cost. For example, this assumption might be violated if some firms have higher technology and labor productivity simply because their owners are more able. Firm owner ability is, in fact, an (approximately) fixed characteristics that arguably should respond little to the quality of the labor force. One question in the WBES 2006 helps to address this concern: "Over the last two years, what were the leading ways in which your establishment acquired technological innovations?". "Embodied in new machinery or equipment" is one of 13 possible choices. Across both rich and poor countries, approximately 60% of firms ranked this as their first choice.

[68] However, beware that whether a right truncation would over or understate the observed dispersion in the data ultimately depends on the shape of the firm distribution.

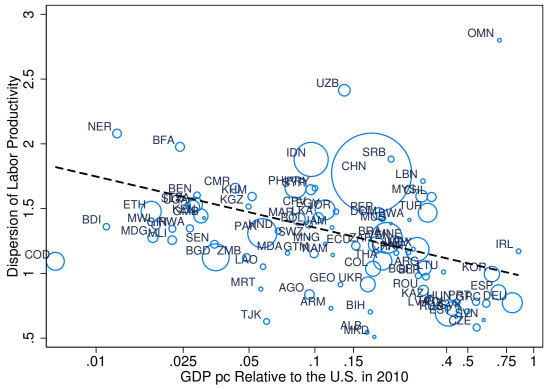
Figure A.8: Complementarity Skill-Technology



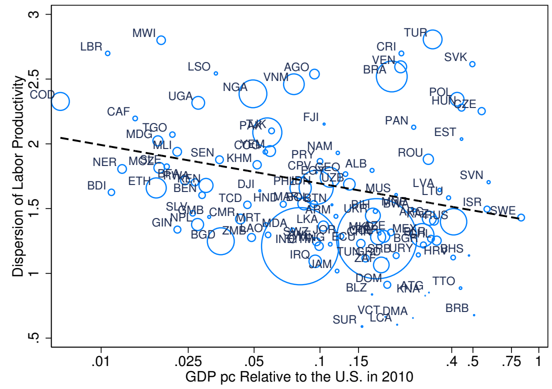
Notes: In each figure I plot the coefficients of a linear regression, computed across firms within a country, on the log of average education of the firm workforce. For example, let me describe the procedure to build the top left panel, since all other ones are identical. I first compute, for each country,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

Figure A.9: Dispersion of Used Technology Across Firms

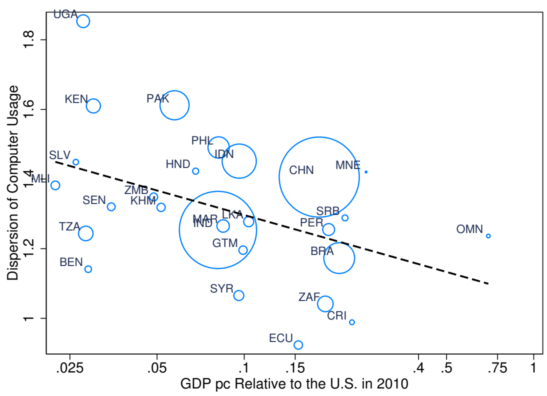
(a) Labor Productivity, WBES 2006



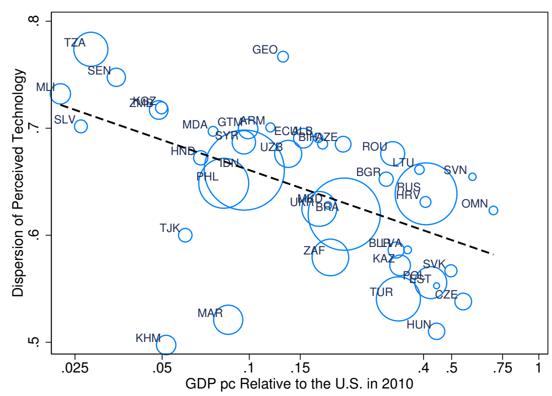
(b) Labor Productivity, WBES 2014



(c) Computer Use



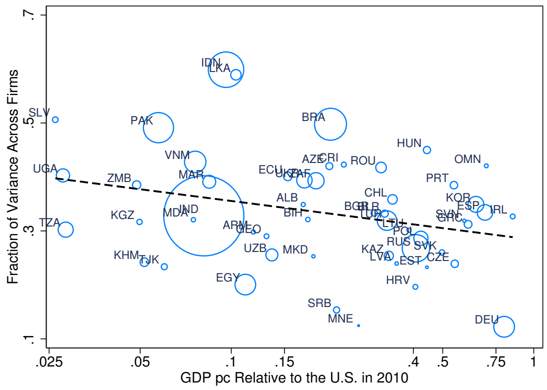
(d) Perceived Technology



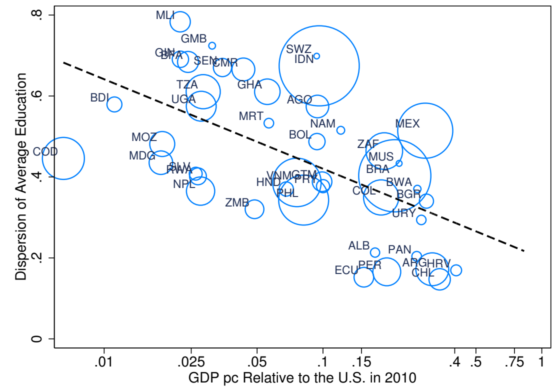
Notes: In each figure I plot standard deviation of log of a measure of the measure of education or technology shown in the figure title. Specifically, for each country I compute the cross-sectional, across firms, standard deviation of logs and then I plot the country estimates as a function of GDP per 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.10: Dispersion of Average Education Across Firms

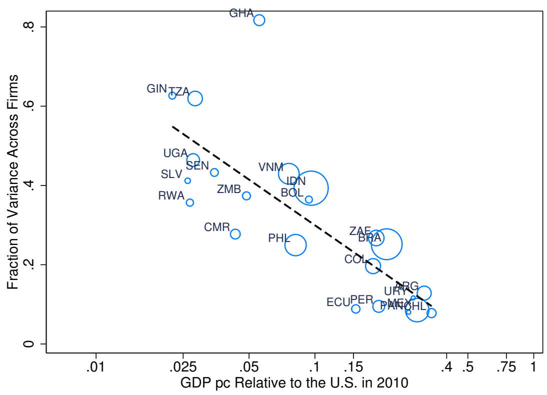
(a) Fraction of Total Variance Explained Across Firms, WBES 2006



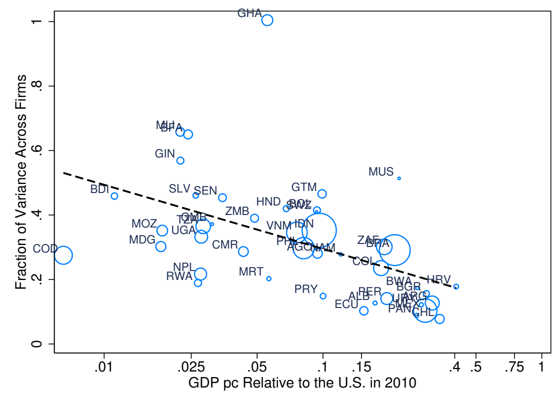
(b) Dispersion of Education, WBES 2014



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

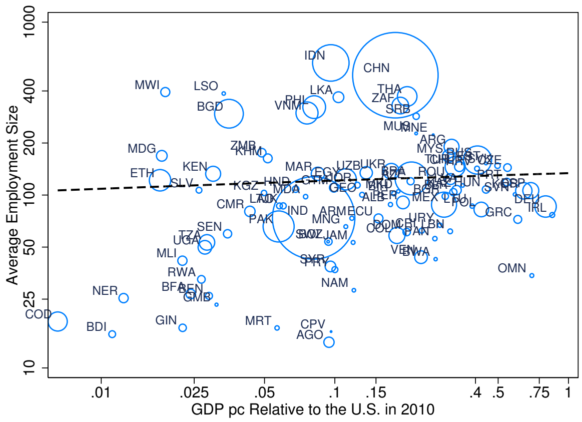


(d) Fraction of Total Variance Explained Across Firms, WBES 2014 (b)

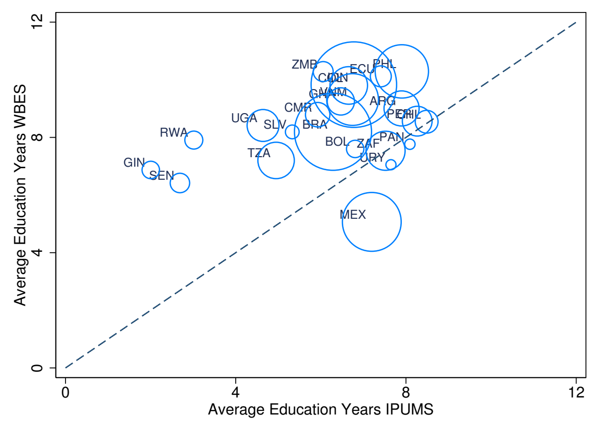


Notes: In the top left panel, I plot for each country as a function of GDP per capita the fraction of total variance of education which is "across firms". Specifically, for each country I compute the cross-sectional variance of education of the individuals hired at the firms in the sample. I can compute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

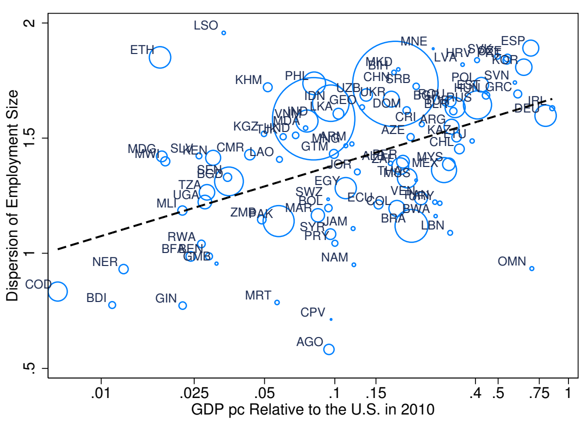
Figure A.11: Robustness for Firms Level Data



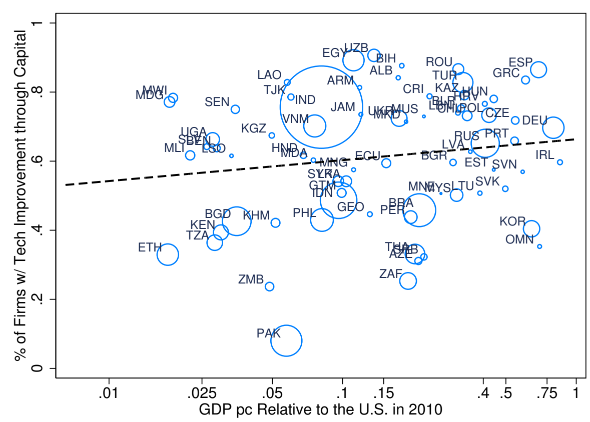
(a) Average Size



(b) Comparison Education in WBES and IPUMS



(c) Dispersion of Size



(d) Technology Embodied in Capital

Notes: In the top left panel I plot the average firm number of employees for each country as a function of the GDP per capita. In the the top right panel I plot the average education of workers in the firms in my data as a function of the average country education taken from IPUMS international micro data as described in Section **??**. 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.

G Computing Algorithm

H Identification

H.1 Average Ability in Agriculture and Non-Agriculture

I Additional Details on the Data

I.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.

Last, few remarks on sample size might be useful. In order to have similarly sized dataset for each country, I limit the sample size in each country-year pair to 500,000 individuals. As discussed, in the main analysis I use males, household head, who are between 18 and 60 years old, and who report their industry and education. This restriciton gives me on average 80,000 observations for each country-year pair.[69] For few countries, I have smaller datasets. But for each one I have at least 10,000 usable observations. On average, I use slightly fewer observations for poor countries than for rich ones. This is due to the fact the age-pyramids are more skewed in poor countries, hence I drop more individuals when I restrict to those between 18 and 60 years old. However, these cross-country differences are small and unlikely to play a major role. In fact, even for the poorest set of countries I have on average 70,000 individuals.

I.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.10a, 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.

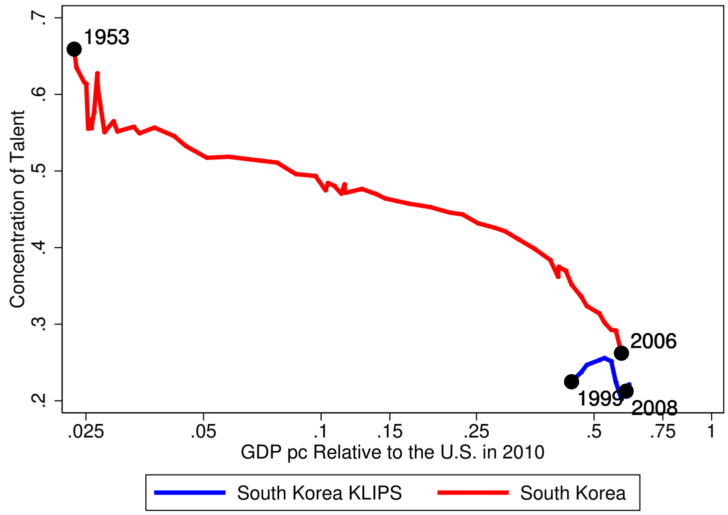
[69] Why only 80,000 on average from the initial 500,000? I consider United States as an example. Of the 500,000 individuals, 243,000 are males. 138,000 have between 18 and 60 years old, and then 72,000 are household heads. Of these 72,000 only 3,000 have missing industry data (no one has missing education data), likely because they are unemployed at the time of the census.

J 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.12 and A.14. 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.13 and shows that the patterns are similar.

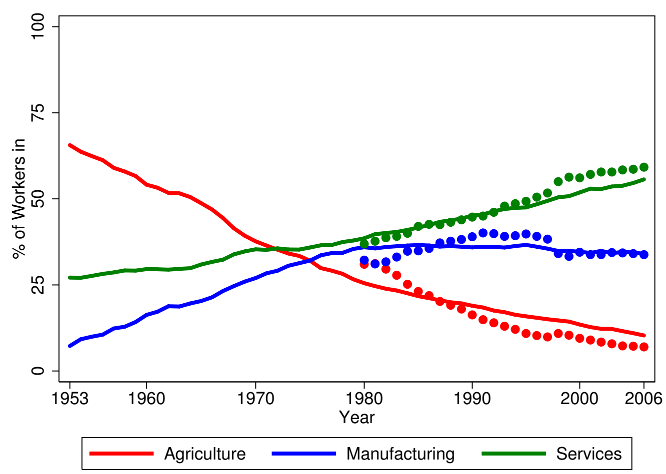
Last, in A.15 I report the disaggregated data from which the measure of concentration of talent is computed from 1960 to 2005. The process of structural transformation is evident in the figure, but we can notice that, throughout the growth miracle, the linear measure of concentration of talent consistently provides a reasonably good fit.

Figure A.12: Comparison between KLOSA and KLIPS dataset



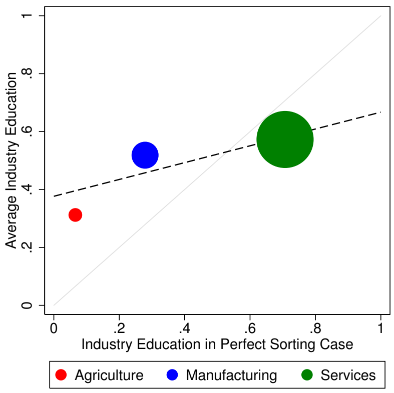
Notes: in the Figure I plot the growth path of concentration of talent across sectors computed from KLOSA data, as shown in Figure 9, and I compare it with the value of concentration of talent computed from the KLIPS data for the period 1999 to 2008.

Figure A.13: Structural Transformation of South Korea

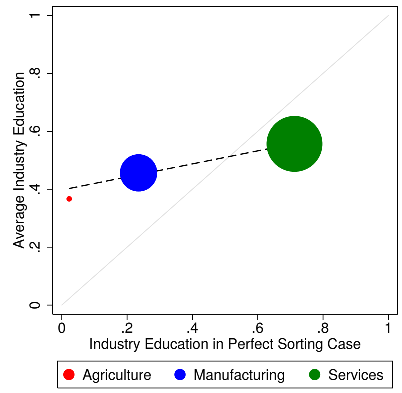


Notes: I plot the fraction of male population employed in either sector. Dots are values from (constructed) cross sections in KLOSA. The solid line are instead aggregate data for the male population from the World Bank Development Indicators.

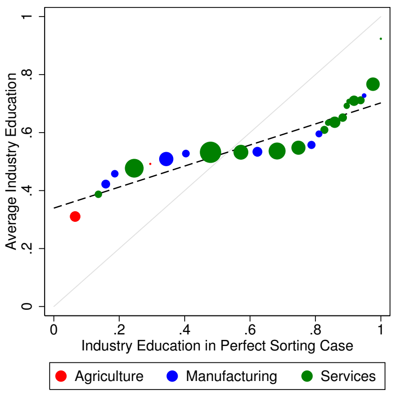
Figure A.14: Concentration of Talent, Comparison between KLOSA and KLIPS dataset in 2005



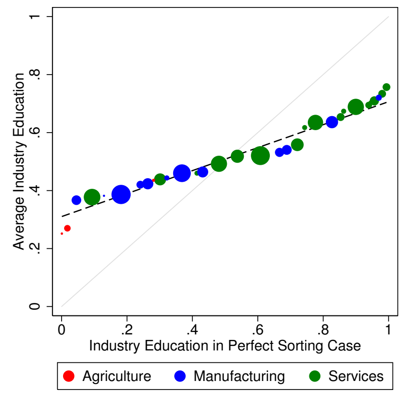
(a) KLOSA, Across Sectors



(b) KLIPS, Across Sectors



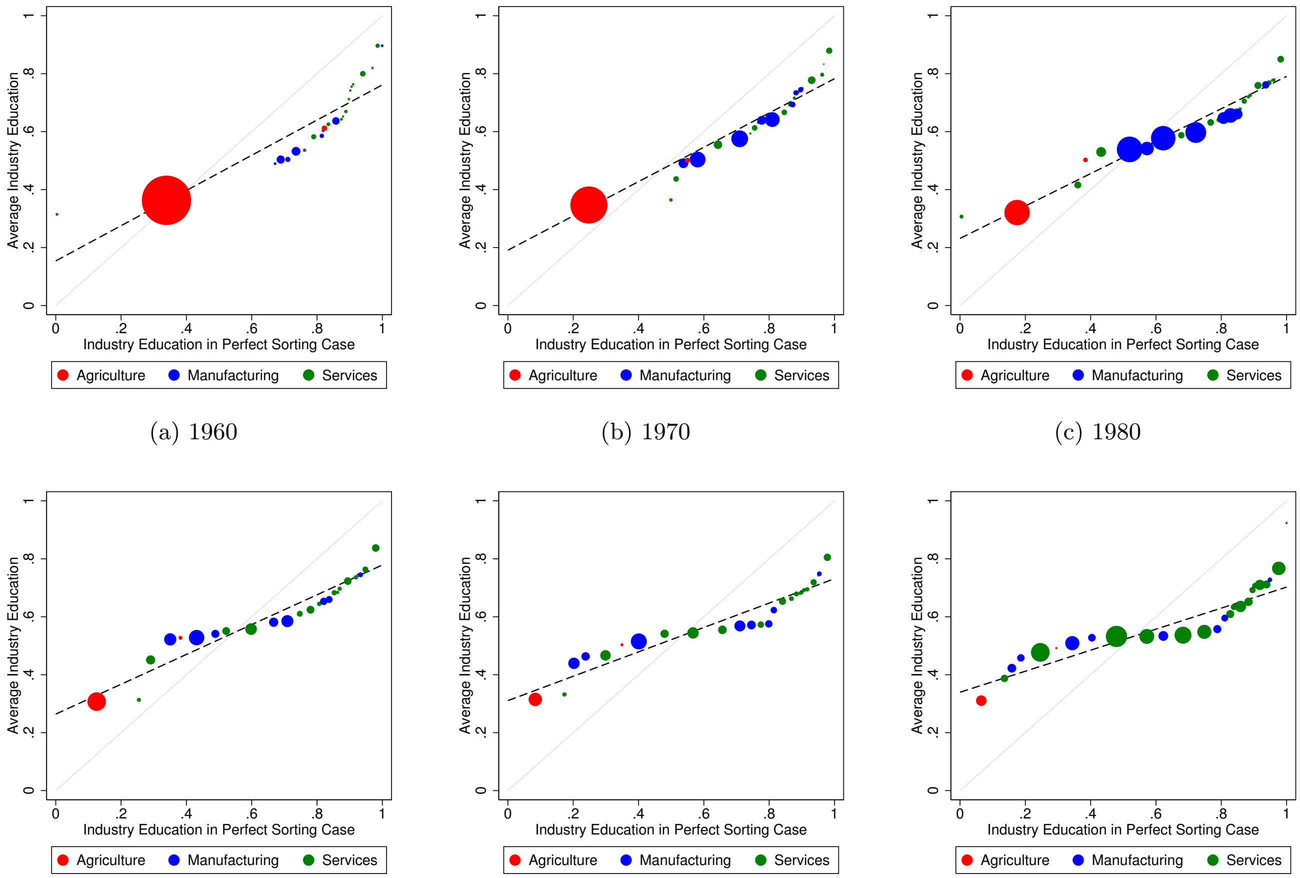
(c) KLOSA, Across Industries



(d) KLIPS, Across Industries

Notes: this figure replicates figures A.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.15: Concentration of Talent Across Industries, South Korea Growth Path



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

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

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