CATCH ME IF YOU LEARN: DEVELOPMENT-SPECIFIC EDUCATION AND ECONOMIC GROWTH

FABIO CERINA
*University of Cagliari
and
CRENoS*

FABIO MANCA
CRENoS

This paper presents a theoretical and empirical investigation of the relationship between human capital composition and economic growth. In the theoretical analysis, we allow for nonconstant returns to scale in technological activities. Differently from previous literature, our results show that, under broad and plausible model parameterizations, the marginal growth effect of skilled workers is increasing with the distance to the frontier for sufficiently poor countries while it is decreasing (in agreement with the existing literature) only for countries close to the technological frontier. Our empirical analysis provides robust evidence for this theoretical prediction by using a 10-year panel of 85 countries for the years in between 1960 and 2000, as well as by using the System Generalized Methods of Moments (GMM) technique to address the problem of endogeneity. Results are robust to different proxies of human capital and different specifications.

Keywords: Technological Frontier, Innovation, Imitation, Human Capital, Skilled, Unskilled, Growth

1. INTRODUCTION

The role played by human capital in generating economic growth has been the focus of a large strand of economic literature for decades. However, in 2001, Lant Pritchett was still wondering: "Where has all the education gone?" when

1652

referring to the weak and sometimes contradictory macroeconomic empirical evidence of a large collection of panel studies.[1] Recent contributions—most notably Vandenbussche et al. (2006) (VAM henceforth), Aghion et al. (2009), and Acemoglu et al. (2006) tried to explain this puzzling evidence by looking at the interplay between an economy's distance to the technological frontier and the composition of its human capital. Their key insight is that different kinds of human capital have each a different effect on the growth rate, depending on the economy's distance to the technology frontier.[2] In particular, an implication of these theoretical models is that skilled human capital should be especially important for the growth of countries at the technology frontier as this type of human capital is key to innovation activity.[3] VAM (by using a panel data set covering 19 developed Organisation for Economic Co-operation and Development (OECD) countries observed every 5 years between 1960 and 2000) and Aghion et al. (2009) (by using US data only) proxy skilled human capital with tertiary educated workers and provide some empirical support to this result.

According to the same models, skilled workers are less relevant for the growth of countries far from the frontier; the reason being that these countries grow out of technology adoption,[4] for which—by assumption—unskilled human capital is deemed to be enough. There is, however, robust microeconomic evidence [see Psacharopoulos (1994), Cohn and Addison (1998), Ichino and Winter-Ebmer (1999), Psacharopoulos and Patrinos (2004)] showing that both private and social returns to tertiary education in low- and middle-income countries are significantly higher than those for high-income countries. This suggests that skilled human capital might play an important role also at lower stages of development.[5] A different strand of literature seems also to support this hypothesis: Mansfield, Schwartz and Wagner (1981), Coe and Helpman (1995), or Behnabib and Spiegel (2005) argue, for instance, that the cost related to the adoption of technologies discovered at the frontier (or in other technological sectors) is positive and that investments in (skilled) human capital are hence needed in order to absorb this foreign-leading technology.[6] Finally, a recent work by Squicciarini and Voigtländer (2015), convincingly shows that the presence of knowledge elites (thick upper tail-skills) in mid-18th century France favored the adoption and efficient operation of innovative industrial technology and was key in enabling entrepreneurs in manufacturing to keep up with advances at the technology frontier.[7]

We contribute to this literature by providing a model that explains why skilled human capital can play a crucial role both for developed countries that grow mainly because of innovation, as well as for developing countries that grow mostly out of technology adoption. Crucially, from the theoretical point of view, our contribution shows that the result proposed by previous literature (for which high skills would mainly foster the growth of countries close to the technology frontier and low skills that of countries farther away from it) boils down to restricting the returns to innovation and imitation activities to be constant.

Once we relax this restrictive assumption, allowing for decreasing returns in both technological activities [following the literature[8] popularized by Jones (1995b)]

while maintaining the reasonable hypothesis for which unskilled workers are more efficient in imitation than innovation, our theoretical model leads to the emergence of a novel effect for which the marginal contribution of an additional skilled worker on the rate of growth increases as we move further away from the frontier.

It turns out that this novel effect is dominant for all the economies lagging sufficiently far from the technology frontier. When, in particular, the comparative advantage of skilled human capital in innovation is strong enough, then the marginal growth effect of an additional skilled worker *decreases* with the proximity of the technological frontier for the set of poor enough countries and then *increases*—in agreement with previous literature—for the set of countries that are sufficiently close to the technology frontier. In other words, our analysis suggests the existence of a U-shaped relationship between the marginal impact on growth of skilled human capital and the proximity to the technological frontier.

The economic intuition for this result[9] lies in the fact that when returns to innovation are decreasing, skilled human capital becomes relatively more valuable when employed in imitation activities, especially for firms operating in lagging countries. When returns to innovation are constant, firms optimally react to an increase in skilled human capital by subtracting resources to imitation and reallocating them in innovation activities. This is true even for firms operating in poor countries, where productivity growth is mainly driven by imitation, as the number of blueprints left to be imitated is very large and innovating upon the frontier is relatively too costly. This is why skilled human capital is always more productive in rich country, which, being closer to the frontier, can take more advantage of skilled human capital employed in innovation activities. Things radically change when returns to innovation are decreasing. In this case—except for very rich countries and despite the strong comparative advantage of skilled workers in innovation—firms optimally respond to an increase in skilled human capital by allocating part of this additional human capital in imitation activities thereby boosting their output. Hence, since imitation activities is extremely productive at very low stages of development, any additional skilled worker has a very large growth effect when employed in a poor country.

As the economy grows and reaches an intermediate stage of development, imitation becomes more difficult as the number of blueprints left to be imitated shrinks and any additional worker placed in this activity (at this intermediate stage of development) brings a relatively lower contribution to growth. Similarly, at intermediate stages of development, the innovation sector is still in its infancy (relative to that of developed countries) and an increase in skilled workers in innovation has also a limited drive on growth. For this reason, at an intermediate stage of development, the growth effect of skilled human capital reaches its minimum.

Once the economy reaches a higher stage of development and despite decreasing returns, innovation becomes more productive and the growth effect of skilled

workers increase again. Therefore, (only) when skilled workers have a sufficiently strong comparative advantage in innovation and are employed relatively rich countries, VAM's main result is confirmed: the growth effect of skilled human capital increases as the economy gets closer to the technological frontier.

Our empirical analysis supports the model's predictions. We estimate VAM's specification by extending the analysis to a much wider sample of countries (85 between developed and developing economies) for a panel at 10-year intervals covering the period between 1960 and 2000. By using tertiary education as a proxy for skilled human capital and secondary and primary education as a proxy for unskilled human capital, we find that the relation between human capital composition and growth changes significantly with the distance to the technological frontier. There exists a cutoff value of the distance to the technological frontier (approximately found around the poorest OECD country) such that the relationship between the marginal growth effect of an additional skilled worker and the distance to the economic frontier turns from positive (for richer countries) to negative (for poorer countries). These empirical results indirectly support the theoretical scenario in which skilled workers are more efficient in innovation than in imitation and the growth effect of skilled workers is U-shaped. The issues of endogeneity between human capital and growth are addressed by using System Generalized Methods of Moments (GMM) techniques as proposed by Arellano and Bover (1995) and Blundell and Bond (1998). Along with that, we provide several robustness checks by introducing additional controls proxying for institutional quality.

The rest of the paper is organized as follows. In Section 2, we describe the analytical framework. Section 3 is dedicated to the theoretical consequences of nonconstant returns to scale on the dynamics of the catching-up behavior. Section 4 includes the empirical analysis, and Section 5 concludes.

2. THE MODEL

2.1. Basic Analytical Framework

The structure of the economy resembles that of VAM with one main generalization: We allow for nonconstant returns to scale in technological activities. As will become clear later, this analysis is not performed only for the sake of generality but because it sheds light on some important mechanisms that are neutralized in the constant returns to scale (CRS) case.

There exists a finite number of economies, each one with entrepreneurs and population workers of size 1. As VAM (2006), we abstract from international trade and labor mobility.[10] Workers have heterogeneous human capital endowment: the economy is endowed with S highly educated (skilled) workers and U less educated (unskilled) units of labor given exogenously and constant over time (they act as our policy instruments). Time is discrete and all agents live for one period only. In every period and in every country, final output y is produced competitively by

using a continuum of mass 1 of intermediate inputs according to the following Cobb–Douglas production function:

$$y_t = \int_0^1 A_{i,t}^{1-\alpha} x_{i,t}^\alpha di,$$

where $\alpha \in (0, 1)$, $A_{i,t}$ is the productivity in sector i at time t and $x_{i,t}$ is the flow of intermediate good i at time t. The final good sector is competitive, so the price of each intermediate good is equal to its marginal product

$$p_{i,t} = \frac{\partial y_t}{\partial x_{i,t}} = \alpha \left(\frac{A_{i,t}}{x_{i,t}}\right)^{1-\alpha}. \tag{1}$$

In each intermediate sector i, one producer can produce good i with productivity $A_{i,t}$ by using the final good as capital according to a one-for-one technology. The local monopolist chooses $x_{i,t}$ in order to solve

$$\max_{x_{i,t}} \left(p_{i,t} x_{i,t} - x_{i,t}\right),$$

which, by using (1), leads to the following profit in the intermediate sector i:

$$\pi_{i,t} = \left(\frac{1}{\alpha} - 1\right) \alpha^{\frac{2}{1-\alpha}} A_{i,t} = \delta A_{i,t}. \tag{2}$$

2.2. Dynamics of Productivity

At the initial stage of each period, firm i decides upon technology choice. A technology improvement results from a combination of two activities: (1) *imitation* aimed at adopting the world frontier technologies; and (2) *innovation* upon the local technological frontier.

Both activities use unskilled and skilled labor as inputs. The dynamics of the productivity of sector i is the following F increasing in its arguments:

$$A_{i,t} - A_{i,t-1} = F\left(\bar{A}_{t-1} - A_{t-1}, A_{t-1}, m\left(u_{m,i,t}, s_{m,i,t}\right), n\left(u_{n,i,t}, s_{n,i,t}\right)\right),$$

where A_{t-1} is the country's technological frontier at time $t - 1$; $\bar{A}_{t-1}$ is the world technological frontier at time $t - 1$ and therefore $\bar{A}_{t-1} - A_{t-1}$ is the distance from the latter; m and n are, respectively, imitation and innovation activities. The output of activity $j = m, n$ is increasing in its input factors $u_{j,i,t}$ and $s_{j,i,t}$, which are the units of, respectively, unskilled and skilled human capital employed in technological activity j by sector i at time t. Technology progress is assumed to be a linear function of imitation m and innovation n activities:

$$A_{i,t} - A_{i,t-1} = \lambda \left[m\left(u_{m,i,t}, s_{m,i,t}\right)\left(\bar{A}_{t-1} - A_{t-1}\right) + \gamma n\left(u_{n,i,t}, s_{n,i,t}\right) A_{t-1}\right], \tag{3}$$

where $\gamma > 0$ measures the relative efficiency of innovation compared to imitation in generating productivity growth, and $\lambda < 0$ measures the efficiency of the overall process of technological improvement.

We use the following Cobb–Douglas specification for the two kinds of technological activities:

$$m\left(u_{m,i,t}, s_{m,i,t}\right) = u_{m,i,t}^{\sigma} s_{m,i,t}^{\beta}, \tag{4}$$

$$n\left(u_{n,i,t}, s_{n,i,t}\right) = u_{n,i,t}^{\phi} s_{n,i,t}^{\theta}, \tag{5}$$

where $\sigma, \beta, \phi, \theta$ are strictly positive parameters. σ and β represent the elasticity of unskilled (resp. skilled) workers in imitation, whereas ϕ and θ are the elasticity of unskilled (resp. skilled) workers in innovation. As for the elasticity of output to each type of worker, we assume that $\sigma > \phi$. This is to say that unskilled workers are assumed to be better suited to imitation than innovation activities. We share this (reasonable) assumption with VAM. Crucially, instead, we depart from their formalization and do not impose $\sigma + \beta$ and $\phi + \theta$ to be necessarily equal to 1. Returns to scale are then allowed to be nonconstant and heterogeneous in imitation and innovation. One important implication is that in CRS the assumption $\sigma > \phi$ implies $\beta = 1 - \sigma < \theta = 1 - \phi$, so that skilled workers are "forced" to be more productive in innovation than in imitation and moreover the value of their relative efficiency in innovation with respect to imitation is constrained. This assumption may be too restrictive, especially if imitation (as suggested by some empirical and theoretical works[11]) is an "easier" activity with respect to innovation.[12]

Since increasing returns to scale in technological activities seem to be implausible, in what follows we restrict our attention to the case where returns to both technological activities are nonincreasing (so that $\theta \leq 1 - \phi$ and $\beta \leq 1 - \sigma$). This slight generalization [motivated by the literature strand inaugurated by Jones (1995), Kortum (1997), Segerstrom (1998)] is sufficient for the mechanism we have in mind to be unveiled. The dynamics of productivity is then governed by

$$A_{i,t} = A_{i,t-1} + \lambda \left[u_{m,i,t}^{\sigma} s_{m,i,t}^{\beta} \left(1 - a_{t-1}\right) + \gamma u_{n,i,t}^{\phi} s_{n,i,t}^{\theta} a_{t-1} \right] \bar{A}_{t-1}, \tag{6}$$

where $a_{t-1} = A_{t-1}/\bar{A}_{t-1}$ is an inverse measure of the country's distance from the frontier. As in VAM, we let $w_{u,t} \bar{A}_{t-1}$ ($w_{s,t} \bar{A}_{t-1}$) be the wage of unskilled (skilled) labor. Total labor cost of productivity improvement by intermediate firm i at time t is then

$$W_{i,t} = \left[w_{u,t} \left(u_{m,i,t} + u_{n,i,t}\right) + w_{s,t} \left(s_{m,i,t} + s_{n,i,t}\right) \right] \bar{A}_{t-1}.$$

Since entrepreneurs live for one period only—and thus maximize current profit net of labor costs—each intermediate good producer i at date t will choose $(u_{m,i,t}, u_{n,i,t}, s_{m,i,t}, s_{n,i,t})$ to solve the following program:

$$\max_{u_{m,i,t}, u_{n,i,t}, s_{m,i,t}, s_{n,i,t}} \delta A_{i,t} - W_{i,t}. \tag{7}$$

All intermediate firms face the same maximization program, so that in equilibrium $u_{j,i,t} = u_{j,t}$ and $s_{j,i,t} = s_{j,t}$, where $j = m, n$. Moreover, since there is a mass 1 of intermediate firms, the labor market equilibrium implies

$$u_{m,t} + u_{n,t} = U, \tag{8}$$

$$s_{m,t} + s_{n,t} = S. \tag{9}$$

Hence, by using (6) and omitting the time suffix, the first-order conditions can be written as

$$(1-a)\sigma \left(\frac{u_m}{s_m}\right)^{\sigma-1} s_m^{\beta+\sigma-1} = \gamma a \phi \left(\frac{U-u_m}{S-s_m}\right)^{\phi-1} (S-s_m)^{\theta+\phi-1}, \tag{10}$$

$$(1-a)\beta \left(\frac{u_m}{s_m}\right)^{\sigma} s_m^{\beta+\sigma-1} = \gamma a \theta \left(\frac{U-u_m}{S-s_m}\right)^{\phi} (S-s_m)^{\theta+\phi-1}. \tag{11}$$

Dividing across equations and rearranging, we find the condition of equality among marginal rate of technical substitution

$$\psi \frac{(U-u_m)}{(S-s_m)} = \frac{u_m}{s_m}, \tag{12}$$

which gives us u_m as a function of s_m

$$u_m = \frac{\psi s_m U}{S + (\psi - 1) s_m}, \tag{13}$$

where $\psi = \frac{\sigma\theta}{\phi\beta}$. Combining (12) and (11), we obtain

$$h(a)U = (S - (\psi - 1)s_m) \left[\frac{s_m^{1-\beta-\sigma}}{(S-s_m)^{1-\theta-\phi}}\right]^{\frac{1}{\sigma-\phi}}, \tag{14}$$

where[13]

$$h(a) = \left(\frac{\beta\psi^\sigma}{\gamma\theta} \frac{1-a}{a}\right)^{\frac{1}{\sigma-\phi}}.$$

Equation (14) defines an implicit function whose solutions represent the equilibrium values for s_m—and then for u_m through (13) and for s_n and u_n through (8) and (9). Since the objective function defined in (7) is convex, then the equilibrium solution given by the systems (13), (14), (8), and (9) is effectively a maximum for each intermediate firm's profit $\delta A_{i,t} - W_{i,t}$.

It is worth focusing on the role of $\psi = \frac{\sigma\theta}{\phi\beta}$. This parameter provides information on which kind of human capital has the *comparative advantage* in each type of technological activity. More precisely, $\psi > 1$ implies $\frac{\sigma}{\beta} > \frac{\phi}{\theta}$, i.e., the ratio between elasticities of unskilled and skilled human capital in *imitation* is larger

than the ratio between the elasticities of unskilled and skilled human capital in *innovation*. If this is the case, skilled human capital has a comparative advantage in innovation, while unskilled human capital has a comparative advantage in imitation. With CRS, the value of ψ collapses to $\frac{\sigma(1-\phi)}{\phi(1-\sigma)}$, so that $\sigma > \phi$ automatically leads to $\psi > 1$. This is not the case if we allow for decreasing returns to scale (DRS) in innovation (and more generally for heterogeneous returns to scale in technological activities): When θ is untied to ϕ, then $\sigma > \phi$ is compatible to $\psi \leq 1$ when the productivity of skilled workers in innovation is relatively low enough, $\frac{\theta}{\beta} < \frac{\phi}{\sigma}$. Although our model can be solved even for $\psi \leq 1$, we do not consider this case a particularly realistic empirical scenario; so that in what follows, we assume $\frac{\theta}{\beta} > \frac{\phi}{\sigma}$ and then let skilled (unskilled) workers keep the comparative advantage in innovation (imitation).[14] Moreover, since the case of CRS in innovation has been already investigated by VAM, in what follows we will focus on the case of strictly decreasing returns to scale in innovation ($\theta < 1 - \phi$), which, as we will see, leads to some very different implications on the catch-up behavior. We collect these assumptions in the following.

Assumption 1. $\frac{\theta}{\beta} \in \left(\frac{\phi}{\sigma}, \frac{1-\phi}{\beta} \right) \wedge (\beta + \sigma \leq 1)$.

3. EQUILIBRIUM ANALYSIS

Thanks to decreasing returns to scale in innovation, the nonlinear term $[\frac{s_m^{1-\beta-\sigma}}{(S-s_m)^{1-\theta-\phi}}]^{\frac{1}{\sigma-\phi}}$ in (14) is unveiled. The latter, which is equal to 1 in the CRS case, represents the main source of the novel results in our model.

An important implication of this nonlinearity is that we cannot find a closed form optimal solution for s_m. A qualitative analysis is nevertheless possible through the implicit function theorem. However, in order for the implicit function theorem to be applied (and for the analysis to be meaningful), we need that the optimal solution of s_m to (1) exists and (2) be unique.

3.1. Existence and Uniqueness: Interior and Corner Solutions

An equilibrium is defined as follows:

DEFINITION 1. *An equilibrium is a vector* $(u_n^*, u_m^*, s_n^*, s_m^*) \in [0, U]^2 \times [0, S]^2 \subset \mathbf{R}_+^4$, *which solves the system of four equations* (13), (14), (8), *and* (9).

As for existence and uniqueness, the following proposition holds.

PROPOSITION 1. *When assumption 1 holds, a unique equilibrium solution,* $(u_n^*, u_m^*, s_n^*, s_m^*)$ *always exists. Moreover,* $\lim_{a \to 0}(u_n^*, u_m^*, s_n^*, s_m^*) = (0, U, 0, S)$ *and* $\lim_{a \to \bar{1}}(u_n^*, u_m^*, s_n^*, s_m^*) = (U, 0, S, 0)$.

Proof. See Appendix A.

This proposition tells us that when assumption 1 holds, then we should not care about problems of nonexistence or multiplicity of equilibria. But it also tells us that for any value $a \in (0, 1)$, firms optimally decide to employ a nonnegative amount of each input in each technological activity, while the equilibrium is a no-imitation one $[(u_n^*, u_m^*, s_n^*, s_m^*) = (U, 0, S, 0)]$ only for countries at the technological frontier ($a = 1$). This is in sharp contrast with the case of CRS where the equilibrium is actually an interior one only for middle-income countries, while it is a "no-innovation" equilibrium for a group of very poor countries and a "no-imitation" one for a group of very rich countries

3.2. Impact of Skilled Workers on Input Allocation and Technological Outputs

This section analyzes the optimal response of firms following an exogenous change in the economy's endowment of skilled workers (S) . We focus only on skilled human capital S both because the latter is our main interest and because the comparative statics with respect to unskilled human capital U and the proximity to the technological frontier a are identical to the CRS case.[15] When non-CRS are allowed for, and assumption 1 holds, the way firms allocate additional skilled workers across technological activities is crucially different from the CRS knife-edge case and varies according to both the proximity to the technological frontier and the relative efficiency of skilled human capital in innovation. We summarize our results in Table 1,[16] wherein we report the direction of the change in the equilibrium allocation of inputs $(u_n^*, u_m^*, s_n^*, s_m^*)$ and in the equilibrium level of outputs across technological activities $[m(u_m^*, s_m^*)$ and $n(u_n^*, s_m^*)]$ following a change in S. The double arrows (upward or downward) identify opposite dynamics with respect to the CRS case (i.e., when the main variable of interest behave differently from the CRS case), while simple arrows are used to describe the dynamics that are qualitatively identical to the CRS case.

As we can see, the dynamics stemming from our generalization are much more complex than those of the CRS case analyzed by VAM and depend on both the degree of efficiency of skilled workers in imitation (shaping the alternative scenarios in Table 1) and the proximity of the economy to the technological frontier (whose key thresholds are described in row 3). However, this additional complexity enables us to reveal a theoretical scenario that is supported by our empirical analysis.

The scenarios we consider all satisfy Assumption 1 in that skilled workers have comparative advantage in innovation ($\frac{\theta}{\beta} > \frac{\phi}{\sigma}$), and returns are assumed to be decreasing in innovation ($\theta + \phi < 1$) and nonincreasing in imitation ($\beta + \sigma \leq 1$) but differ according to the different values of the ratio $\frac{\theta}{\beta}$, which, as already argued, measures the relative efficiency of skilled workers in imitation with respect to innovation. We consider two main scenarios. In Scenario 1 (columns 1 and 2), the relative efficiency of skilled workers in innovation is not very high, so their comparative advantage in innovation is limited ($\frac{\theta}{\beta} \leq \frac{1}{\beta+\sigma}$). As a result,

TABLE 1. Equilibrium responses to changes in S for different values of $\frac{\theta}{\beta}$

		Scenario 1: $\frac{\theta}{\beta} \leq \frac{1}{\beta+\sigma}$ weak comparative advantage		Scenario 2: $\frac{\theta}{\beta} > \frac{1}{\beta+\sigma}$ strong comparative advantage					CRS
		(1) case a $\frac{\phi}{\sigma} < \frac{\theta}{\beta} \leq 1$	(2) case b $1 < \frac{\theta}{\beta} \leq \frac{1}{\beta+\sigma}$	(3) case a $\frac{1}{\beta+\sigma} < \frac{\theta}{\beta} \leq \frac{1-\sigma}{\beta}$		(4) case b $\frac{1-\sigma}{\beta} < \frac{\theta}{\beta} \leq \frac{1-\phi}{\beta}$			(5) $\frac{\theta}{\beta} = \frac{1-\phi}{1-\sigma}$
	a	$(0, 1)$	$(0, 1)$	$(0, \hat{a}]$	$(\hat{a}, 1)$	$(0, \hat{a}]$	$(\hat{a}, a^*]$	$(a^*, 1)$	$(\underline{a}, \bar{a})$
(1)	s_m^*	⇑	⇑	⇑		⇑		↓	↓
(2)	u_m^*	⇑	↓	↓		↓			↓
(3)	$m\left(s_m^*, u_m^*\right)$	⇑	⇑	⇑	↓	⇑		↓	↓
(4)	s_n^*	↑	↑	↑		↑			↑
(5)	u_n^*	⇓	↑	↑		↑			↑
(6)	$n\left(s_n^*, u_n^*\right)$	↑	↑	↑		↑			↑

returns to innovation are strongly decreasing. In Scenario 2 (columns 3 and 4), skilled workers are more efficient in innovation and their comparative advantage in innovation is strong ($\frac{\theta}{\beta} > \frac{1}{\beta+\sigma}$). As a result, returns to innovation are only slightly decreasing.

As can be appreciated from Table 1, these two main scenarios are associated with different responses of the output of imitation activities m, following a change in S (row 3). Each of these two main scenarios encompasses two subcases in which the qualitative behavior of m does not change but where some differences in the response in the equilibrium allocation of inputs ($u_n^*, u_m^*, s_n^*, s_m^*$) can be observed. In column (5), we report the CRS case, $\theta + \phi = \beta + \sigma = 1$, which results in an even stronger comparative advantage of skilled human capital in innovation, to highlight how our generalization leads to crucially different dynamics with respect to the knife-edge case analyzed by VAM.

As for the proximity to the technological frontier, the key thresholds change across scenarios. In Scenario 1, the qualitative behavior of the variable considered does not change as the economy gets closer to the frontier. By contrast, the second main scenario is more complex as the qualitative behavior of the equilibrium values is affected by the value of a as well. More precisely, Scenario 2a (column 3) introduces a threshold $\hat{a}$ such that the response of intermediate firms in poor enough countries, $a \in (0, \hat{a}]$, is such that the output of imitation increases after an increase in S, while firms' response in sufficiently rich countries, $a \in (\hat{a}, 1)$, leads to a decrease in the output of imitation activities. While in this subscenario, firms always employ more skilled workers in imitation activities after an increase in S (s_m^* is always increasing in S), Scenario 2b (column 4) introduces an additional threshold, $a^* > \hat{a}$, such that very rich countries, $a \in (a^*, 1)$, reallocate both unskilled and skilled workers from imitation to innovation activities after an increase in S and then, straightforwardly, the amount of output produced by imitation activities is reduced.

Scenario 2b is actually the *only case* where every variable considered behaves exactly as in the CRS case. In any other case, the optimal share of skilled workers employed in imitation (s_m^* in row 1) increases at all stages of development following an increase in S. This is in sharp contrast to the CRS case where, instead, the share of skilled workers engaged in imitation activities always decreases in response to an increase in the overall endowment of skilled workers in the economy. This important difference is at the basis of more complex dynamics that are neglected in VAM (see the detailed description of the two main scenarios discussed below).

Generally speaking, the generalization proposed in our paper shows that for poor and developing countries an increase in the endowment of skilled workers leads to its reallocation to both imitation and innovation activities rather than solely to innovation activities as in VAM.

Finally, we also note that, as far as innovation activities are concerned, the only deviation from the CRS case lies in the negative effect of S on the amount of *unskilled* workers allocated in innovation in Scenario 1a (row 5, column 1). However, this difference does not affect the sign of the effect of an additional unit

of skilled human capital on the output of innovation activities, which is always positive (as in the CRS case) in every subcase considered (row 6). For these reasons, our description of Table 1 will be focused basically on the impact of S in the allocation of inputs and in the resulting output of imitation activities. Nonetheless, such differences lead to drastic consequences on the growth prospects.

Scenario 1: Skilled workers' efficiency in imitation is relatively high, $\frac{\theta}{\beta} \leq \frac{1}{\beta+\sigma}$. Scenario 1 (columns 1 and 2) assumes that the efficiency of skilled workers in imitation is relatively high and that, for this reason, imitation activities may be a relatively profitable activity for maximizing firms. As long as $\frac{\theta}{\beta} \leq \frac{1}{\beta+\sigma}$, it can be shown that an increase in the economy's endowment of skilled workers leads to the allocation of a positive fraction of these workers to imitation activities *at all stages of economic development*, i.e., $\frac{\partial s_m^*}{\partial S} > 0$ for any $a \in (0, 1)$ (row 1). The increase in the fraction of skilled workers devoted to imitation activities leads to a contextual increase in the economy's imitation output [$\frac{\partial m}{\partial S} > 0$ for any $a \in (0, 1)$, row 3]. This is not the case in the CRS case (column 5) where all skilled workers would have been employed solely in innovation activities and no increase in imitation output would have been observed.

In Scenario 1a (column 1), the efficiency of skilled workers in imitation is assumed to be particularly high, actually higher than their efficiency in innovation. When this is so, firms optimally decide to support imitation activities by reallocating a share of unskilled workers from innovation to imitation (row 2), leading to a straightforward increase in the output of imitation activities (row 3).

In Scenario 1b (column 2), the relative efficiency of skilled workers in imitation is lower but still high. In this scenario, following an increase in S, unskilled workers will move out from the imitation sector to be employed in innovation (row 2). Despite the decrease in share of unskilled workers devoted to imitation, the inflow of skilled workers into the imitation sector (row 1) is sufficient to ensure an overall increase in the output of the imitation sector at any stage of development (row 3).

Scenario 2: Skilled workers' efficiency in imitation is relatively low, $\frac{\theta}{\beta} > \frac{1}{\beta+\sigma}$. Scenario 2 (columns 3 and 4) examines the possibility that the efficiency of skilled workers in imitation, as opposed of that in innovation, may be low. This scenario leads to slightly more complex dynamics since, in this case, the allocation of skilled workers to imitation and innovation activities is linked to both the relative efficiency of skilled workers across technological activities and to the relative development stage of the economy under consideration.

A key feature of this scenario is the existence of a value $\hat{a} \in (0, 1)$ of the proximity to the technological frontier, above which the impact of an additional skilled worker on the output of imitation activities (the derivative $\frac{\partial m}{\partial S}$) changes from positive to negative.

Scenario 2a (column 3) considers the case in which the efficiency of skilled workers in innovation is relatively larger than that of Scenario 1b: $\frac{\theta}{\beta} \in (\frac{1}{\beta+\sigma}, \frac{1-\sigma}{\beta}]$.

Following an increase in S, a fraction of the new skilled workers in the economy will be allocated to imitation activities (row 1 as in Scenario 1b). This time, however, skilled workers being assumed to be more efficient in innovation than in Scenario 1b, ceteris paribus, the decrease in u_m^* will be more pronounced and sufficient to lead to a decrease in the output of the imitation sector but only for countries that are sufficiently close to the technological frontier, $a > \hat{a}$, where the innovation sector is well developed and skilled workers experience higher returns to innovation (row 3).

Scenario 2b (column 4) considers the case in which the efficiency of skilled workers in imitation is the lowest possible compatible with decreasing returns to scale in innovation, i.e., $\theta \in (1 - \sigma, 1 - \phi)$. This scenario is identical to 2a with the only difference that, for countries very close to the technology frontier $(a > a^*)$, the increase in the endowment of skilled workers will be only allocated to innovation activities (rather than being split into imitation and innovation activities). Also, some of the skilled workers originally employed in imitation will be reallocated to innovation leading to a reduction in the overall output of imitation activities (row 3). As we can see by comparing the last two columns of Table 1, this is the only case where the impact of a change in S on inputs' allocation is identical to the case of CRS.

This result only applies, however, to countries very close to the technology frontier. For middle-income countries, $a \in [\hat{a}, a^*)$, performing innovation at this stage of development is a relatively unproductive activity regardless of the fact that skilled workers have a strong comparative advantage in innovation. As a result of this, part of the increase in S will be devoted to imitation (row 1). For this set of countries, the increase in s_m^* is nevertheless not sufficient to offset the effect of reduction in the other imitation input, u_m^* (row 2), on the output of technological improvements coming from imitation activities m, which will be then still decreasing in S for any $a \in [\hat{a}, a^*)$ (row 3).

Similar dynamics apply to countries at very low stages of development $a \in [0, \hat{a})$. When a country is poor, innovation is even less productive relative to imitation. In this case, the increase in s_m^* is relatively large enough to more than offset the reduction in u_m^* and therefore leading to an *increase* in m.

The deviation in the sign of $\frac{\partial m}{\partial S}$ with respect to the CRS case, which occurs in Scenarios 1a, 1b, and 2a, is the main determinant of the differences in the dynamics of the catch-up as we will see in the following section.

4. GROWTH ANALYSIS

The main theoretical result of VAM (Lemma 2) is that the growth effect of an additional skilled worker $(\frac{\partial g}{\partial S})$ is always larger for economies that are closer to the frontier, so that $\frac{\partial^2 g}{\partial a \partial S}$ is positive *at any distance from the technological frontier*. The purpose of this section is to analyze the sign of $\frac{\partial^2 g}{\partial a \partial S}$ as a function of a in a more general context and to show that a slight relaxation of the CRS assumption is

able to match better with empirical data. By using (3), and since each intermediate firm behaves the same in equilibrium, we can write the equilibrium growth rate as

$$g \equiv \frac{A_t - A_{t-1}}{A_{t-1}} = \lambda m(u_m^*, s_m^*) \left(\frac{1-a}{a}\right) + \lambda \gamma n\left(u_n^*, s_n^*\right), \tag{15}$$

where $m(\cdot)$ and $n(\cdot)$ are defined by (4) and (5). This expressions clearly shows that—irrespectively from the nature of the returns to technological activities—close to the frontier (for a close to 1), growth is basically innovation driven, while far from the frontier (for a close to zero), growth is mainly imitation driven. By using (13), (14), (8), and (9) in (15) and a bit of algebra, we can express the growth rate as function of s_m^* only[17]:

$$g = \lambda \gamma h\left(a\right)^{-\phi} \left[\frac{s_m^{*\,(1-\beta-\sigma)\phi}}{(S - s_m^*)^{(1-\theta-\phi)\sigma}}\right]^{\frac{1}{\sigma-\phi}} \left(S - \frac{\beta-\theta}{\beta}s_m^*\right). \tag{16}$$

Differentiating (15) with respect to S, we can write[18]

$$\frac{\partial g}{\partial S}/\lambda = \frac{\partial m}{\partial S}\left(\frac{1-a}{a}\right) + \gamma\frac{\partial n}{\partial S}, \tag{17}$$

which, by using again (13), (14), (8), and (9) and after some algebra[19] can be written as

$$\frac{\partial g}{\partial S}/\lambda = \theta\gamma h\left(a\right)^{-\phi}\left[\frac{s_m^{*\,(1-\beta-\sigma)\phi}}{(S - s_m^*)^{(1-\theta-\phi)\sigma}}\right]^{\frac{1}{\sigma-\phi}} > 0. \tag{18}$$

Equations (17) and (18) are again two different ways of representing the marginal growth effect of an additional skilled worker. By looking at (17), we note that, due to the linearity of the Nelson–Phelps productivity equation (3), an argument similar to the growth rate also applies here: For poor economies (a close to 0), $\frac{1-a}{a}$ is very large and therefore $\gamma\frac{\partial n}{\partial S}$ is relatively unimportant. As a consequence, $\frac{\partial g}{\partial S}/\lambda$ is almost equal to $\frac{\partial m}{\partial S}\left(\frac{1-a}{a}\right)$. The opposite happens for rich economies since in this case $\frac{1-a}{a}$ is very small and therefore the weight of the marginal effect of an additional skilled worker in imitation has a very small growth effect because rich economies basically grow out of innovation. Differentiating (18) with respect to a and again by using (18), we find, after some cumbersome computations,[20]

$$\frac{\partial^2 g}{\partial a\,\partial S} = -\frac{\partial g}{\partial S}\frac{h'\left(a\right)}{h\left(a\right)}\left[\phi - \frac{(1-\beta-\sigma)\,\phi\left(S - s_m^*\right) + \sigma\left(1-\theta-\phi\right)s_m^*}{F\left(S, U, a\right)}\right], \tag{19}$$

where

$$F(S, U, a) = \frac{(\psi - 1)(\sigma - \phi)(S - s_m^*) s_m^*}{(S + (\psi - 1) s_m^*)}$$
$$+ (1 - \theta - \phi) s_m^* + (1 - \beta - \sigma)(S - s_m^*) > 0$$

is clearly positive with nonincreasing returns to scale in technological activities. This expression represents the core of our analysis. It shows that the sign of $\frac{\partial^2 g}{\partial S \partial a}$ depends on the difference between the two terms in the parenthesis. The first term, ϕ, is clearly positive. As for the second term, $\frac{(1 - \beta - \sigma)\phi(S - s_m^*) + \sigma(1 - \theta - \phi)s_m^*}{F(S, U, a)}$, it is clearly nonnegative for any $a \in (0, 1)$ when returns to technological activities are nonincreasing, being zero if and only if $\theta + \phi = \beta + \sigma = 1$, i.e., in the CRS case. Hence, the presence of this second term, unveiled by decreasing returns to scale in innovation (which emphasizes the relative productivity of skilled workers in imitation activities especially for poor countries) makes the sign of $\frac{\partial^2 g}{\partial S \partial a}$ a priori ambiguous, being positive (as in VAM) if the first effect dominates and negative otherwise. So what are the determinants of the sign of $\frac{\partial^2 g}{\partial S \partial a}$ and then of the relative strength of these two effects? A complete answer is in the following proposition.

PROPOSITION 2. *The sign of $\frac{\partial^2 g}{\partial a \partial S}$ is governed by the following rules:*

$$\frac{\theta}{\beta} \in \left(\frac{\phi}{\sigma}, \frac{1}{\beta + \sigma} \right] \Rightarrow \frac{\partial^2 g}{\partial a \partial S} < 0, \forall a \in (0, 1)$$

$$\frac{\theta}{\beta} \in \left(\frac{1}{\beta + \sigma}, \frac{1 - \phi}{\beta} \right) \Rightarrow \begin{cases} \frac{\partial^2 g}{\partial a \partial S} \leq 0, \forall a \in (0, \hat{a}] \\ \frac{\partial^2 g}{\partial a \partial S} \geq 0, \forall a \in (\hat{a}, 1]. \end{cases}$$

Hence,

$$\text{sign} \frac{\partial^2 g}{\partial a \partial S} = -\text{sign} \frac{\partial m}{\partial S}, \forall a \in (0, 1).$$

Proof. See Appendix A.

Proposition 2 contains the main result of this paper and delivers three main messages. First, it tells us that whenever returns to innovation are allowed to be strictly decreasing,[21] there is always a threshold level for the proximity to the technological frontier $\hat{a}$, such that any country below this threshold is characterized by a negative value of $\frac{\partial^2 g}{\partial a \partial S}$.

Second, Proposition 2 confirms the importance of the role of $\frac{\theta}{\beta}$ as it tells us that when skilled workers' relative efficiency in imitation is high ($\frac{\theta}{\beta}$ is smaller than $\frac{1}{\beta + \sigma}$), then $\frac{\partial^2 g}{\partial a \partial S}$ is negative *for any country, irrespective of its distance from the technological frontier*. By contrast, if skilled workers' relative efficiency in imitation is low ($\frac{\theta}{\beta}$ is larger than $\frac{1}{\beta + \sigma}$), then $\frac{\partial^2 g}{\partial a \partial S}$ is positive for countries close to the technological frontier (as in VAM) but negative for the rest of the countries.
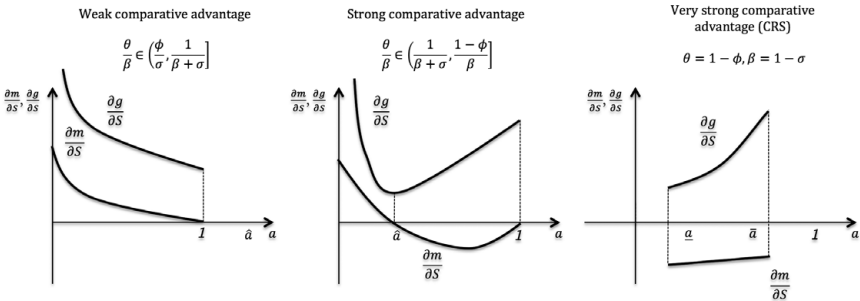
FIGURE 1. The marginal effect of S on growth and imitation outputs for different levels of comparative advantage of skilled human capital in innovation.

Third, Proposition 2 establishes an intimate link between the shape of growth effect of S during the catching-up process and the marginal productivity of S in imitation: Whenever—for an economy located at a given distance to the technology frontier—an additional unit of skilled human capital increases the amount of technological improvements of imitation activities ($\frac{\partial m}{\partial S} > 0$), then the marginal growth effect of this additional skilled worker is decreasing with respect to the proximity to the technology frontier and therefore $\frac{\partial^2 g}{\partial a \partial S}$ is negative. The opposite happens when, as in the CRS case, $\frac{\partial m}{\partial S}$ is negative. As an implication, *the marginal growth effect of an additional skilled workers reaches a minimum when the marginal productivity of skilled workers in imitation is null.*

Figure 1 shows the different behavior of the marginal growth effect of skilled human capital *and* the marginal productivity of skilled workers in imitation both as a function of the proximity to the technological frontier in the distinct scenarios described in Proposition 2. Notice that when $\frac{\theta}{\beta} \in (\frac{1}{\beta+\sigma}, \frac{1-\phi}{\beta})$, then $\frac{\partial g}{\partial S}$ is graphically represented by a U-shaped curve that horizontally moves to the right or to the left according to whether θ approaches, respectively, $(1-\sigma)$ or $(1-\phi)$. Also notice that when $\frac{\theta}{\beta} \leq \frac{1}{\beta+\sigma}$, then $\frac{\partial g}{\partial S}$ is monotonically decreasing in $a \in (0, 1)$ while it is monotonically increasing in $a \in (\underline{a}, \overline{a})$ only when $\theta = 1 - \phi$ and $\beta = 1 - \sigma$.

What is the intuition for such a big difference between the CRS and the DRS case due to an even infinitesimal reduction in θ? And why the sign of $\frac{\partial^2 g}{\partial a \partial S}$ is so intimately linked to the sign of $\frac{\partial m}{\partial S}$? A clear answer can be if we differentiate (17) with respect to a to find

$$\frac{\partial^2 g}{\partial S \partial a} = \frac{1}{\lambda}\left[-\frac{\partial m}{\partial S}\frac{1}{a^2} + \frac{\partial^2 m}{\partial S \partial a}\left(\frac{1-a}{a}\right) + \gamma\frac{\partial^2 n}{\partial S \partial a}\right]. \tag{20}$$

This expression shows that the value of $\frac{\partial^2 g}{\partial a \partial S}$ can be considered as the sum of three different effects. The first $-\frac{\partial m}{\partial S}\frac{1}{a^2}$, being clearly the most important when a

is close to zero. This term has a clear meaning: It tells us that the growth effect of an additional unit of skilled human employed in imitation—the term $-\frac{\partial m}{\partial S}\frac{1-a}{a}$ in (17)—is large for economies far from the technological frontier and small otherwise. Since, as shown in Table 1, when decreasing returns to innovation and nonincreasing return to imitation are assumed, an additional unit of skilled worker induces firms in countries at low stages of development (where $a < \hat{a}$) to increase the output of imitation activities ($\frac{\partial m}{\partial S} > 0$), then $\frac{\partial^2 g}{\partial a \partial S}$ is reasonably negative when $a < \hat{a}$ and therefore the growth effect of S is decreasing in a. The magnitude of this term is comparatively large and it dominates the other two effects at any distance to the technological frontier.

Hence, the sharp difference between the constant and the decreasing returns to scale in innovation activities lies in the sign taken by $\frac{\partial m}{\partial S}$. When returns to technological activities are constant, the impact of skilled human capital on imitation is *always negative* as profit-maximizing firms always react by reallocating both kinds of human capital from imitation to innovation activities, regardless the distance from the technological frontier (see Table 1). Since the growth effect driven by the imitation component in (20) is comparatively large and of opposite sign with respect to $\frac{\partial m}{\partial S}$, then the value of $\frac{\partial^2 g}{\partial S \partial a}$ is clearly positive. As soon as the value of θ is lower than $1 - \phi$ (Scenario 2), and then returns to innovation activities are strictly decreasing, the prediction of the model changes dramatically as the impact of skilled human capital on imitation activities for very poor countries ($a \to 0$) becomes strictly positive and equal to $\lim_{a \to 0} \frac{\partial m}{\partial S} = (\frac{U}{S})^{\sigma}(1 - \sigma) > 0$. Accordingly, the value of $\frac{\partial^2 g}{\partial S \partial a}$ is negative and large.

The empirical analysis suggests the existence of a U-shaped relationship between the growth effect of skilled human capital and the proximity to the technological frontier[22] and, hence, it indirectly supports the Scenarios 2a and 2b of Table 1 (columns 3 and 4) according to which $\frac{\theta}{\beta} \in (\frac{1}{\beta + \sigma}, \frac{1 - \phi}{\beta})$. Our model provides a rationale for this empirical finding: When $\frac{\theta}{\beta} \in (\frac{1}{\beta + \sigma}, \frac{1 - \phi}{\beta})$ and then skilled workers' comparative advantage in innovation is strong enough, the economic intuition for the U-shaped relationship between the growth effect of skilled human capital and the proximity to the technological frontier can be explained as follows.

Below $\hat{a}$, firms optimally allocate any additional resources of skilled human capital to both imitation and innovation activities, leading to an increase in the output of the imitation activities. Since imitation is the main driver of the growth of poor countries, the increase in its output also increases growth the more countries are farther away from the frontier.

When the country reaches a higher stage of development, $\hat{a}$, the marginal growth effect of S reaches its minimum value and moreover, by Proposition 2, an infinitesimal change in S will leave imitation output unchanged. This is so since the positive effect of an increase in the employment of skilled workers in imitation is fully compensated by the negative effect due to reallocation of unskilled workers from imitation to innovation.

As an economy closes the gap with the technological leader, imitation activities contribute less to economic growth relative to innovation. Above $\hat{a}$, the marginal productivity of a skilled worker employed in imitation turns to be negative, despite firms continue to allocate part of additional skilled workers in imitation. But now the increase of skilled workers employed in imitation is too small compared to the reduction of unskilled workers. This leads to a reduction in the output of imitation activities while that of innovation increases even more. This is all the more true once the economy reaches a^*, above which firms start to reallocate skilled workers from imitation to innovation activities following an increase in S. As a consequence of this dynamics, the marginal growth effect of an increment in S will be increasing as an economy gets closer to the frontier through its effect on innovation activities. An exogenous increase in S leads to an increase in the growth rate of the economy through the increase in the output of innovation activities up to a point where, very close to the frontier, almost only innovation is performed.

The differences in policy implications between our model and previous literature are, hence, noteworthy. Our theoretical results, in fact, emphasize the fundamental role of skilled human capital for countries at low development stages, even if these mainly perform imitation activities and little (or none) innovation activities.

5. EMPIRICAL ANALYSIS

5.1. Empirical Model and the Treatment of Endogeneity

We follow VAM[23] and test the predictions of our theoretical model with the following empirical specification for total factor productivity (TFP) growth:

$$g_{j,t} = \alpha_{0,j} + \alpha_1 z_{j,t-1} + \alpha_2 f_{j,t-1} + \alpha_3 z_{j,t-1} * f_{j,t-1} + \epsilon_{j,t}, \tag{21}$$

where $g_{j,t} = \ln A_{j,t} - \ln A_{j,t-1}$ is TFP growth and $A_{j,t}$ represents the TFP in country j at period t. The variable $z_{j,t-1} = \ln a_{j,t-1} = \ln A_{j,t-1} - \ln \bar{A}_{t-1}$ is the log of the proximity to the TFP frontier[24] in the initial period (this is a negative number) while $f_{j,t-1}$ represents human capital that (depending on the empirical specification under consideration) will be proxied by the (i) fraction(s) of the workforce with a specific education attainment level, or (ii) average number of years of schooling (in tertiary, secondary, or primary education). Our empirical specification, hence, fully resembles that of VAM.

The estimation of the empirical model in (21) poses a number of econometric challenges. On the one hand, as argued by Nickell (1981), a "dynamic panel bias" may arise when lagged values of the dependent variable are correlated to the fixed effect in the error term.[25] This positive correlation violates a necessary assumption for the consistency of ordinary least-squares estimators that are, hence, not valid for inference. On the other hand, an additional source of bias might arise, as

pointed out by Bils and Klenow (2000), due to the positive correlation between the explanatory variables [i.e., the educational variables in equation (21)] and the error term creating additional severe endogeneity problems.

An intuitive first attack to these issues is to draw the fixed effect out of the error term by entering dummies for each individual through the so-called Least Squares Dummy Variables (LSDV) estimator as well as instrumenting all the (endogenous) right-hand side variables by their lagged values. As argued by Aghion et al. (2009), however, the use of LSDV does not solve a variety of problems that are intrinsic to the estimation of the empirical model in equation (21). To start with, the use of the lagged realization of education variables or the use of education spending lagged 10 years as instruments for education levels may still conduce to biases due to the instrument's potential correlation to omitted variables specific to each country.[26]

Additionally, as argued by Kiviet (1995) and Bond (2002), the within-groups transformation does not fully eliminate dynamic panel bias. Kiviet (1995) devises a strategy to correct for this bias. This correction, however, works only in the context of balanced panels and, crucially, does not address the potential endogeneity of other regressors as it would be needed, instead, in our case due to the potential simultaneous relation between educational variables and TFP.

Last but not least, educational variables are not only endogenous to the dependent variable, they are also persistent over time. Fixed-effect estimators that exploit the within-country variation in the data do not seem to represent, hence, the most appropriate choice in this context due to the limited power of lagged explanatory variables to be used as instruments.

As a solution to these above-mentioned issues, the Arellano and Bover (1995) and Blundell and Bond (1998) GMM estimator builds a system of equations by exploiting the assumption that first differences of instrument variables are uncorrelated with the fixed effects. As argued by Roodman (2009b) " for random walk–like variables, past changes may indeed be more predictive of current levels than past levels are of current changes so that the new instruments are more relevant" (p. 28). System GMM estimators, then, prove to be of the highest advantage with persistent series in which the lagged levels of explanatory variables are weak instruments for subsequent changes and when both dynamic panel bias and additional endogeneity biases of covariates are likely to affect the estimation.

The validity of GMM estimates, however, depends on the assumption that the idiosyncratic disturbance terms are not serially correlated as well as on the paucity of the instrumental set employed to fit the endogenous regressors. Regarding the first condition, Arellano and Bond (1991) developed a test of autocorrelation of the second order, which checks for the validity of lagged variables as instruments. About the second requirement, the works of Andersen and Soerensen (1996), Bowsher (2002), and Roodman (2009a) provide an in-depth discussion of how instrument proliferation (easily obtained with the system of equations built for the SYSGMM estimators) vitiates the estimation of the Hansen test, providing unreliable information on the robustness of the instrumental set and on the overall

validity of GMM estimations. Limiting the lag depth (i.e., collapsing the instrument) is, hence, a necessary, though usually overlooked, condition in order to avoid false positive. Roodman (2009a) suggests that the instrumental count should be kept as parsimonious as possible and especially that this, as a general rule of thumb, should not exceed the number of groups in the SYSGMM regression. In what follows, hence, we will estimate the impact of human capital composition on growth through SYSGMM estimators while carefully taking into consideration all the above-mentioned estimation issues.

5.2. The Data

The data used to test the empirical model in equation (21) cover 85 countries for 10-years time spans over the period 1960–2000. The information we use comes from different sources. As for the GDP data, we rely on the Penn World Tables provided by Heston et al. (2002). Since capital stock data are not available in this database, a common solution is to build estimates by applying the Perpetual Inventory Method (PIM) to time series investment data. Even though the PIM is a well-established method in the empirical literature, it is not without its concerns. These relate to the possible measurement error affecting the estimation of the initial capital stock year that could arise if the investment data do not go back far enough in time. In a recent study, Baier et al. (2006) build capital stock estimates through the PIM by exploiting long investment time series (in some cases dating back to the 18th century) that are provided in Mitchell (1998). Investment data prior to 1992 are measured by using the (i) International Historical Statistics: The Americas 1750–1993, (ii) International Historical Statistics: Africa, Asia, and Oceania 1750–1993, and (iii) International Historical Statistics: Europe 1750–1993 so that the measurement error on the initial capital stock is of virtually no concern in these estimates. We use Baier et al. capital stock estimates and follow VAM to build TFP as output per worker minus capital per worker times capital share. Hence, we compute the proximity to the technological frontier as the ratio of each country's TFP level to that of the United States.

To proxy for human capital types, we employ Cohen and Soto (2007) data that provide information about the share of the workforce aged 25 or more having completed tertiary, secondary, or primary education for a large sample of countries at 10-years intervals, based on both census and enrolment data collected in the UNESCO Statistical Yearbook as well as in the United Nations Demographic Yearbook.[27]

The descriptive statistics for the variables of interest are given in Table 2. The average TFP proximity of the OECD sample with respect to the US[28] is 0.69, while it is only 0.22 for the subsample of developing countries. As expected, there are also substantial differences in human capital endowment across countries, with the average number of years of tertiary schooling in OECD countries standing at 0.51 compared to 0.22 for the developing countries subsample.[29]

TABLE 2. Descriptive statistics

Variable	Obs	Mean	Std. dev.	Min	Max
All countries					
TFP gap	401	0.34	0.27	0	1
Fraction tertiary	401	0.06	0.07	0	0.35
Fraction secondary	401	0.12	0.13	0	0.56
Fraction primary	401	0.24	0.19	0.01	0.85
Mean years tertiary	401	0.24	0.28	0	1.4
Mean years secondary + primary	401	4.39	2.9	0.08	11.01
Mean years secondary	401	1.45	1.13	0.04	5.13
Mean years primary	401	0.75	0.8	0	3.36
OECD countries					
TFP gap	104	0.69	0.17	0.21	1
Fraction tertiary	104	0.13	0.08	0.01	0.35
Fraction secondary	104	0.25	0.15	0.01	0.56
Fraction primary	104	0.39	0.22	0.03	0.85
Mean years tertiary	104	0.51	0.32	0.04	1.4
Mean years secondary + primary	104	7.68	1.49	2.75	11.01
Mean years secondary	104	2.32	1.34	0.18	5.13
Mean years primary	104	1.53	0.9	0.08	3.36
Developing countries					
TFP gap	297	0.22	0.17	0	0.9
Fraction tertiary	297	0.04	0.04	0	0.32
Fraction secondary	297	0.08	0.09	0	0.48
Fraction primary	297	0.19	0.15	0.01	0.76
Mean years tertiary	297	0.15	0.18	0	1.3
Mean years secondary + primary	297	3.24	2.33	0.08	9.93
Mean years secondary	297	1.15	0.87	0.04	4.54
Mean years primary	297	0.47	0.54	0	2.89

5.3. Empirical Predictions of the Theoretical Model

The theoretical model developed in the preceding sections predicts a positive marginal effect on growth of skilled workers [$\frac{\partial g}{\partial S}$ in equation (18)]. In the empirical model, expressed by (21) this theoretical prediction would translate into the following:

$$\frac{\partial g_{j,t}}{\partial f_{j,t-1}} = \alpha_2 + \alpha_3 z_{j,t-1} > 0. \tag{22}$$

The overall effect of a marginal increase in human capital on the growth rate is then proxied by a linear function of $z_{j,t-1}$ and so it may change according to a country's relative stage of development with respect to the world productivity frontier. More precisely, given the presence of the interaction term $z_{j,t-1} \times f_{j,t-1}$,

the overall effect of an additional $f_{j,t-1}$ (the share of tertiary educated population) could be graphically represented by a straight line taking values for $z_{j,t-1} \in \mathbf{R}^-$, where α_2 is the vertical intercept and α_3 is the slope.

As for how the growth effect of skilled workers changes with respect to the proximity to the technological frontier, Proposition 2 predicts two different behavior for $\alpha_2 + \alpha_3 z_{j,t-1}$ according to whether Scenario 1 or 2 in Table 1 applies. In the first case, when skilled workers' efficiency in imitation is large enough, we know that the growth effect of skilled workers is always decreasing as the economy approach the technological frontier. By contrast, in Scenario 2 when skilled workers' efficiency in imitation is relatively small enough, then $\frac{\partial g}{\partial S}$ is U-shaped with respect to the proximity to the technological frontier. These results suggests that the empirical analysis should lead to a value of $\alpha_2 + \alpha_3 z_{j,t-1}$ that cannot be increasing in $z_{j,t-1}$ for every subset of values of the latter and should be certainly decreasing for intervals of $z_{j,t-1}$, where the later takes sufficiently small values.

As for the expected sign of the coefficient α_2 notice that, for countries very close to the world frontier, the value of $z_{j,t-1}$ is close to zero and then the marginal growth effect of human capital for these developed countries can be approximated by the value of α_2 only. In other words, our model predicts a positive value for α_2 for countries close enough to the technology frontier:

$$\lim_{A_{j,t-1} \to \bar{A}_{t-1}} \frac{\partial g_{j,t}}{\partial f_{j,t-1}} = \alpha_2 > 0. \tag{23}$$

This is not necessarily true for developing countries. For countries far away from the frontier, in fact, the value of the coefficient α_2 could be negative while still being consistent with the theoretical predictions of our model of a positive effect of skilled workers on growth as described above. This is so if the term $\alpha_3 z_{j,t-1}$ is positive and relatively larger in absolute value than α_2. Notice that, being $z_{j,t-1}$ negative by construction, a necessary condition for this to happen is that the coefficient α_3 is also negative.

As for this latter, α_3 represents the empirical counterpart of the cross-derivative $\frac{\partial^2 g}{\partial a \partial S}$ that has been analyzed in Proposition 2. From an empirical point of view this is shown as

$$\frac{\partial^2 g_{j,t}}{\partial f_{j,t-1} \partial z_{j,t-1}} = \alpha_3. \tag{24}$$

As detailed in preceding sections, we already know that, in the knife-edge case of CRS, $\frac{\partial^2 g}{\partial a \partial S}$ is always positive, hence predicting a positive value for α_3. This is not necessarily true in our theoretical generalization, where $\frac{\partial^2 g}{\partial a \partial S}$ can either assume positive or negative values as a result of different combinations of parameter-elasticities associated to human capital in innovation and imitation activities and fundamentally depending on the actual distance of the economy from the technological frontier. More precisely, as already argued in Section 4, the model predicts α_3 should be negative for countries sufficiently far from the technological

frontier. By contrast, for more developed countries, the model predicts that the sign of α_3 is ambiguous and that this will depend on the relative efficiency of skilled human capital in innovation with respect to imitation: A positive sign is expected if this efficiency is relatively larger or a negative one otherwise.

To sum up the theoretical predictions presented above are as follows: (1) a positive value of the overall effect of tertiary human capital ($\alpha_2 + \alpha_3 z_{j,t-1}$) and either monotonically decreasing or U-shaped with respect to $z_{j,t-1}$ as we consider groups of increasingly richer countries; (2) a positive value of α_2 for the groups of countries closer to the frontier; and (3) a negative value of α_3 for *less* developed countries while an ambiguous (positive or negative) value of α_3 for developed countries depending on whether skilled workers are relatively more or less efficient in innovation activities.

5.4. Empirical Results

In order to empirically test the hypothesis on the development-specific impact of human capital composition on growth, we estimate the model in (21) on the whole sample of 85 countries as well as on different subsamples of countries grouped at different stages of development and hence compute the implied elasticities of growth w.r.t. tertiary education for the different subsamples. In columns 2 and 3 of Tables 3–6, we split the whole sample into high-income countries (21 OECD economies) and developing economies (64 economies) while in columns 4 to 7, we repeat the analysis by grouping countries belonging to the top 25% of the GDP distribution (representing the countries at the frontier) vs. those with a GDP level below 75, 50, and 25% of the sample (representing groups of increasingly less-developed countries).

First specification: fractions.   We start our analysis by proxying for skilled human capital through the fraction of workforce with tertiary education in each economy. Our theoretical model predicts a wide array of empirical results. Some of them, as we detailed before, crucially differ from previous literature and, we will show next, find robust confirmation in our empirical tests. Results are given in Table 3.

The empirical results showed in Table 3 strongly support the predictions of the model and confirm that the dynamics governing the impact of skilled labor on growth for the economies close to the technology frontier crucially differ from those arising, instead, at lower stages of development.

First, notice that coefficient associated to the share of tertiary educated workforce, α_2, is positive and strongly significant for the subsample of the OECD countries while negative and statistically significant for those economies farther away from the frontier (in columns 3 and 5–7). If, on the one hand, the positive coefficient α_2 is consistent with the empirical results found in VAM, on the other hand, the negative value for the developing countries fits with our theoretical generalization as long as also α_3 is estimated to be negative. Indeed, the latter

TABLE 3. Fractions

Dependent variable: TFP growth	(1) All	(2) OECD	(3) Developing	(4) > 25%	(5) < 75%	(6) < 50%	(7) < 25%
Proximity	−0.009	−0.064***	−0.004	−0.076***	−0.001	0.010	0.008
	(0.006)	(0.019)	(0.006)	(0.019)	(0.006)	(0.009)	(0.013)
Tertiary fraction	0.136*	0.109*	−0.327*	0.184**	−0.492***	−1.717***	−4.125*
	(0.076)	(0.056)	(0.192)	(0.072)	(0.182)	(0.584)	(2.241)
Proximity*tertiary fraction	0.036	0.235*	−0.238*	0.298**	−0.335**	−0.886***	−1.761**
	(0.085)	(0.123)	(0.136)	(0.121)	(0.136)	(0.259)	(0.797)
Constant	−0.011	−0.003	−0.003	−0.021	0.002	0.027	0.025
	(0.014)	(0.010)	(0.015)	(0.014)	(0.014)	(0.026)	(0.042)
Observations	308	83	225	100	208	144	86
Number of IDs	85	21	64	26	59	41	24
Hansen stat. p-value	0.551	0.145	0.122	0.559	0.232	0.233	0.554
Hansen stat.	19.53	9.545	28.66	19.41	25.36	25.34	19.48
H-test excluding group	11.81	0.382	0.258	0.403	0.399	0.574	0.383
p-value	0.461	0.09	0.124	0.652	0.174	0.095	0.672
Number of instruments	28	13	28	28	28	28	28
AR(2) p-value	0.521	0.699	0.804	0.621	0.502	0.521	0.284
AR(2) stat.	0.642	0.387	0.248	−0.494	0.671	0.642	1.072
Implied total effect of S	0.08	0.01	0.12	0.04	0.17	0.35	0.86

Robust standard errors in brackets ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

TABLE 4. Fractions with time-invariant institutional controls

Dependent variable: TFP growth	(1) All	(2) OECD	(3) Developing	(4) > 25%	(5) < 75%	(6) < 50%	(7) < 25%
Proximity	−0,004	−0.067***	−0,003	−0.071***	0,001	0,01	0.029*
	(0.006)	(0.016)	(0.007)	(0.020)	(0.006)	(0.007)	(0.015)
Tertiary fraction	0,066	0.096*	−0.379*	0.178**	−0.594***	−1.781***	−6.509**
	(0.076)	(0.052)	(0.213)	(0.072)	(0.196)	(0.537)	(2.407)
Proximitytertiary fraction	−0,041	0.247**	−0.321**	0.294**	−0.411***	−0.926***	−2.575***
	(0.086)	(0.107)	(0.142)	(0.135)	(0.143)	(0.253)	(0.849)
Constant	−0,003	−0,001	−0,004	−0,019	0,006	0,026	0.088*
	(0.013)	(0.010)	(0.016)	(0.014)	(0.014)	(0.021)	(0.043)
Observations	308	83	225	100	208	144	86
Number of IDs	85	21	64	26	59	41	24
Hansen stat. p-value	0.0338	0.376	0.0762	0.666	0.17	0.267	0.513
Hansen stat.	35.56	9.696	34.5	13,1	29.32	26.75	7.222
H-test excluding group	0.176	0.283	0.048	0.429	0.077	0.53	0.402
p-value	0.035	0.519	0.407	0.731	0.61	0.129	0.55
Number of instruments	29	16	31	23	30	30	15
AR(2) p-value	0.509	0.793	0.814	0.733	0.487	0.504	0.278
AR(2) stat.	0.66	0.262	0.235	−0.342	0.694	0.668	1.085
Implied total effect of S	0.13	0	0.23	0.04	0.22	0.38	0.77

Robust standard errors in brackets ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

TABLE 5. Years

Dependent variable: TFP growth rate	(1) All	(2) OECD	(3) Developing	(4) >25%	(5) <75%	(6) <50%	(7) <25%
Proximity	−0.013	−0.077	0.028	−0.026	0.046	0.041	0.001
	(0.026)	(0.060)	(0.030)	(0.057)	(0.032)	(0.029)	(0.019)
L.yearsT	0.042*	0.042*	−0.149*	0.038*	−0.211**	−0.360**	−0.512
	(0.022)	(0.022)	(0.085)	(0.019)	(0.088)	(0.154)	(0.612)
L.yearsPS	−0.000	0.004	−0.014	−0.001	−0.024*	−0.024	−0.019
	(0.007)	(0.005)	(0.012)	(0.006)	(0.014)	(0.016)	(0.012)
ProximityT	0.005	0.086**	−0.103*	0.056*	−0.149**	−0.229***	−0.312
	(0.023)	(0.035)	(0.057)	(0.030)	(0.068)	(0.078)	(0.217)
ProximityPS	0.001	−0.001	−0.004	−0.005	−0.007*	−0.006	−0.002
	(0.003)	(0.006)	(0.003)	(0.006)	(0.004)	(0.004)	(0.002)
Constant	−0.014	−0.04	0.082	−0.008	0.129	0.122	0.024
	(0.057)	(0.051)	(0.080)	(0.051)	(0.086)	(0.089)	(0.060)
Observations	308	83	225	100	208	144	86
Number of IDs	85	21	64	26	59	41	24
Hansen stat. p-value	0.181	0.908	0.177	0.984	0.351	0.918	0.222
Hansen stat.	44.67	13.82	44.81	21.02	39.70	25.73	9.450
H-test excluding group	0.566	0.661	0.275	0.828	0.249	0.756	0,852
p-value	0.059	0.886	0.199	0.989	0.554	0.898	0,99
Number of instruments	46	31	46	46	46	46	16
AR(2) p-value	0.523	0.834	0.479	0.572	0.326	0.495	0.954
AR(2) stat.	0.639	−0.209	0.707	−0.566	0.982	0.683	0.0578
Implied total effect of S	0.03	0.01	0.05	0.01	0.08	0.17	0.37
Implied total effect of U	0	0	−0.01	0	−0.01	−0.01	−0.01

Robust standard errors in brackets ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

TABLE 6. Years with time-invariant institutional controls

Dependent variable: TFP growth rate	(1) All	(2) OECD	(3) Developing	(4) >25%	(5) <75%	(6) <50%	(7) <25%
Proximity	−0.008	−0.023	0.002	0.004	0.013	0.019	−0.005
	(0.011)	(0.049)	(0.013)	(0.060)	(0.016)	(0.020)	(0.031)
L.yearsT	0.031*	0.026	−0.092	0.031	−0.146**	−0.300**	0.049
	(0.016)	(0.017)	(0.058)	(0.018)	(0.055)	(0.141)	(0.069)
L.yearsPS	−0.000	−0.001	−0.004	−0.003	−0.012	−0.016	−0.014
	(0.004)	(0.004)	(0.007)	(0.006)	(0.009)	(0.013)	(0.009)
ProximityT	−0.007	0.061**	−0.086**	0.043	−0.119**	−0.206***	0.080
	(0.021)	(0.027)	(0.043)	(0.027)	(0.051)	(0.069)	(0.055)
ProximityPS	0.000	−0.006	−0.001	−0.008	−0.003	−0.003	−0.012*
	(0.001)	(0.005)	(0.001)	(0.007)	(0.002)	(0.003)	(0.007)
Constant	−0.009	0.006	0.011	0.013	0.044	0.061	−0.005
	(0.026)	(0.043)	(0.038)	(0.054)	(0.045)	(0.061)	(0.043)
Observations	308	83	225	100	208	144	118
Number of IDs	85	21	64	26	59	41	34
Hansen stat. p-value	0.0912	0.905	0.239	0.989	0.285	0.810	0.392
Hansen stat.	53.51	16.33	45.96	23.12	43.53	31.17	25.25
H-test excluding group	0.199	0.889	0.169	0.987	0.274	0.697	0.998
p-value	0.05	0.575	0.839	0.633	0.396	0.99	0.99
Number of instruments	50	34	49	50	48	48	33
AR(2) p-value	0.544	0.975	0.742	0.594	0.439	0.500	0.725
AR(2) stat.	0.606	0.0310	0.329	−0.533	0.774	0.674	−0.352
Implied total effect of S	0.04	0	0.07	0.01	0.09	0.18	0.37
Implied total effect of U	0	0	0	0	−0.01	−0.01	−0.01

Robust standard errors in brackets ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

is strongly significant for all subsamples and shows opposite signs for the sub-sample of OECD and that of developing countries (resp. positive and negative) as expected. Hence our empirical results also show that the growth impact of tertiary educated labor increases with the proximity to the technological frontier for those group of countries sufficiently close to the technology frontier. The results for the OECD countries are in fact, qualitatively the same as those proposed by VAM. This said, however, our empirical analysis claims that for the subsample of lagging economies, the effect of tertiary education increases *as we move far away* from the frontier, in contrast to the predictions of previous literature.

Finally, the overall effect of tertiary education on economic growth $\alpha_2 + \alpha_3 z_{j,t-1}$[30] (presented at the bottom of the table) is consistent with our predictions being *positive and significant* for the all the subsample considered. Interestingly, we observe that the magnitude by which a marginal increase in tertiary education affects growth is very much heterogeneous across countries at different stages of development. For the OECD sample, the estimated average value of $\alpha_2 + \alpha_3 z_{j,t-1}$ is of 0.01 while that for developing countries is of 0.12. The relative larger overall impact of tertiary education on the growth of developing vis à vis developed economies is robust to different samplings. The implied average overall effect of tertiary educated workers on growth for countries at the top 25% of the GDP distribution is of 0.04 while that for increasingly lower development stages (countries below the second, third, and fourth quartile of GDP in columns 4–7) show increasingly larger impacts as of 0.17, 0.35, and 0.86, respectively. This result confirms the predictions of the model that excludes the case in which the growth impact of skilled human capital is monotonically increasing with the proximity to the technological frontier throughout the sample.

As for the robustness of our econometric specification, tests are all passed. The Hansen overidentification tests reports the acceptance of the null of instruments exogeneity for all the specifications proposed in Table 3 suggesting that the model is correctly specified. A similar result is obtained by the difference-in-Hansen.[31] Interestingly, the recent contribution by Ang et al. (2011), uses a similar empirical approach to ours in order to estimate the impact of different educational level on economic growth while, however, finding somehow different results.[32] It is worth noticing, however, that their Hansen p-values are almost always suspiciously high and close to unity (as of 0.99) and that the authors do not report the instrumental count. As extensively argued in recent empirical literature the use of an excessive number of instruments can cause the p-value of the Hansen test to get close to unity and lead to the incorrect acceptance the null of instruments exogeneity. We carefully check that the instrumental set in our estimates does not over-fit the endogenous variables as suggested by Roodman (2009a). The AR(2) test, checking that the error terms in the first-differenced regression exhibit no second-order serial correlation is also passed by all the specifications proposed in Table 3.

As a robustness check of the results, we introduce time-invariant institutional controls into the empirical specification in Table 4. As pointed out by Roodman (2009b), "In system GMM, one can include time-invariant regressors, which would

disappear in difference GMM. Asymptotically, this does not affect the coefficient estimates for other regressors because all instruments for the levels equation are assumed to be orthogonal to fixed effects, indeed to all time-invariant variables. In expectation, removing them from the error term does not affect the moments that are the basis for identification" (p. 30). These controls do not appear in the table since they are treated as standard instruments in the SYSGMM estimation and for which one column for each variable is built in the instrument matrix. The results of such a robustness checks are presented in Table 4. The additional exogenous country-specific institutional variables are the legal origin variables proposed by la Porta et al. (2008), where a country legal origin ranges from English to Socialist.

Our results are robust after controlling for legal origin while the differences in the implied total effect of skilled workers on growth slightly increases.[33]

If any, our empirical analysis implicitly supports (1) the assumption of nonincreasing returns to technological activities (and strictly against that of constant returns, as in VAM), and (2) the case according to which skilled workers' comparative advantage in innovation is strong enough. The fact that α_3 is positive and significant for sufficiently rich countries while is negative and significant for developing countries is indeed the exact empirical translation of case 2 in Proposition 2 according to which, when returns to innovation are strictly decreasing, when returns to imitation are nonincreasing and when the relative efficiency of skilled workers in innovation with respect to imitation is large enough, the cross derivative $\frac{\partial^2 g}{\partial a \partial S}$ is positive for rich enough countries and negative otherwise.

There are several reasons to believe that this scenario is a sensible one. Previous empirical and theoretical literature already argued (and our work adds to these contributions) that innovation activities would encounter diminishing returns in their inputs. See, for instance, Griliches (1990), Kortum (1993), Jones (1995a, 1995b), Davidson and Segestrom (1998), and Segestrom (1998) for whom a sustained growth in TFP can be only obtained by increasing growth in research and development (R&D) inputs. Similarly, as for the efficiency of skilled workers in innovation activities, this is actually the same scenario employed by VAM, which, however, with DRS has very different implications.

These empirical results are confirmed by the analysis by using years of schooling as proposed below.

Second specification: years   We now move to a specification where the stock of skilled and unskilled labor can vary independently. For this, we calculate the average years of schooling of tertiary educated labor and that of secondary- and primary-educated people in each country. We build the indicators for the average number of years of schooling in the two categories as follows:

$$\text{Years}T \equiv pT \times nT,$$
$$\text{Years}PS \equiv pP \times nP + pS \times nS,$$

where pT, pS, and pP are the fractions of population having achieved tertiary, secondary, and primary education, respectively, while nT, nS, and nP are the

number of extra years of education that an individual has accumulated over the preceding level. Empirical results are presented in Table 5.

Our estimates suggest again the crucial role of tertiary education for economic growth. This said, the results confirm the increasing importance of tertiary education for countries increasingly farther away from the frontier. The implied total effect of skilled workers (this time proxied by the average number of years of tertiary education in each country) is shown to increase at lower development stages as predicted by our theoretical model. The elasticity of TFP growth associated to an increase in tertiary education in the OECD countries is estimated to be of around 0.01 vis a vis 0.05 for the developing countries subsample. Similarly, when we disaggregate the whole sample and compare the 25% top part of the GDP distribution with that of increasingly poorer countries (below the 75, 50, and 25% of the sample distribution), the estimated total effect of tertiary education goes from 0.01 to 0.37 for the subsample of poorest countries.

The effect of primary and secondary education seems instead to be either nonsignificant or close to zero. The coefficients associated to the secondary and primary average years of schooling, in fact, do not reach statistical significance in almost all the specification proposed. Similar results are obtained when (Table 6) we control for institutional quality differences through legal origin time-invariant characteristics.

Our estimates are again robust to a wide array of robustness checks on the quality of the instrumental set (the Hansen and difference-in-Hansen test) as well to the AR(2) test of second-order serial correlation in the errors.

6. CONCLUSIONS

Anecdotic evidence frequently reports how increasing the share of high-skilled and trained workers could play an extremely important role in the absorption of technology in poor countries and alleviate poverty.

Our study proposes a rationale for this view and provides compelling and robust evidence regarding the heterogeneous impact of human capital composition on the growth of countries at different stages of development. In contrast to earlier theoretical and empirical literature that argued for the "primacy" of high skills at higher stages of development (when countries are closer to the technology frontier and perform technology innovation), our work shows—both theoretically and empirically—that tertiary education is especially important for the growth of those countries that are lagging behind and far away from the technology frontier. By contrast, its relative impact on middle-income economies appears to be substantially weaker, becoming again more important for developed economies.

We contribute to the existing literature in a number of ways. First, we generalize the theoretical settings proposed by Vandenbussche et al. (2006) by encompassing the case of decreasing returns to scale in the production of technological improvements. This generalization is crucial to unveil a distinctively more complex dynamics linking tertiary education to economic growth of economies found at

very different stages of development while leaving the case of CRS, analyzed by VAM, as a special one.

Unlike previous literature and under less restrictive assumptions, our model shows that the marginal effect of an increase in skilled workers for least developed countries is growth enhancing the more the economies are found farther away from the frontier. Even if so, for those close to the technology frontier, our model provides results that are qualitatively similar to those proposed in the literature and analyzed by VAM.

These theoretical results are supported by empirical investigation. We estimated the empirical model proposed by VAM-addressing endogeneity between educational variables and economic growth through System GMM estimators for a 10-years intervals dynamic panel 85 countries (developed and developing) in between the year 1960 and 2000. Our empirical results, while confirming VAM's results for the subset of OECD countries, show the increasingly larger effect of tertiary education on the growth of lagging economies as consistently predicted by our theoretical model and in contrast to previous theoretical and empirical literature.

All in all, our results point to the importance of tertiary education in the explanation of growth while, at the same time, showing that its effect on growth is heterogeneous across countries found at different stages of development. Our results suggest the relatively more important role of tertiary education for the growth of countries for which, instead, the primacy of lower educational levels has been usually advocated as main engine of growth and development.

NOTES

1. The work by Krueger and Lindhal (2001), Benhabib and Spiegel (1994), and Temple (2001) are among those supporting this puzzling evidence and arguing that the role of human capital on economic growth might have been quite overstated.

2. This hypothesis is based on the assumption that different types of human capital (resp. skilled vs. unskilled workers) perform different tasks (resp. innovation vs. imitation) depending on the relative distance of the economy to the technology frontier (resp. when close or far away from the technological leader).

3. "It seems safe to say that, if our model is right, the graduate education that occurs in research universities should be most growth-enhancing in states that are close to the technological frontier" [Aghion et al. (2009, pp. 2–3)].

4. The terms *adoption* and *imitation* are used interchangeably in this paper.

5. Empirical results on this issue are surveyed and collected by Psacharopoulos and Patrinos (2004) according to which social returns to higher education in low-income countries are 11.2% versus 11.3% for middle-income countries and only 9.5% in high-income countries. Differences are even more striking if we consider private returns on tertiary education: 26% in low-income countries, 19.3% in middle-income countries, and only 12.4% in high-income countries.

6. In particular, Mansfield et al. (1981) point out how, over 48 different products in chemical, drug, electronics, and machinery U.S. industries, the costs of imitation lied between 40% and 90% of the costs of innovation.

7. Squicciarini and Voigtländer (2015) argue that "France, in its role as a follower country, initially depended largely on the adoption of British technology" (p. 8) and that an interest in science by the

knowledge elites "helped entrepreneurs both to learn about these techniques in the first place, and to understand the underlying principles needed to implement and run them" (p. 2).

8. There is extensive literature focusing on decreasing returns to R&D activities (which includes both innovation and imitation activities). From the empirical perspective, Griliches (1990) and Kortum (1993) note that, in the United States, the ratio of the number of patent applications to the scientists and engineers involved in R&D has fallen over time in the post-war period, while Jones (1995a, 1995b) points out that the economy growth rates have remained constant and even declined despite an increase in the amount of R&D effort. The assumption of diminishing returns to R&D activities has then been used in many important theoretical papers aiming at proposing a theoretical solution to the observed absence of relation between the scale of R&D effort and growth. Those theoretical papers include Jones (1995b), Kortum (1997), Segestrom (1998), Young (1998), Davidson and Segestrom (1998), and Cheng and Tao (1999). In the last two papers, in particular, R&D activity includes both imitation and innovation and both activities are assumed to have decreasing returns, just like in our paper.

9. Which has been recently supported empirically by Papakonstantinou (2014).

10. Aghion et al. (2009) introduces the possibility of migration for skilled workers.

11. See, for instance, Arrow (1962), Mansfield, Schwartz, and Wagner (1981), Barro and Sala-i-Martin (1997), and Benhabib and Spiegel (1994, 2005).

12. When $\beta + \sigma > \theta + \phi$, a case that is excluded by the CRS case, imitation can be considered to be an "easier" activity in the sense that, following an equal percentage change in each production factor, the induced percentage change in the contribution by imitation activities will be larger than the percentage change in the contribution by innovation activities. Formally, it is easy to see that with Cobb–Douglas specification, when $\frac{\partial u_n}{u_n} = \frac{\partial u_m}{u_m} = \frac{\partial s_n}{s_n} = \frac{\partial s_m}{s_m}$ and total differentiating m and n, we have that $\frac{\partial m}{m} > (<) \frac{\partial n}{n}$ implies $\sigma + \beta > (<) \phi + \theta$.

13. Notice that

$$h'(a) = -\frac{1}{\sigma - \phi} \left(\frac{\beta \psi^\sigma}{\gamma \theta} \frac{1-a}{a} \right)^{\frac{1}{\sigma - \phi} - 1} \left(\frac{\beta \psi^\sigma}{\gamma \theta} \frac{1}{a^2} \right) < 0,$$

so that the negativity of $h'(a)$ is not affected by nonconstant returns to scale in imitation and innovation but it only depends on the assumption $\sigma > \phi$.

14. Clearly enough, it looks reasonable to assume that unskilled workers cannot outperform skilled workers in both technological activity and therefore $\beta > \sigma$ and $\theta > \phi$. However, our results are completely independent from this assumption. In other words, the dynamics of catch-up are governed only by comparative advantages (i.e., relative efficiencies) and not by absolute advantages.

15. When returns are nonincreasing in innovation, the sign of $\frac{\partial s_m^*}{\partial U}$ and $\frac{\partial s_m^*}{\partial a}$ is, respectively, positive and negative at any distance to the technological frontier and for any parameter value, exactly as in the CRS case already analyzed by VAM.

16. Proofs of the signs of each derivative for any relevant value of a and $\frac{\theta}{\beta}$ are provided in Appendix A.

17. We emphasize that under CRS ($\theta = 1 - \phi$), the above expression boils down to $g/\lambda \gamma = h(a)^{-\phi} (1 - \phi) S + h(a)^{1-\phi} \phi U$, which is the same as in VAM 2006.

18. Interestingly, decreasing returns to scale in innovation do not affect the impact of unskilled human capital on growth, which is positive and identical to the CRS case, i.e., $\frac{\partial g}{\partial U}/\lambda = \phi \gamma h(a)^{1-\phi}$.

19. The value of the derivatives $\frac{\partial u_j}{\partial S}$ and $\frac{\partial s_j}{\partial S}$ for $j = m, n$ are computed in Appendix A.

20. Computations are available at request.

21. It is easy to see that the results stated in Proposition 2 still hold when returns to imitation are kept constant. In this case, the threshold is simply $\theta > (\leq) 1 - \sigma$, which has a clear economic interpretation: Skilled workers are more (less) efficient in imitation than in imitation.

22. A recent and still unpublished paper by Papakostantinou (2014) obtains a similar finding.

23. We choose this specification because the empirical model tested has to be fully consistent with the one used by VAM in order to allow the comparison of our results, run on a much larger set of countries, with those in VAM, run only on countries at the frontier.

24. The TFP of the leader (at the frontier) is denoted by $\bar{A}$.

25. This happens since the lagged value $A_{j,t-1}$ enters within $a_{j,t-1}$ as a regressor for the growth rate of TFP.

26. See Aghion et al. (2009): "Instrumenting with lagged spending does not overcome biases caused by omitted variables such as institutions" (p. 5).

27. Our analysis focuses on the long-run trends in the effect of human capital composition on economic growth and it abstracts from the turmoil associated to the great recession. Cyclical factors may well be affecting our estimates and, for this reason, time dummies are always used to address this potential bias. Given the deep economic downturn experienced by many countries during the recent financial economic crisis started in 2007, the sample used in the current analysis covers only the long period in between 1960 and 2000, so as to be able to compare our empirical results to those obtained in VAM.

28. The choice of the United States, the technological leader, is common in this kind of literature. We follow this approach in order to ensure the comparability of our results with previous studies. The United States is, however, the leader in our own TFP estimates as well.

29. The statistics referring to the OECD subsample are fully in line with those presented by VAM both for the TFP and human capital measures.

30. Notice that when j does not refer to a country but to a *group* of countries, then $z_{j,t-1}$ is computed as the arithmetic mean of the variable z for all the countries k belonging to group j: $z_j = \frac{1}{N_j} \sum_i^{N_j} z_k$, where $k = 1, 2, \ldots, N_j$.

31. The difference in Hansen test also points to the exogeneity of the instrument subsets with the null hypothesis that the subsets of instruments are exogenous. See Roodman (2009b), for more details on this.

32. The authors analyze the effect of tertiary education on the growth of countries at different stages of development. However, differently from us, they find a positive effect of tertiary education only at middle and higher stages of development. Part of this result, as we argue above, it may be caused by an incorrect specification of the lag structure in their System GMM estimation.

33. As suggested by one of the referee, we run several additional robustness test controlling for (i) the size of government expenditures (e.g., taxes, government spending and transfers, and subsidies as a percentage of GDP), (ii) the access to sound money (e.g., inflation and money growth as well as freedom to own foreign currency bank accounts), or (iii) market and trade openness (e.g., taxes on international trade, trade barriers, nontariff trade barriers). The results of these additional robustness checks, which are not significantly different from those obtained in the previous specifications, can be found in Appendix B.

REFERENCES

Acemoglu, D., P. Aghion, and F. Zilibotti (2006) Distance to frontier, selection and economic growth. *Journal of the European Economic Association* 4, 37–74.

Aghion, P., L. Boustan, C. Hoxby, and J. Vandenbussche (2009) The Causal Impact of Education on Economic Growth: Evidence from U.S. Brooking Papers on Economic Activity. Available at http://www.brookings.edu/economics/bpea/~/media/Files/Programs/ES/BPEA/2009_spring_bpea_papers/2009_spring_bpea_aghion_etal.pdf.

Andersen, T. G. and B. E. Soerensen (1996) GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics* 14, 328–352.

Ang, J., J. Madsen, and M. Islam (2011) The effects of human capital composition on technological convergence. *Journal of Macroeconomics* 33, 465–476.

Arellano, M. and S. Bond (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment Equations. *Review of Economic Studies* 58, 277–297.

Arellano, M. and O. Bover (1995) Another look at the instrumental variables estimation of error components models. *Journal of Econometrics* 68, 29–51.

Arrow, K. J. (1962) The economic implications of learning by doing. *Review of Economic Studies* 29, 155–173.

Baier, S. L., G. P. Dwyer, and R. Tamura (2006) How important are capital and total factor productivity for economic growth? *Economic Enquiry* 441, 23–49.

Barro, R. J. and Xavier Sala-i-Martin (1997) Technological diffusion, convergence, and growth. *Journal of Economic Growth* 2, 1–26.

Benhabib, J. and M. Spiegel (1994) The role of human capital in economic development evidence from aggregate cross-country data. *Journal of Monetary Economics* 34, 143–173.

Benhabib, J. and M. Spiegel (2005) Human capital and technology diffusion. In P. Aghion and S. Durlauf (eds.), *Handbook of Economic Growth*, pp. 935–966. Amsterdam: Elsevier.

Bils, M. and P. J. Klenow (2000) Does schooling cause growth? *American Economic Review* 90, 1160–83.

Blundell, R. and S. Bond (1998) Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115–143.

Bond, S. (2002) Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice. Working paper 09/02, Institute for Fiscal Studies.

Bowsher, C. G. (2002) On testing overidentifying restrictions in dynamic panel data models. *Economics Letters* 77, 211–220.

Chen, L. K. and Z. Tao (1999) The impact of public policies on innovation and imitation: The role of R&D technology in growth models. *International Economic Review* 40, 187–207.

Coe, D. and E. Helpman (1995) International R&D spillovers. *European Economic Review* 39, 859–887.

Cohen, D. and M. Soto (2007) Growth and human capital: Good data, good results. *Journal of Economic Growth* 12, 51–76.

Cohn, E. and J. T. Addison (1998) The economics returns to lifelong learning in OECD countries. *Education Economics* 6, 253–30.

Davidson, C. and P. Segestrom (1998) R&D subsidies and economic growth. *Rand Journal of Economics* 29, 548–577.

Griliches, Z. (1990) Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 28, 1661–1707.

Kiviet, J. F. (1995) On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68, 53–78.

Kortum, S. (1993) Equilibrium R&D and the decline in the patent-R&D ratio. *American Economic Review: Papers and Proceedings* 83, 450–457.

Kortum, S. (1997) Research, patenting, and technological change. *Econometrica* 65, 1389–1419.

Krueger, A. and M. Lindahl (2001) Education for growth: Why and for whom? *Journal of Economic Literature* 39, 1101–1136.

Ichino, A. and R. Winter-Ebmer (1999) Lower and upper bounds of returns to schooling: An exercise in IV estimation with different instruments. *European Economic Review* 43, 889–901.

Jones, C. (1995a) Time series tests of endogenous growth models. *Quarterly Journal of Economics* 110, 495–525.

Jones, C. (1995b) R&D-based models of economic growth. *Journal of Political Economy* 103, 759–784.

La Porta, R., F. Lopez-de-Silanes, and A. Shleifer (2008) The economic consequences of legal origins. *Journal of Economic Literature* 46, 285–332.

Mansfield, E., M. Schwartz, and S. Wagner (1981) Imitation costs and patents: An empirical study. *The Economic Journal* 91, 907–918.

Mitchell, B. R. (1998) *International Historical Statistics: Europe, The Americas, Africa, Asia and Australia 1750–1992*. New York: Stockton Press.

Nickell, S. (1981) Biases in dynamic models with fixed effects. *Econometrica* 49, 1417–1426.

Papakonstantinou, M. (2014) Composition of Human Capital, Distance to the Frontier and Productivity. IARIW papers. Available at http://www.iariw.org/papers/2014/PapakonstantinouPaper.pdf.

Psacharopoulos, G. (1994) Returns to investment in education: A global update. *World Development* 22, 1325–43.

Psacharopoulos, G. and H. Patrinos (2004) Returns to investment in education: A further update. *Education Economics* 12, 111–134.

Romer, P. (1990) Endogenous technological change. *Journal of Political Economy* 98, S71–S102.

Roodman, D. (2009a) A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71, 135–158.

Roodman, D. (2009b) How to do xtabond2: An introduction to difference and system GMM in stata. *Stata Journal* 9, 86–136.

Segerstrom, P. S. (1998) Endogenous growth without scale effects. *American Economic Review* 88, 1290–1310.

Squicciarini, M. and N. Voigtlaender (2015) Human capital and industrialization: Evidence from the age of enlightenment. *Quarterly Journal of Economics*, forthcoming.

Temple, J. (2001) Growth effects of education and social capital in the OECD countries. *OECD Economic Studies* 2, 57–101.

Vandenbussche, J., P. Aghion, and C. Meghir (2006) Growth, distance to frontier and composition of human capital. *Journal of Economic Growth* 11, 97–127.

Young, A. (1998) Growth without scale effects. *Journal of Political Economy* 106, 41–63.

APPENDIX A: PROOFS

A.1. PROOF OF PROPOSITION 1

We first prove existence and uniqueness by focusing only on the equilibrium value of s_m, and then we derive the equilibrium values of u_m, s_n, and u_n by using (13), (8), and (9). Now consider (14) and define the function $k(s_m, S, U, a)$ as follows:

$$k(s_m, S, U, a) = h(a) U - (S - (\psi - 1)s_m) \left(\frac{s_m^{1-\beta-\sigma}}{(S - s_m)^{1-\theta-\phi}} \right)^{\frac{1}{\sigma-\phi}},$$

where

$$h(a) = \left(\frac{\beta \psi^\sigma}{\gamma \theta} \frac{1-a}{a} \right)^{\frac{1}{\sigma-\phi}},$$

$$\psi = \frac{\sigma\theta}{\beta\phi}.$$

Since with nonincreasing returns in technological activities the second-order conditions satisfy the convexity requirement, a necessary and sufficient condition for a value $s_m^* = s_m$ to be an equilibrium is that $k(s_m^*, S, U, a) = 0$. Now, we know that the equilibrium value of s_m should be such that $s_m^* \in [0, S]$. Computing the limit of $k(s_m, S, U, a)$ for $s_m^* = 0$ and for $s_m^* = S$ we find, by Assumption 1,

$$\lim_{s_m \to 0} k(s_m, S, U, a) = h(a) U > 0,$$

$$\lim_{s_m \to S} k(s_m, S, U, a) = -\infty < 0.$$

So that existence is proved as there is, by continuity of $k(s_m, S, U, a)$ for $s_m \in (0, S)$, at least one value of $s_m = s_m^* \in (0, 1)$ such that $k(s_m^*, S, U, a) = 0$.

To prove that this equilibrium value is unique, compute the partial derivative of $k(s_m, S, U, a)$ with respect to s_m to find

$$\frac{\partial k(s_m, S, U, a)}{\partial s_m} = -(S - s_m)^{-\frac{1-\theta-\phi}{\sigma-\phi}} s_m^{\frac{1-\beta-\sigma}{\sigma-\phi}}$$

$$\times \left[(\psi - 1) s_m + \frac{1-\theta-\phi}{\sigma-\phi} \frac{(S + (\psi-1) s_m)}{(S - s_m)} + \frac{1-\beta-\sigma}{\sigma-\phi} \frac{(S + (\psi-1) s_m)}{s_m} \right] < 0$$

so that, by Assumption 1, $k(s_m, S, U, a)$ is monotonically decreasing in s_m and therefore, for any $a \in (0, 1)$, there is only one value of s_m^* such that $k(s_m, S, U, a) = 0$.

The equilibrium values of u_m, s_n, and u_n are easily and univocally derived as functions of s_m^* by using (13), (8), and (9). So, we can write

$$\{u_n^*, u_m^*, s_n^*, s_m^*\} = \left\{ \frac{U(S - s_m^*)}{S + (\psi-1) s_m^*}, \frac{\psi s_m^* U}{S + (\psi-1) s_m^*}, S - s_m^*, s_m^* \right\} \in (0, U)^2$$

$$\times (0, S)^2 \text{ for any } a \in (0, 1).$$

Finally, notice that

$$\lim_{a \to 0} (u_n^*, u_m^*, s_n^*, s_m^*) = (0, U, 0, S),$$
$$\lim_{a \to 1} (u_n^*, u_m^*, s_n^*, s_m^*) = (U, 0, S, 0).$$

A.2. PROOF OF TABLE 1

We start by computing $\frac{\partial s_m^*}{\partial S}$ by using the implicit function theorem and then we use (13), (8), and (9), to compute $\frac{\partial u_m^*}{\partial S}$, $\frac{\partial u_n^*}{\partial S}$, and $\frac{\partial s_n^*}{\partial S}$. Finally, by using derivative of s_m^* with respect to S and then, by using (4) and (5), we compute the derivatives $\frac{\partial m}{\partial S}$ and $\frac{\partial n}{\partial S}$.

The sign of $\frac{\partial s_m^*}{\partial S}$. Consider (14). At $s_m = s_m^*$, it must be

$$(S + (\psi - 1) s_m^*) (S - s_m^*)^{-\frac{1-\theta-\phi}{\sigma-\phi}} s_m^{*\frac{(1-\beta-\sigma)}{\sigma-\phi}} \equiv h(a) U$$

for any $a \in (0, 1)$. Differentiating both sides with respect to S and dividing by $s_m^{*\frac{(1-\beta-\sigma)}{\sigma-\phi} -1} (S - s_m^*)^{-\frac{1-\theta-\phi}{\sigma-\phi}-1}$, we find

$$\left(1 + (\psi - 1) \frac{\partial s_m^*}{\partial S}\right) (S - s_m^*) s_m^* - \frac{1-\theta-\phi}{\sigma-\phi} (S + (\psi-1) s_m^*) s_m^* \left(1 - \frac{\partial s_m^*}{\partial S}\right)$$

$$+ \frac{(1-\beta-\sigma)}{\sigma-\phi} (S + (\psi-1) s_m^*) (S - s_m^*) \frac{\partial s_m^*}{\partial S} = 0,$$

which yields

$$\frac{\partial s_m^*}{\partial S} = \frac{\left[(S - s_m^*)(1 - \theta - \sigma) + (1 - \theta - \phi) s_m^* \psi\right] s_m^*}{(\psi - 1)(\sigma - \phi)(S - s_m^*) s_m^* + \left((\beta + \sigma - \theta - \phi) s_m^* + (1 - \beta - \sigma) S\right)(S + (\psi - 1) s_m^*)}.$$

$$\text{(A.1)}$$

Let us study the sign of this derivative in two different scenarios: (1) $\frac{\theta}{\beta} \in (\frac{\phi}{\sigma}, \frac{1-\sigma}{\beta}]$; and (2) $\frac{\theta}{\beta} \in [\frac{1-\sigma}{\beta}, \frac{1-\phi}{\beta})$.

(1) $\frac{\theta}{\beta} \in (\frac{\phi}{\sigma}, \frac{1-\sigma}{\beta}]$

When $\frac{\theta}{\beta} < \frac{1-\sigma}{\beta}$ and then $\theta \leq 1 - \sigma$, both the numerator and the denominator are clearly positive. The numerator since $1 - \theta - \phi > 0$ by Assumption 1 and $S > s_m^*$ by definition. The denominator is positive as we can see by writing it as

$$(\psi - 1)(\sigma - \phi)(S - s_m^*)s_m^* + ((\beta + \sigma - \theta - \phi)s_m^*$$
$$+ (1 - \beta - \sigma)S)(S + (\psi - 1)s_m^*)$$
$$= (\psi - 1)(\sigma - \phi)(S - s_m^*)s_m^* + (1 - \theta - \phi)(S + (\psi - 1)s_m^*)s_m^*$$
$$+ (1 - \beta - \sigma)(S + (\psi - 1)s_m^*)(S - s_m^*).$$

Hence, $\frac{\partial s_m^*}{\partial S}$ is strictly positive for any $a \in (0, 1)$.

(2) $\frac{\theta}{\beta} \in [\frac{1-\sigma}{\beta}, \frac{1-\phi}{\beta})$

Again, the denominator is clearly positive for the same reason as above. As a consequence, the sign of $\frac{\partial s_m^*}{\partial S}$ only depends on the sign of the numerator, which is a priori ambiguous. More precisely

$$\frac{\partial s_m^*}{\partial S} > (\leq) 0 \Leftrightarrow s_m^* > (\leq) S \frac{(\sigma + \theta - 1)}{(\sigma - \phi) + (\psi - 1)(1 - \theta - \phi)} \in (0, S).$$

Hence, when $\frac{\theta}{\beta} \in [\frac{1-\sigma}{\beta}, \frac{1-\phi}{\beta})$, and so when $\theta \in (1 - \sigma, 1 - \phi)$, there is always at an interior equilibrium value of s_m^* above which $\frac{\partial s_m^*}{\partial S}$ turns from positive to negative. To see how this equilibrium value changes with a, compute $\frac{\partial s_m^*}{\partial a}$ by differentiating both sides of (14) evaluated at $s_m = s_m^*$ to obtain

$$\frac{\partial s_m^*}{\partial a} =$$

$$\frac{(S - s_m^*)^{1 + \frac{1-\theta-\phi}{\sigma-\phi}} s_m^{*1 - \frac{(1-\beta-\sigma)}{\sigma-\phi}} h'(a) U(\sigma - \phi)}{((\sigma - \phi)(\psi - 1)(S - s_m^*)s_m^* + (1 - \theta - \phi)(S + (\psi - 1)s_m^*)s_m^* + (1 - \beta - \sigma)(S + (\psi - 1)s_m^*)(S - s_m^*))} < 0.$$

Hence, s_m^* is monotonically decreasing in a and then there is only one value of a, call it a such that $s_m^*(a^*) = S \frac{(\sigma+\theta-1)}{(\sigma-\phi)+(\psi-1)(1-\theta-\phi)}$ and then $\frac{\partial s_m^*}{\partial S}\big|_{a=a^*} = 0$. Therefore, we can write

$$\frac{\partial s_m^*}{\partial S} > (\leq) 0 \Leftrightarrow a < (\geq) a^* \in (0, 1).$$

Finally, note that by setting $\theta = 1 - \phi$ and $\beta = 1 - \sigma$, we find

$$\frac{\partial s_m^*}{\partial S} = -\frac{1}{\psi - 1} < 0, \forall a \in (a, \bar{a}).$$

The argument developed so far shows the directions of the arrows in row (1) of Table 1.

The sign of $\frac{\partial u_m^*}{\partial S}$.   To obtain the values of $\frac{\partial u_m^*}{\partial S}$, evaluate (13) at $s_m = s_m^*$ and differentiate both sides of (13) with respect to S to obtain

$$\frac{\partial u_m^*}{\partial S} = \psi U \frac{\frac{\partial s_m^*}{\partial S} S - s_m^*}{(S + (\psi - 1)s_m^*)^2}. \tag{A.2}$$

We then have

$$\frac{\partial u_m^*}{\partial S} > (\leq) \, 0 \Leftrightarrow \frac{\partial s_m^*}{\partial S} > (\leq) \, \frac{s_m^*}{S}.$$

Now, substitute for the value of $\frac{\partial s_m^*}{\partial S}$ to find

$$\frac{\partial u_m^*}{\partial S} > (\leq) \, 0$$

$$\Leftrightarrow \frac{\left((S - s_m^*)(1 - \theta - \sigma) + (1 - \theta - \phi) s_m^* \psi\right) s_m^*}{(\psi - 1)(\sigma - \phi)(S - s_m^*) s_m^* + \left((\beta + \sigma - \theta - \phi) s_m^* + (1 - \beta - \sigma) S\right)(S + (\psi - 1) s_m^*)} > (\leq) \, \frac{s_m^*}{S},$$

which, with some algebra, can be simplified as follows:

$$\frac{\partial u_m^*}{\partial S} > (\leq) \, 0 \Leftrightarrow \theta < (\geq) \, \beta, \, \forall a \in (0, 1).$$

Hence, for any $a \in (0, 1)$, u_m^* is monotonically increasing in S for any $\frac{\theta}{\beta} \in (\frac{\phi}{\sigma}, 1]$ and monotonically decreasing and for any $\frac{\theta}{\beta} \in (1, \frac{1-\phi}{\beta}]$. We have then proved the directions of the arrows in row (2) Table 1.

The sign of $\frac{\partial m}{\partial S}$.   As for the effect of S on the equilibrium imitation output $m(u_m^*, s_m^*)$, differentiating (4) with respect to S, we find

$$\frac{\partial m}{\partial S} / m = \sigma \frac{\frac{\partial u_m^*}{\partial S}}{u_m^*} + \beta \frac{\frac{\partial s_m^*}{\partial S}}{s_m^*}, \tag{A.3}$$

so that $\frac{\partial m}{\partial S}$ is certainly negative for any $a \in (0, 1)$ and for $\frac{\theta}{\beta} \in (\frac{\phi}{\sigma}, 1]$ (Scenario 1a) when both $\frac{\partial u_m^*}{\partial S}$ and $\frac{\partial s_m^*}{\partial S}$ are negative. A similar argument can be used for the case when $\theta \in (1 - \sigma, 1 - \phi)$ and $a \in (a^*, 1)$ when we know that $\frac{\partial u_m^*}{\partial S}$ and $\frac{\partial s_m^*}{\partial S}$ are both negative so that $\frac{\partial m}{\partial S}$ is negative too.

We then need to understand what happens to $\frac{\partial m}{\partial S}$ for any $a \in (0, 1)$ when $\frac{\theta}{\beta} \in (1, \frac{1-\sigma}{\beta}]$ and for $a \in (0, a^*)$ when $\theta \in (\frac{1-\sigma}{\beta}, \frac{1-\phi}{\beta}]$. Substitute for (A.2) in (A.3) we can write, after some algebra,

$$\frac{\partial m \left(u_m^*, s_m^*\right)}{\partial S} / m$$

$$= \frac{(\beta - \theta (\sigma + \beta))(S - s_m^*) + \beta (1 - \theta - \phi) \psi s_m^*}{(\psi - 1)(\sigma - \phi)(S - s_m^*) s_m^* + \left((\beta + \sigma - \theta - \phi) s_m^* + (1 - \beta - \sigma) S\right)(S + (\psi - 1) s_m^*)}$$

$$\tag{A.4}$$

so that

$$\frac{\partial m}{\partial S} > (\leq) \, 0 \Leftrightarrow s_m^* > (\leq) \, \hat{s}_m^* = S \frac{\theta (\sigma + \beta) - \beta}{(1 - \theta) \beta (\psi - 1)}.$$

We can then conclude that

$$\frac{\theta}{\beta} < \frac{1}{\beta + \sigma} \Rightarrow \frac{\partial m}{\partial S} > 0, \, \forall a \in (0, 1).$$

As for the sign of $\frac{\partial m}{\partial S}$ when $\frac{\theta}{\beta} \in (\frac{1}{\beta+\sigma}, \frac{1-\sigma}{\beta}]$, in this case we have that $\hat{s}_m^* = S \frac{\theta(\sigma+\beta)-\beta}{(1-\theta)\beta(\psi-1)} > 0$ and since s_m^* takes all values between S and 0 as a goes from 1 to 0, then there is an a, call it $\hat{a}$, such that $s_m^*(\hat{a}) = \hat{s}_m^* = S \frac{\theta(\sigma+\beta)-\beta}{(1-\theta)\beta(\psi-1)}$ and therefore

$$\left. \frac{\partial m}{\partial S} \right|_{a=\hat{a}} = 0$$

with

$$a < (>)\hat{a} \Leftrightarrow \frac{\partial m}{\partial S} > (<)0.$$

As for $\theta \in (1-\sigma, 1-\phi)$, we have that the value of $s_m^* = S \frac{(\sigma+\theta-1)}{(\sigma-\phi)+(\psi-1)(1-\theta-\phi)}$ such that $\frac{\partial s_m^*}{\partial S} = 0$ is smaller than the value of $s_m^* = S \frac{\theta(\sigma+\beta)-\beta}{(1-\theta)\beta(\psi-1)}$ because, doing some algebras, we yield

$$S \frac{(\sigma+\theta-1)}{(\sigma-\phi)+(\psi-1)(1-\theta-\phi)} < S \frac{\theta(\sigma+\beta)-\beta}{(1-\theta)\beta(\psi-1)} \Leftrightarrow \theta > \beta,$$

which is always true by definition when $\theta \in (1-\sigma, 1-\phi)$. As a consequence, since $\frac{\partial s_m^*}{\partial a} < 0 \; \forall a \in (0, 1)$, we conclude that

$$\hat{a} < a^*, \text{ for } \theta \in (1-\sigma, 1-\phi).$$

We can then completely characterize the sign of $\frac{\partial m}{\partial S}$ for every $a \in (0, 1)$ and for every feasible value of θ:

$$\frac{\theta}{\beta} \in \left(\frac{\phi}{\sigma}, \frac{1}{\beta+\sigma}\right] \Rightarrow \frac{\partial m}{\partial S} > 0, \forall a \in (0, 1)$$

$$\frac{\theta}{\beta} \in \left(\frac{1}{\beta+\sigma}, \frac{1-\phi}{\beta}\right) \Rightarrow \begin{cases} \frac{\partial m}{\partial S} \geq 0, \forall a \in (0, \hat{a}] \\ \frac{\partial m}{\partial S} < 0, \forall a \in (\hat{a}, 1]. \end{cases}$$

Finally, when $\theta = 1-\phi$ and $\beta = 1-\sigma$, the CRS case, from (A.4) we clearly have

$$\frac{\partial m}{\partial S} = -\frac{m}{(\psi-1)s_m^*} < 0, \forall a \in (\underline{a}, \overline{a}).$$

The argument developed so far shows the directions of the arrows in row (3) of Table 1.

The sign of $\frac{\partial s_n^*}{\partial S}$ and $\frac{\partial u_n^*}{\partial S}$. As for $\frac{\partial s_n^*}{\partial S}$, by (9) we have

$$\frac{\partial s_n^*}{\partial S} = 1 - \frac{\partial s_m^*}{\partial S},$$

which is always positive for any feasible value of $a \in (0, 1)$. That happens because, as it is straighforward from (A.1), $\frac{\partial s_m^*}{\partial S} < 1$ for any value of $a \in (0, 1)$. We have then shown the reason for the upward directions of all the arrows in row (4) of Table 1.

As for $\frac{\partial u_n^*}{\partial S}$, by (8) we have $\frac{\partial u_n^*}{\partial S} = -\frac{\partial u_m^*}{\partial S}$ so that when S increases, since U is constant, a change in u_n^* is always associated to an opposite change u_m^*. Hence, the directions of the arrows in row (5) of Table 1 are simply the opposite of that of row (2).

The sign of $\frac{\partial n}{\partial S}$.   As for the change in the equilibrium output of innovation $n\left(u_n^*, s_n^*\right)$ after a change in S, since we can write

$$\frac{\partial n}{\partial S}/n = \phi\frac{\partial u_n^*}{\partial S}/u_n^* + \theta\frac{\partial s_n^*}{\partial S}/s_n^*,$$

and $\frac{\partial u_n^*}{\partial S}$ and $\frac{\partial s_n^*}{\partial S}$ are both strictly positive when $\frac{\theta}{\beta} \in (1, \frac{1-\phi}{\beta})$, then in this case $\frac{\partial n}{\partial S}$ is certainly strictly positive (Scenarios 1b, 2a, and 2b). The only doubt concern Scenario 1a when $\frac{\theta}{\beta} \in (\frac{\phi}{\sigma}, 1]$. In this case, $\frac{\partial u_n^*}{\partial S}$ is negative and $\frac{\partial s_n^*}{\partial S}$ is positive so that the sign of $\frac{\partial n\left(u_n^*, s_n^*\right)}{\partial S}$ is a priori ambiguous. By using (8) and (9), we can write

$$\frac{\partial n}{\partial S}/n = -\phi\frac{\partial u_m^*}{\partial S}/\left(U - u_m^*\right) + \theta\left(1 - \frac{\partial s_m^*}{\partial S}\right)/\left(S - s_m^*\right).$$

Now by using (13) and (A.2) and doing some algebra, we yield

$$\frac{\partial n}{\partial S}/n = \frac{\left[\theta\left(S - s_m^*\right) + \psi\left(\theta s_m^* + \phi S\right)\right]}{\left(S - s_m^*\right)\left(S + (\psi - 1) s_m^*\right)}\left[\frac{\theta\left(S - s_m^*\right) + \psi s_m^*(\theta + \phi)}{\theta\left(S - s_m^*\right) + \psi s_m^*\left(\theta + \phi\frac{S}{s_m^*}\right)} - \frac{\partial s_m^*}{\partial S}\right].$$

Now substitute for $\frac{\partial s_m^*}{\partial S}$ by using (A.1) to yield, after some cumbersome computations,

$$\frac{\partial n}{\partial S}/n$$

$$= \theta\frac{(1 - \beta - \sigma)S + (1 - \beta)(\psi - 1)s_m^*}{(\psi - 1)(\sigma - \phi)\left(S - s_m^*\right)s_m^* + \left((1 - \theta - \phi)s_m^* + (1 - \beta - \sigma)(S - s_m)\right)\left(S + (\psi - 1)s_m^*\right)},$$

which is clearly positive for any value of a and $\frac{\theta}{\beta}$. We have then shown the reason why the arrows in row (6) in Table 1 are all upward.

A.3. PROOF OF PROPOSITION 2

$\frac{\partial^2 g}{\partial a \partial S} > (< 0)$ implies

$$\phi > (<)\frac{(1 - \beta - \sigma)\phi\left(S - s_m^*\right) + \sigma(1 - \theta - \phi)s_m^*}{F(S, U, a)},$$

where

$$F(S, U, a) = \frac{(\psi - 1)(\sigma - \phi)\left(S - s_m^*\right)s_m^*}{\left(S + (\psi - 1)s_m^*\right)} + (1 - \theta - \phi)s_m^* + (1 - \beta - \sigma)\left(S - s_m^*\right) > 0.$$

Now notice that the right-hand side is always positive because of nonincreasing returns in imitation and strictly decreasing in innovation. We can then multiply by $F(S, U, a)$ and solve for s_m^* to yield leads to

$$\frac{\partial^2 g}{\partial a \partial S} > (< 0) \Leftrightarrow s_m^* > (<)\hat{s}_m^* = S\frac{(\psi\phi - (1 - \theta))}{(\psi - 1)(1 - \theta)} = S\frac{(\theta(\beta + \sigma) - \beta)}{\beta(\psi - 1)(1 - \theta)}.$$

Now notice that the sign of $\hat{s}_m^*$ is a priori ambiguous. While the denominator $(\psi - 1)(1 - \theta)$ is always positive, the numerator is positive (negative) only if $\theta > 1 - \psi\phi$ and since $\psi = \frac{\theta\sigma}{\phi\beta}$, we have

$$\hat{s}_m^* \geq 0 \Leftrightarrow \frac{\theta}{\beta} \geq \frac{1}{\beta + \sigma}.$$

Then, we can write

$$\frac{\theta}{\beta} \in \left(\frac{\phi}{\sigma}, \frac{1}{\beta + \sigma}\right] \Rightarrow \frac{\partial^2 g}{\partial a \partial S} < 0, \forall s_m^* \in (0, S),$$

$$\frac{\theta}{\beta} \in \left(\frac{1}{\beta + \sigma}, \frac{1 - \phi}{\beta}\right) \Rightarrow \begin{cases} \frac{\partial^2 g}{\partial a \partial S} \geq 0, \forall s_m^* \in \left(0, \hat{s}_m^*\right] \\ \frac{\partial^2 g}{\partial a \partial S} < 0, \forall s_m^* \in (0, S] \end{cases}.$$

Now focus on the case $\frac{\theta}{\beta} \in \left(\frac{1}{\beta+\sigma}, \frac{1-\phi}{\beta}\right)$. We know that s_m^* is monotonically decreasing in a being $\frac{\partial s_m^*}{\partial a} < 0$ and since $\lim_{a \to 0} s_m^* = S$ and $\lim_{a \to \bar{a}} s_m^* = 0$, there is one and only one value of $a = \hat{a}$ such that $s_m^*(S, U, \hat{a}) = \hat{s}_m^* = S\frac{(\psi\phi - (1-\theta))}{(\psi - 1)(1-\theta)}$. To show that the value of $a = \hat{a}$ such that $s_m^*(S, U, \hat{a}) = \hat{s}_m^*$ and then $\frac{\partial^2 g}{\partial a \partial S} = 0$ is the same for which $\frac{\partial m}{\partial S} = 0$, we substitute the value $\hat{s}_m^* = S\frac{(\psi\phi - (1-\theta))}{(\psi - 1)(1-\theta)}$ in the expression for $\frac{\partial m}{\partial S}$ as a function of s_m^* to obtain

$$\frac{\partial m}{\partial S}\bigg|_{s_m^* = \hat{s}_m^*} = \frac{m\left(\hat{s}_m^*\right)\left((\sigma + \beta) S + \beta(\psi - 1)\hat{s}_m^*\right)}{\hat{s}_m^*\left(S + (\psi - 1)\hat{s}_m^*\right)}$$

$$\times \left(\frac{\partial s_m}{\partial S}\bigg|_{s_m^* = \hat{s}_m^*} - \frac{\sigma \hat{s}_m^*}{(\sigma + \beta) S + \beta(\psi - 1)\hat{s}_m^*}\right).$$

Now

$$\frac{\partial s_m}{\partial S}\bigg|_{s_m^* = \hat{s}_m^*}$$

$$= \frac{\left((S - \hat{s}_m^*)(1 - \theta - \sigma) + (1 - \theta - \phi) s_m^* \psi\right)\hat{s}_m^*}{\left((\psi - 1)(\sigma - \phi)(S - s_m^*)\hat{s}_m^* + ((\beta + \sigma - \theta - \phi)\hat{s}_m^* + (1 - \beta - \sigma) S)(S + (\psi - 1)\hat{s}_m^*)\right)} = 1 - \theta.$$

Substituting for the exact value of $\hat{s}_m^* = S\frac{(\theta(\beta+\sigma) - \beta)}{\beta(\psi-1)(1-\theta)}$, we yield

$$\frac{\partial m}{\partial S}\bigg|_{s_m^* = \hat{s}_m^*} = \frac{m\left(\hat{s}_m^*\right)\left((\sigma + \beta) S + \beta(\psi - 1)\hat{s}_m^*\right)}{\left(S + (\psi - 1)\hat{s}_m^*\right)}(1 - \theta - 1 + \theta) = 0.$$

Hence, according to Table 1, we conclude that

$$a = \hat{a} \Rightarrow \frac{\partial m}{\partial S}\bigg|_{s_m^* = \hat{s}_m^*} = \frac{\partial^2 g}{\partial a \partial S}\bigg|_{s_m^* = \hat{s}_m^*} = 0$$

and therefore

$$\text{sign}\frac{\partial^2 g}{\partial a \partial S} = -\text{sign}\frac{\partial m}{\partial S}, \forall a \in (0, 1).$$

APPENDIX B: ADDITIONAL ROBUSTNESS CHECKS

In this appendix, we report the results of several additional robustness test controlling for: (i) the size of government expenditures (e.g., taxes, government spending and transfers, and subsidies as a percentage of GDP), (ii) the access to sound money (e.g., inflation and money growth as well as freedom to own foreign currency bank accounts), or (iii) market and trade openness (e.g., taxes on international trade, trade barriers, nontariff trade barriers).

Data for the robustness checks come from the Economic Freedom of the World index (EFW) produced by the Fraser institute (see http://www.freetheworld.com/release.html). Cross-country data collected for the EFW index are based on survey data from two widely known publications: the Global Competitiveness Report and the International Country Risk Guide. Results are shown in Table B.1.

After controlling for the size of government expenditure, access to sound money, and market and trade openness, our results remain unchanged. Results by using these controls, along with those for differences in institutional and legal origin, do not significantly change the pattern (or the statistical significance) of the results we obtained in the previous specifications. This is so either when these controls are used altogether in the econometric specification or when they are used one by one in more parsimonious empirical specifications.

TABLE B.1. Additional robustness checks: (1) access to sound money, (2) trade openness, (3) regulations of credit and business

Dependent variable: TFP growth	(1) All	(2) OECD	(3) Dev	(1) All	(2) OECD	(3) Dev	(4) All	(5) OECD	(6) Dev	(7) All	(8) OECD	(9) Dev	(10) All	(11) OECD	(12) Dev
Proximity	−0.008**	−0.079***	−0.004	−0.005	−0.069***	−0.003	−0.008	−0.078***	−0.006	−0.008**	−0.071***	−0.003	−0.005	−0.060***	−0.005
	(0.004)	(0.016)	(0.004)	(0.004)	(0.019)	(0.006)	(0.005)	(0.011)	(0.008)	[0.004]	[0.019]	[0.006]	[0.004]	[0.015]	[0.005]
Tertiary fraction	−0.055	0.134*	−0.453**	0.068	0.105	−0.413**	−0.026	0.078*	−0.371	0.017	0.079*	−0.452**	0.066	0.169*	−0.341
	(0.076)	(0.073)	(0.184)	(0.057)	(0.069)	(0.174)	(0.090)	(0.046)	(0.262)	[0.074]	[0.043]	[0.214]	[0.104]	[0.085]	[0.228]
Proximity*tertiary fraction	−0.059	0.252**	−0.291**	−0.048	0.249**	−0.336***	−0.067	0.275***	−0.285*	−0.009	0.226**	−0.296**	−0.014	0.200*	−0.267*
	(0.078)	(0.107)	(0.135)	(0.079)	(0.119)	(0.112)	(0.098)	(0.080)	(0.175)	[0.075]	[0.101]	[0.135]	[0.095]	[0.110]	[0.150]
Size of government	−0.002	−0.003	0.001	−0.001	−0.001	−0.000									
	(0.002)	(0.005)	(0.004)	(0.002)	(0.004)	(0.003)									
Access to sound money	0.003**	0.002	0.002				0.007***	0.005*	0.006*						
	(0.001)	(0.003)	(0.002)				(0.003)	(0.002)	(0.003)						
Trade openness	0.004	0.005	0.004							0.006	0.013	0.004			
	(0.003)	(0.006)	(0.003)							[0.004]	[0.011]	[0.004]			
Regulations of credit and business	0.004	−0.005	0.001										0.000	−0.011	−0.003
	(0.004)	(0.007)	(0.004)										[0.007]	[0.009]	[0.006]
Constant	−0.065**	−0.033	−0.055*	0.002	0.002	−0.002	−0.057***	−0.045**	−0.050**	−0.044	−0.112	−0.024	−0.003	0.062	0.011
	(0.028)	(0.051)	(0.029)	(0.016)	(0.013)	(0.028)	(0.018)	(0.018)	(0.023)	[0.027]	[0.093]	[0.024]	[0.035]	[0.054]	[0.033]
Observations	260	83	177	301	83	218	307	83	224	288	83	205	268	83	185
Number of IDs	84	21	63	85	21	64	85	21	64	85	21	64	84	21	63
Hansen *p*-value	0.266	0.998	0.612	0.0336	0.447	0.158	0.123	0.368	0.343	0.0173	0.552	0.183	0.0200	0.381	0.164
Hansen stat.	63.23	6.987	52.41	49.34	11.99	39.95	42.57	13.02	34.64	52.39	10.73	39.04	51.75	12.84	39.71
Sargan *p*-value	0.117	0.123	0.105	0.0288	0.163	0.0514	0.216	0.215	0.287	0.0244	0.287	0.0211	0.0228	0.177	0.0291
Sargan stat.	69.91	28.65	69.59	50.06	16.65	46.06	39.07	15.51	35.99	50.84	14.22	50.25	51.15	16.31	48.78
Number of instruments	68	32	67	41	20	40	41	20	40	41	20	40	41	20	40
AR(2) *p*-value	0.132	0.581	0.646	0.444	0.653	0.593	0.363	0.606	0.432	0.292	0.654	0.402	0.125	0.978	0.870
AR(2) stat.	1.505	−0.552	0.460	0.765	0.450	0.534	0.909	−0.516	0.787	1.053	−0.449	0.838	1.536	−0.0281	0.163

Robust standard errors in brackets ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.