GEE

GEE Papers

Number 116

December 2018

Entrepreneurial Human Capital and Firm Dynamics

Francisco Queiró

■■■ GEE

Entrepreneurial Human Capital and Firm Dynamics[1]

Francisco Queiró[2]

Abstract:

This paper shows that entrepreneurial human capital is a key driver of firm dynamics using administrative panel data on the universe of firms and workers in Portugal. Firms started by more educated entrepreneurs are larger at entry and exhibit higher growth throughout the life cycle. The differences are driven by productivity, are particularly strong in the upper tail of the distribution, and do not hold for more educated workers in general. In addition, they do not appear to be driven by omitted ability or selection. Combining these findings with cross-country data to calibrate a simple model of heterogeneous firms, I find that accounting for the effect of entrepreneurial human capital on firm-level productivity increases the fraction of cross-country income differences explained by human and physical capital from 40% to 65%-76%.

JEL Classification Codes: I2; L2; O4

Keywords: Entrepreneurship; Human Capital; Firm Dynamics; Productivity

Note: This article is sole responsibility of the authors and do not necessarily reflect the positions of GEE or the Portuguese Ministry of Economy.

---

[1]A previous version of this paper was circulated under the title "The Effect of Manager Education on Firm Growth". I am very grateful to Andrei Shleifer, Lawrence Katz, Josh Lerner and Alberto Alesina for their support and guidance. I thank Ruchir Agarwal, Pedro Bordalo, Rui Castro, Itzik Fadlon, Miguel Ferreira, Nicola Gennaioli, Danial Lashkari, Joana Naritomi, Clara Raposo, Ariel Stern, Geoffrey Tate, Peter Thompson, Ian Tomb, Gal Wettstein, Tom Zimmerman and Eric Zwick for comments and advice. I also thank Joao Leão and Teresa Feliciano for the opportunity to work with the Gabinete de Estratégia e Estudos (GEE) in Portugal and for their generous help. This project was funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209)

[2]Nova School of Business and Economics (francisco.queiro@novasbe.pt)

Contents

List of Figures

List of Tables

1. Introduction

Recent research suggests that the prevalence of high growth firms plays an important role in development. Hsieh and Klenow (2014) show that manufacturing firms in the U.S. experience stronger life cycle growth in employment and productivity than those in Mexico and India, and that this can account for sizeable differences in the level of aggregate total factor productivity (TFP) between the three countries. Eslava et al. (2018) report that life cycle growth in Colombia is also lower than in the U.S., and not far from that in Mexico. Examining a larger group of developing countries, La Porta and Shleifer (2008) find that growth is extremely rare among the mass of informal firms that account for much of economic activity. Yet little is known about what underlies these differences in firm growth.

One view is that institutional barriers, such as taxes and regulation (Parente and Prescott, 1994), financial frictions (King and Levine, 1993) or contract enforcement (Albuquerque and Hopenhayn, 2004; Acemoglu et al., 2007), may discourage investments in TFP improvements. Another view, going back to the work of Nelson and Phelps (1966), is that the speed of technology adoption is driven by a firm's own human capital, in particular by the human capital of entrepreneurs and managers. In their words, "production management is a function requiring adaptation to change and [...] the more educated a manager is, the quicker will he be to introduce new techniques of production". Nelson and Phelps focus on production efficiency, but the idea can be extended to other drivers of TFP highlighted in recent research, such as management practices (Bloom and Van Reenen, 2007), organizational design (Bloom et al., 2012) or product quality and demand (Foster et al., 2008).

Using administrative data on the universe of firms and workers in Portugal, this paper presents new evidence on entrepreneurial human capital and firm life cycle dynamics, and examines its implications for understanding differences in aggregate TFP. A key challenge in connecting firm dynamics to entrepreneurial characteristics has been the limited availability of comprehensive, high quality data. This paper combines employer-employee matched data, from which I identify entrepreneurs and their characteristics, with financial statements data, from which I measure firm performance, in a twenty-year panel. In addition, Portugal is a particularly attractive setting because all schooling levels from primary school to college are well represented among entrepreneurs.

I start by introducing a simple model of heterogeneous firms based on Hsieh and Klenow (2014) to guide the analysis. Input and output allocations are a function of productivity and idiosyncratic distortions, which may lead to misallocation, and productivity dynamics are determined by the education of entrepreneurs.[3] Aggregate TFP, in turn, depends on the educational attainment of the population, on the distribution of firm productivity by entrepreneur schooling level, and on allocative efficiency.

Turning to the data, I find that both firm size at entry and life cycle growth increase with entrepreneur schooling. As an example, consider figure 1a, which groups firms by entrepreneur years of schooling and plots, for each group, average employment by age for all firms in the 1995 cohort, the oldest cohort of entrants in the data.[4] At entry, firms in the top group, whose entrepreneurs have 15 or more years of schooling, are 23% larger than those in the bottom group, whose entrepreneurs have less than six years of schooling. By age 20, they are 75% larger. The remaining groups fall in between, with growth increasing monotonically with entrepreneur schooling. The same pattern holds when pooling across cohorts. The differences are stronger for gross output and value added, and they are driven by survivor growth, not selection from higher exit rates among smaller firms. If anything, survival rates also increase marginally with entrepreneur schooling. Moreover, they are specific to entrepreneurs. The average schooling of other workers appears to matter much less for firm dynamics.

Are these patterns driven by underlying productivity dynamics? Less educated entrepreneurs might, for example, face stronger financial constraints which limit their growth despite high productivity. I find that productivity dynamics closely resemble

[3]The model can be restated in terms of heterogeneous quality or demand, rather than efficiency, with equivalent observational implications (see Appendix II in Hsieh and Klenow, 2009), and there is evidence that these factors play a major role in accounting for size differences across firms (Foster et al., 2008; Hottman et al., 2016). In the absence of firm-level price data, I cannot distinguish between these channels, and use the term “productivity” loosely to refer to their combined effects.

[4]Each group includes one of the five main levels of educational attainment recorded in the data, corresponding to 4, 6, 9, 12 and 17 years of schooling. Employment differences are conditional on average entrepreneur experience, average non-entrepreneur schooling and experience, and sector.

those for employment and output. Firms in the top group by entrepreneur schooling are 19% more productive at entry than those in the bottom group, and twice as productive by age 20. In addition, the average revenue product of inputs, which increases with the extent of distortions in the model, also increases slightly with entrepreneur schooling. This suggests that size differences actually understate productivity differences, and that firms with more educated entrepreneurs would be even larger in the absence of misallocation.

The comparisons across schooling levels just described impose no parametric restrictions on the relationship between firm performance and entrepreneur schooling. The literature on labor market returns to schooling going back to Mincer (1974) has consistently found that the relationship between earnings and schooling is approximately log-linear, conditional on experience. I also find that a log-linear specification can accurately summarize the relationship between firm productivity and entrepreneur schooling across the life cycle. On average, productivity rises by approximately 5% per year of schooling. The coefficient is higher for older firms, in line with the evidence on life cycle growth, and in more technology-intensive industries, which is consistent with Nelson and Phelp's focus on technology adoption. It is also remarkably stable over time.

A key finding is that the relationship between entrepreneur schooling and productivity is significantly stronger in the upper tail of the distribution than at the mean. In quantile regressions, the coefficient on entrepreneur schooling is close to zero in the left tail of the distribution, and rises to more than twice the OLS coefficient in the right tail.[5] To get a sense of magnitude, the estimates imply that the average firm among entrepreneurs with a college degree is 2.6 times more productive than the average firm among those with no schooling. This ratio rises to 5.6 times at the 99.1th percentile of the distribution, and to 8.7 times at the 99.9th percentile. In other words, entrepreneurial human capital seems to be particularly valuable in fostering the emergence of the large, highly productive firms that account for a substantial fraction of employment and output in developed countries. These differences have important aggregate implications, precisely because firms in the upper tail of the productivity

[5] The same holds for labor market returns to schooling, although the differences there are smaller (Buchinsky, 1994; Martins and Pereira, 2004).

distribution have high market shares, and therefore have a disproportionate impact on aggregate TFP. I assess these implications below.

Two potential sources of bias in these findings are omitted ability and selection into entrepreneurship. I employ two strategies to evaluate these concerns. The first leverages the fact that the employer-employee data reports labor market earnings for entrepreneurs who worked in other occupations before starting their own firms during the sample period. These earnings can be used as a proxy for ability, although earnings in other occupations also increase with schooling, which introduces a negative over-controlling bias in the schooling coefficient. But this bias can be corrected using estimates of the labor market return to schooling, on which there is a large literature (e.g. Card, 1999). When I implement this strategy, the OLS coefficient on entrepreneur schooling is marginally lower but very similar to the baseline estimate, suggesting that ability bias plays a limited role in accounting for the findings. The same holds for the quantile coefficients in the upper tail of the productivity distribution. These findings are in line with the literature on labor market returns to schooling, which also finds that ability bias is small (Card, 1999).

The second concern is that the results might be biased by selection into entrepreneurship. More educated individuals might be more selective when deciding whether to pursue an entrepreneurial opportunity, perhaps because the return to schooling is higher in other occupations. I follow an approach introduced by Combes et al. (2012) to distinguish agglomeration and selection effects on firm productivity in the context of cities. The effect of schooling is parametrized as a shift and dilation of an underlying productivity distribution of potential entrants, while selection generates an endogenous exit threshold that truncates that distribution from the left. I estimate the combination of a shift, dilation and truncation that can best explain the differences between productivity distributions across levels of entrepreneur schooling and find that shifting and dilating the productivity distribution for entrepreneurs with no schooling can almost perfectly replicate the distributions for higher levels of schooling, with no role for truncation. Moreover, the estimated parameters are stable over time. This suggests the results are not driven by selection either.

I conclude by exploring the implications of a causal interpretation of these findings for understanding differences in aggregate TFP and development. Cross-country regressions suggest that human capital plays a major role in explaining output differences across countries (Mankiw et al., 1992), but the results could be biased by omitted factors such as the quality of institutions, culture or geography, among others. The development accounting approach yields a much smaller role for human capital by exploiting within-country labor market returns to schooling (Klenow and Rodriguez-Clare, 1997; Hall and Jones, 1999), but excludes any effect of human capital on TFP.[6] In order to overcome this trade-off, I combine the within-country effect of entrepreneur schooling on firm productivity estimated in Portugal with cross-country data on educational attainment to infer the contribution of schooling to cross-country differences in aggregate TFP.

Using data from Caselli (2005) to facilitate comparison, I implement the variance decompositions proposed by Caselli (2005) and by Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999). I find that accounting for the effect of entrepreneur schooling on TFP increases the fraction of cross-country income variation that can be explained by human and physical capital from 40 percent to between 65 and 76 percent. As mentioned above, the stronger effect of entrepreneur schooling on the upper tail impacts aggregate TFP beyond its effect on mean firm productivity because upper tail firms have higher market shares in equilibrium. This differential upper tail effect accounts for slightly over half of the increase, with the effect on mean productivity accounting for the remainder.

This paper contributes to the large literature on the determinants of aggregate TFP. In particular, it links the long-standing debate on the role of human capital in development (see Erosa et al. (2010), Schoellman (2012), Lagakos et al. (2012), Caselli and Ciccone (2013), Jones (2014), Manuelli and Seshadri (2014) and Hendricks and Schoellman (2018) for recent contributions) to the emerging literature on cross-country differences in firm dynamics, which has mostly focused on misallocation and institutional factors to date (Hsieh and Klenow, 2014; Cole et al., 2016; Bento and Restuccia, 2017; Akcigit et al., 2018). A closely related study is Gennaioli et al. (2013),

[6] Klenow and Rodríguez-Clare (2005) and Córdoba and Ripoll (2008) indirectly infer this contribution from model calibrations, but reach very different conclusions about its magnitude.

who find that the human capital of entrepreneurs increases output at the firm and regional levels, but treat it as a conventional input that complements physical capital and worker human capital in a constant returns production function, rather than a driver of firm productivity. They argue for a large role for human capital in development through larger returns to schooling for entrepreneurs than for other workers, a different channel than the one investigated here.[7] In addition, this paper builds on recent studies that exploit relative size and growth to indirectly infer a broad-based measure of firm-level innovation and technology adoption (Garcia-Macia et al., 2016; Aghion et al., 2017; Eslava and Haltiwanger, 2018).

The rest of the paper is structured as follows. Section 2 describes the data. Section 3 introduces a simple model of heterogeneous firms to guide the analysis. Section 4 presents the empirical findings. Section 5 examines aggregate implications, and section 6 concludes.

2. Data

The data used in the paper comes from two sources. The first is *Quadros de Pessoal* (QP), a matched employer-employee administrative panel data set that covers the universe of firms in Portugal with at least one employee and their workers, including employers and unpaid family workers. The survey combines firm-level information, such as total employment, sales and date of incorporation, with a wide range of worker characteristics, which I use to identify and characterize entrepreneurs.

The second data source is *Sistema de Contas Integradas das Empresas* (SCIE), an administrative dataset that reports financial statements for the universe of firms, covering the period from 2004 to 2015. I rely on this dataset to obtain measures of firm performance beyond employment, namely productivity. The two datasets share a firm identifier and I obtain a match for 94% of firm-year observations in *Quadros de Pessoal* for the period in which the two datasets overlap. This section defines the variables and sample used in the analysis and provides summary statistics.

[7]For additional evidence on returns to education among entrepreneurs, see **?** and the survey by **?**.

2.1. Variable Definitions

Entrepreneurs: An important challenge in the entrepreneurship literature is the identification of entrepreneurs in the data. A standard approach is to define entrepreneurs as those that are self-employed, but this misses entrepreneurs who decide to incorporate and become employees of the firm, which arguably includes those with the highest potential. Azoulay et al. (2018) make progress by leveraging IRS data in the U.S. to define as entrepreneurs those who own equity and report wages from the firm, but ownership information is unfortunately not available for those who incorporate as C-corporations, which again might include the highest potential entrepreneurs.

This paper exploits the rich occupational data reported in QP to define as entrepreneurs the top managers of the firm at entry.[8] This simple definition matches the conventional notion of an entrepreneur as someone who starts and operates a new business, regardless of ownership, compensation or employment status. For some of the results, I extend the sample to include older firms that entered before 1995. For these results, entrepreneurs are defined as the top managers at the time of the *first* observation of the firm in the data. As shown below, top manager schooling is a highly persistent firm characteristic, which suggests this is a reasonable approach.

I identify top managers using the occupational classification in QP, which is based on the International Standard Classification of Occupations (ISCO).[9]. The ISCO provides a three-layer hierarchy of managers, starting with directors, chief executives and general managers, followed by production and operations managers, and then by managers of narrower functional departments, such as HR, finance or sales. I define as top managers those at the highest layer that the firm reports. To maximize coverage, I take two additional steps. First, the data also report a separate hierarchy based on worker qualifications, and the top layer in this classification primarily comprises managerial qualifications. If a firm does not report any managers under ISCO, I define

[8]For firms that do not report a top manager at age zero, I use those reported at age one. Firms that do not report a top manager in the first two years of life are excluded from the sample. The results are robust to different procedures, such as using the first top manager that the firm reports, regardless of the firm's age, or the highest paid employee(s) at entry when the firm does report a top manager.

[9]In particular, it follows ISCO-88 between 1995 and 2009, and ISCO-08 from 2010 onwards. I use the correspondence tables provided by ISCO to match the two

as entrepreneurs the workers assigned to this top layer at entry. Second, if the firm does not report any managers under ISCO or the top qualifications layer, I define as entrepreneurs the workers whose employment status is reported as "employer" at entry. 88 percent of the entrepreneurs in the main sample are identified through the first step, and all results are robust to excluding those identified in steps two and three.

Entrepreneur Schooling: Educational attainment is measured as years of schooling completed. QP reports the highest level of schooling attained by each worker, where the levels are: no schooling, 4th grade, 6th grade, 9th grade, 12th grade, *bacharelato* and *licenciatura*. The *bacharelato* and *licenciatura* are higher education degrees typically lasting three and five years, respectively.[10] The distinction is similar to that between associate and bachelor's degrees in the U.S..

Entrepreneur schooling is defined as average years of schooling of the firm's entrepreneurs. Figure 3 shows a histogram of entrepreneur schooling in the full sample. Over 80 percent of observations cluster at the five main schooling levels reported in the data: four, six, nine, twelve and seventeen years of schooling. In several figures I sort firms by entrepreneur schooling into five groups, each including one of these main levels: zero to less than six years of entrepreneur schooling, six to less than nine, nine to less than twelve, twelve to less than fifteen and fifteen and over.

Firm outcomes: The paper examines the relationship between entrepreneur schooling and several firm outcomes. *Employment* is defined as the number of workers reported in QP, including both entrepreneurs and non-entrepreneurs, regardless of employment status and including unpaid workers. *Gross Output* corresponds to the firm's revenue, also from QP. *Value Added* (py in the model in section 3) is reported in SCIE. I assume *worker human capital* (h in the model) is defined as $le^{r\bar{s}}$, where l is the number of non-entrepreneurs the firm employs, $\bar{s}$ is average worker years of schooling and $r = 0.08$, the mid-point in the 6 to 10 percent return to schooling typically found in the labor literature (Card, 1999). The results are unchanged when experience is

[10]The higher education system changed in 2006 with the European Union's Bologna Accords, which shortened the typical duration of a *licenciatura* to three years, with many students under the new system completing a two-year masters immediately afterwards. The first graduates under the new system entered the labor market in 2009 at the earliest. I assume a duration of five years throughout the sample.

also accounted for in the measurement of human capital. Finally, productivity and the average revenue product of inputs (A and τ in the model) are defined in section 3. In addition to py and h, their measurement uses data on the firm's *physical capital* (k in the model), which is obtained from the book value of assets in SCIE, including both tangible and intangible assets.

Controls: *Sectors* are two-digit industries as reported in QP, *year* represents calendar years and *firm age* is calculated from the firm's reported year of incorporation in QP. When a firm reports different sectors or years of incorporation over time I use the mode of its reports. *Entrepreneur experience* represents the average potential experience of the firm's entrepreneurs, regardless of whether they stayed at the firm over time. It is measured as entrepreneur age at entry, reported in QP, minus years of schooling, minus six, plus the firm's age. *Non-entrepreneur schooling and experience* are the average years of schooling and experience for the workers that the firm employed at entry, and are both calculated in the same way as for entrepreneurs. For the results where I extend the sample to include firms that entered before 1995, non-entrepreneurs are defined as the firm's workers at the time of the *first* observation of the firm in the data.

2.2. Analysis Sample

QP data are available for the period from 1985 to 2015. The occupational classification system used to identify a firm's entrepreneurs was introduced in 1995. Firm age is available starting in 1994. I therefore restrict the analysis to the period between 1995 and 2015.[11]

In addition, the focus of the paper is on private-sector firms. I exclude state-owned firms, defined as those that take the legal form of *Empresa Publica* (state-owned company) or where the state has an equity stake of at least 50 percent. I also exclude government agencies, which are covered when they employ workers under private sector labor law, and non-profits. A number of large privatizations occurred

[11]It should be noted that data on worker characteristics was not collected in 2001. QP consists of three databases: a firm-level database (covering firm-level information such as firm age and total employment), an establishment-level database (e.g. location, employment) and a worker-level database (e.g. schooling, occupation). The worker-level database is not available in 2001. For firms that entered in 2001, I define the entrepreneurs as the top managers that the firm reports in 2002.

during the sample period, involving significant mergers, breakups and downsizings. I exclude these firms by also dropping all private firms that were state-owned at any point in time.[12] Altogether, I exclude 2.6 percent of firms with these filters.

The main sample used in the paper focuses on firms that entered in 1995 or later. These firms are present in the data from entry, which enables me to identify their entrepreneurs. The sample includes all firms that report at least one entrepreneur and one non-entrepreneur, as defined above, along with their year of incorporation and sector. The sample is further restricted to sectors and cohorts where firms from each of the five entrepreneur schooling groups defined above are present, to minimize extrapolation. The impact of this restriction is negligible. The final sample consists of 1.2 million observations, corresponding to 189 thousand firms, and the top left panel in table 1 presents summary statistics. The median firm's entrepreneur has the ninth grade, and the standard deviation across firms is 4.3 years of schooling. Non-entrepreneur schooling has a lower median of 7.75 years of schooling and displays less variation. Average employment is 7 workers, with a median of 4. The firm at the 90^{th} percentile employs 14 workers. Firms have 1.4 entrepreneurs on average, and the majority have just one.

For all results on value added and productivity, the sample is restricted to the time period for which these outcomes are available from SCIE: 2004 to 2015. The sample satisfying these additional restrictions comprises 691 thousand observations from 133 thousand firms. Summary statistics are presented in the bottom left panel of table 1. Firms in this sample are marginally larger, with 8.37 workers on average, but are very similar in terms of entrepreneur and non-entrepreneur characteristics to those in the larger sample.

For some results I extend the sample to include firms that entered before 1995, preserving the remaining restrictions and defining entrepreneurs and non-entrepreneurs as described above. This larger sample consists of 3.9 million observations, corresponding to 457 thousand firms. For comparison, the top and bottom right panels in table 1

[12]In some cases the privatized firms were reincorporated and show up as new firms in the data. To identify these cases, I follow the procedure in Braguinsky et al. (2011): I take all entering firms with over 50 employees and identify those where a majority of workers worked at state-owned firms in the previous year. This procedure identifies an additional 49 firms that I exclude.

display summary statistics for all years and for the shorter period when value added and productivity data are available, respectively. Firms in this sample are significantly older on average, as expected, somewhat larger, and have similarly educated but more experienced entrepreneurs and non-entrepreneurs.

3. Model

This section outlines a simple model of heterogeneous firms based on (Hsieh and Klenow, 2014). I first describe the static equilibrium of input and output allocations, which will be used to measure productivity and distortions in the data and to frame the firm-level empirical analysis. I then impose a simple dynamic structure such that the effect of entrepreneur schooling on firm productivity estimated in Portugal can be combined with cross-country data on educational attainment to perform a development accounting exercise.

3.1. Production and Equilibrium Allocations

The final consumption good consists of a CES aggregate of intermediate goods indexed by ω, competitively produced by a representative firm

$$Y = \left(\int y(\omega)^{\frac{\sigma-1}{\sigma}} d(\omega) \right)^{\frac{\sigma}{\sigma-1}} \tag{1}$$

Intermediate goods are produced by a continuum of firms under monopolistic competition, and each firm faces a demand function given by

$$y = Y \left(\frac{p}{P} \right)^{-\sigma} \tag{2}$$

where p is the price of the firm's output and P is the price of a unit of aggregate output, hereafter normalized to one.

There is a fixed mass L of infinitely lived agents in the economy. Each agent is endowed with a schooling level s, and the fraction of agents at each level of schooling is given by θ_s. Agents are divided into workers and entrepreneurs. Workers inelastically supply their human capital, consisting of e^{rs} effective units of labor, to intermediate goods firms, and are paid a wage w per unit of human capital. Each entrepreneur runs an intermediate goods firm, and is compensated out of firm profits.

A firm with productivity A and employing physical capital k and worker human capital h produces output

$$y = zAk^{\alpha}h^{1-\alpha} \tag{3}$$

where z is an aggregate productivity parameter that affects all firms identically, reflecting factors such as technological spillovers across firms or institutions. Firm productivity A evolves over the life cycle and this evolution is driven by the entrepreneur's schooling. At age a, the productivity of a firm with entrepreneur schooling s is drawn from a density $g_{s,a}$. Productivity here measures process efficiency but, as Hsieh and Klenow (2009) show, the model can be generalized to include differences in product quality and demand as well, with equivalent observational implications.

In addition, firms face idiosyncratic output distortions, denoted by τ_y. As Hsieh and Klenow note, distortions can arise for multiple reasons, such as taxes, financial constraints, markups, transportation costs, and regulation. I use a single output distortion for simplicity but this choice has no impact on the measurement of A, which is the same as in Hsieh and Klenow (2014)'s model with distortions in output and input markets. Given the distortions they face and unit costs u and w, entrepreneurs choose prices and inputs to maximize profits

$$\pi(A, s) = \max_{p,k,h}(1 - \tau_y)py - uk - wh \tag{4}$$

subject to (2) and (3).

Equilibrium input and output allocations will be given by

$$h \propto A^{\sigma-1}\tau^{-\sigma} \tag{5}$$

$$py \propto \left(\frac{A}{\tau}\right)^{\sigma-1} \tag{6}$$

where $\tau \equiv (1-\tau_y)^{-1}$. In the absence of input-specific distortions, the capital-labor ratio will be equalized across firms. Input and output allocations increase with A and decrease with τ. Conditional on τ, a firm with higher A sets a lower price since its production costs are lower, enabling it to expand. Conditional on A, a firm with higher τ sets a higher price to compensate for the implicit taxes it faces in output and input markets, and operates below its efficient scale. Combining (2) and (3) gives the following expression for A

$$A \propto \frac{y}{k^{\alpha}h^{1-\alpha}} \propto \frac{(py)^{\frac{\sigma}{\sigma-1}}}{k^{\alpha}h^{1-\alpha}} \tag{7}$$

while (6) and (7) yield

$$\tau \propto \frac{py}{k^{\alpha}h^{1-\alpha}} \tag{8}$$

A corresponds to "physical" productivity, or TFPQ, while τ is proportional to the average revenue product of labor and capital and corresponds to "revenue" productivity, or TFPR (Foster et al., 2008). As these expressions show, A and τ can be inferred from value added and input data, combined with values for σ and α. I follow Hsieh and Klenow (2014) in setting $\sigma = 3$ in the baseline specification, but examine the sensitivity of the results to setting $\sigma = 4$ and $\sigma = 5$. $1-\alpha$, in turn, is measured using sector labor shares in Portugal from the EU KLEMS database (Jaeger, 2017).

These equilibrium allocations will be used to guide the empirical analysis of the relationship between entrepreneur schooling, firm input and output dynamics, and underlying productivity and distortions in section 4.

3.2. Steady-state Aggregate Output and TFP

In section 5, I use the firm-level findings to construct counterfactuals of steady-state aggregate TFP as a function of entrepreneur schooling, and use these counterfactuals to perform a development accounting exercise. With that goal in mind, I impose a simple dynamic structure with exogenous entry and exit.

Each period, a fraction δ of entrepreneurs randomly exit and become workers, shutting down their firms, and a fraction γ of workers randomly quit their jobs and become entrepreneurs, starting new firms with productivity drawn from $g_{s,0}$. The steady-state fraction of entrepreneurs is therefore given by $\frac{\gamma}{\gamma+\delta}$. Since the transition probabilities are independent of s, θ_s represents the steady-state fraction of agents with schooling s both in the population and among entrepreneurs. This enables me to use data on the distribution of schooling in the population to proxy for the distribution among entrepreneurs in each country. While this equality does not hold exactly in the Portuguese data, where entrepreneurs tend to have slightly higher levels of education than non-entrepreneurs, it is a reasonable approximation. See table 1 for firm-level summary statistics.

In steady-state equilibrium, all aggregate variables are constant and the productivity and firm size distributions are stationary, as is the firm age distribution.[13] Firms exhibit life cycle dynamics that are specific to each s, and the steady-state productivity distribution for each s can be written as

$$\mu_s(A) = \delta \sum_{a=0}^{\infty} (1-\delta)^a g_{s,a}(A) \tag{9}$$

Exogenous entry and exit imply that μ_s depends only on the $g_{s,a}$ distributions and δ. The same applies to the steady-state distributions of relative employment and output, through equations (5) and (6). In particular, μ_s is independent of schooling shares θ_s. This enables me to use the μ_s estimated in the Portuguese data to construct aggregate TFP counterfactuals of changes in the distribution of schooling, under the assumption that $g_{s,a}$ and δ are the same across countries.

These exogenous dynamics are of course a simplification. An alternative approach would be to endogenize entry, calibrate the model using the estimated μ_s and then use the calibrated model to simulate the effect of changes in the distribution of schooling, accounting for the endongenous response of entry. Hsieh and Klenow (2014) follow this approach when computing the effect of changes in firm life cycle growth on aggregate

[13]Note that there is no growth, since the focus of the analysis is on differences in levels of aggregate output and TFP, but a simple way to introduce a balanced growth path would be to add constant exogenous growth in z.

TFP, and find that it leads to results that are similar to those of a simple model with exogenous entry. They reach the same conclusion both when holding entrant quality fixed, as in Hopenhayn (1992), and, perhaps more surprisingly, also when endogenizing the productivity threshold for entry, as in Lucas (1978). In the endogenous entrant quality model, stronger incumbent growth increases the threshold for entry and this raises average firm productivity, but this selection effect is largely offset by changes in product variety resulting from reduced entry.[14] Motivated by their findings, I adopt the simpler model with exogenous dynamics. In addition, I present evidence below that suggests selection into and out of entrepreneurship plays a limited role in explaining differences in productivity across levels of entrepreneur schooling.

To express aggregate outcomes, let $H = L\sum_s \theta_s e^{rs}$ denote the worker human capital of all agents in the economy, including those employed as entrepreneurs, and let K represent the exogenous stock of physical capital. In addition, define $N \equiv \frac{L\gamma}{\gamma+\delta}$ as the steady-state mass of entrepreneurs in the economy. Then steady-state aggregate output can be written as

$$Y = \text{TFP } K^{\alpha} H^{1-\alpha} \tag{10}$$

This expression corresponds to the standard Cobb-Douglas aggregate production function used in development accounting, except that TFP is partly driven by entrepreneur schooling and given by

$$\text{TFP} = zN^{\frac{1}{\sigma-1}}\left(\frac{\delta}{\delta+\gamma}\right)^{1-\alpha}\left[\sum_s \theta_s \int\int \left(A\frac{\bar{\tau}}{\tau}\right)^{\sigma-1} f(\tau|A)\mu_s(A)dAd\tau\right]^{\frac{1}{\sigma-1}} \tag{11}$$

where $\bar{\tau}$ is average τ across firms, weighted by relative value added shares,[15] and f is the distribution of τ conditional on A.

Aggregate TFP depends on aggregate productivity z, variety $N^{\frac{1}{\sigma-1}}$, the share of agents employed as workers $\frac{\delta}{\delta+\gamma}$, and average firm productivity, given by the term in brackets. Average firm productivity, in turn, is a power mean of firm productivity across

[14] See rows one to three in tables IV and V of their paper.

[15] $\bar{\tau} \propto \left[\sum_s \theta_s \int\int \tau^{-1} \frac{py}{PY} f(\tau|A)\mu_s(A)dAd\tau\right]^{-1}$

schooling levels s, with weights given by the share of entrepreneurs at each s.[16] The exponent $\sigma - 1 > 0$ reflects the fact that, in equilibrium, firms that are more productive or face less distortions are also larger, and therefore receive heavier weight. Following Hsieh and Klenow (2017), aggregate TFP in (11) can be decomposed into separate contributions from firm productivity and from allocative efficiency:

$$\text{TFP} = zN^{\frac{1}{\sigma-1}} \left(\frac{\delta}{\delta+\gamma}\right)^{1-\alpha} \times \tilde{A} \times E \tag{12}$$

where

$$\tilde{A} \equiv \left[\sum_s \theta_s \int A^{\sigma-1}\mu_s(A)dA\right]^{\frac{1}{\sigma-1}}$$

$$E \equiv \left[\sum_s \theta_s \int\int \left(\frac{A}{\tilde{A}}\frac{\bar{\tau}}{\tau}\right)^{\sigma-1} f(\tau|A)\mu_s(A)dAd\tau\right]^{\frac{1}{\sigma-1}}$$

Here, $\tilde{A}$ represents average firm productivity in the absence of misallocation, while allocative efficiency is captured by E and depends on the dispersion of average revenue products τ across firms. For the counterfactuals in section 5, I assume E is unchanged, in line with the prevailing view that allocative efficiency is determined by policy and institutional distortions (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009). In other words, I focus on the effect of schooling on TFP through its effect on $\tilde{A}$, and assign any cross-country variation in E and the remaining terms in (12) to unexplained residual TFP.

4. Entrepreneur Schooling and Firm Dynamics

4.1. Life Cycle

I start by presenting graphical evidence on firm life cycle dynamics by level of entrepreneur schooling. Throughout this section, firms are sorted into five groups by average entrepreneur schooling – zero to less than six years of schooling, six to less than nine, nine to less than twelve, twelve to less than fifteen and fifteen and over. Each

[16] $g_{s,a}$, f and δ are assumed to be such that these integrals are finite for each s.

group includes one of the major schooling levels reported in the data, as shown in figure 3. For each figure, I estimate an OLS regression of the following form

$$\ln Q_{i,t} = \sum_s \sum_a \beta_{s,a} c_{s,i} d_{a,i} + \phi X_{i,t} + \epsilon_{i,t} \tag{13}$$

where Q is an outcome of interest, $c_{s,i}$ and $d_{a,i}$ are dummies indicating whether firm i belongs to entrepreneur schooling group s and is of age a, respectively, and X is a vector of controls including the average schooling of non-entrepreneurs, a quadratic in average entrepreneur and non-entrepreneur experience, sector fixed effects, and also year fixed effects in the figures which track more than one cohort over time. I then plot $e^{\hat{\beta}_{s,a}+\hat{\phi}\bar{X}}$, which corresponds to the estimated geometric mean of Q for each schooling group at each age, evaluated at the sample mean of X. It may be argued that the schooling and experience of non-entrepreneurs are choice variables for the entrepreneur and part of the effect of entrepreneur schooling, in which case they should be omitted from the regression. While this is true from the perspective of an individual entrepreneur, entrepreneurs as a whole are constrained by the aggregate supply of human capital in the economy. I therefore exclude this matching channel by controlling for non-entrepreneur characteristics, but show below that this choice has little impact on the results. It should be noted that none of these controls address concerns with omitted ability and selection, which I also confront separately below.

I first present evidence from tracking a single cohort over time. This ensures that the observed patterns reflect true life cycle dynamics, unconfounded by cohort effects. Figure 1a plots employment by age for each entrepreneur schooling group for all firms in the 1995 cohort, the oldest cohort of entrants in the data. The sample period ends in 2015, so I am able to track these firms up to age 20. Firms start small across entrepreneur schooling groups, with around four to five employees on average, and average employment tends to grow with age. But size at entry and especially growth increase with entrepreneur schooling. At entry, firms in the top group, whose entrepreneurs have 15 or more years of schooling, are 23% larger than those in the bottom group, whose entrepreneurs have less than six years of schooling. By age 20, they are 75% larger. The remaining groups fall in between, with growth increasing monotonically with entrepreneur schooling. Figure 1b shows that the same pattern holds for gross output, with stronger differences. Firms in the top group are 69% larger at entry than those in the bottom group, and 2.6 times larger at age 19.[17]

These differences could be driven by survivor growth or by selection out of entrepreneurship, as emphasized in the models of Jovanovic (1982) and Hopenhayn (1992). If small firms are relatively more likely to exit among more educated entrepreneurs, then the pattern in figure 1a could emerge in the absence of differences in firm growth. To distinguish these channels, figure 2a plots average employment for firms who survived until age 20 versus average employment for all firms, for the top and bottom groups by entrepreneur schooling. In both groups, survivors are on average larger at all ages except age 20, where they are the same by construction. This indicates the presence of selection. But the *differences* between the two groups are clearly driven by differences in survivor growth. If anything, differences in size at entry between the top and bottom groups are smaller for surviving firms than for all firms, which suggests selection might be somewhat stronger for the bottom group in the early stages of the life cycle.

Another question is whether more educated entrepreneurs are simply pursuing riskier strategies, with lower probabilities of survival but higher growth conditional on survival. Figure 2b plots cumulative survival rates by age for the same groups, and shows that this is not the case.[18] In fact, firms in the top group are slightly more likely to survive, while firms in the remaining groups experience very similar survival rates. The cumulative survival rate for 20-year old firms in the top group is 32%, which compares with 23% for the bottom group. Differences in survivor growth do not seem to be driven by increased risk taking.

Is this relationship specific to entrepreneurs or does it hold for more educated workers in general? Figure 2c is constructed analogously to figure 1b but sorts firms in the 1995 cohort by average non-entrepreneur schooling instead, and shows that there are no differences in gross output across groups. This evidence suggests that it is the human capital of entrepreneurs, in particular, that matters for firm dynamics, very much in line with the ideas of Nelson and Phelps (1966), Welch (1970) and Schultz (1975). I focus on

[17]Gross output in QP in reported with a one-year lag, so data is only available up to age 19.

[18]For this graph, equation (13) is estimated in levels, not logs.

output instead of employment for this comparison because non-entrepreneur schooling and the number of non-entrepreneurs are likely to be substitutes in production, and in fact performing the same exercise for employment shows that it is negatively correlated with non-entrepreneur schooling.

A key question is whether these differences in employment and output dynamics are driven by underlying productivity dynamics or by misallocation. In the absence of distortions, employment and output dynamics would be entirely driven by productivity in the model presented in section 3. This is also the standard interpretation of firm dynamics in canonical models such as Hopenhayn (1992) or Klette and Kortum (2004). But in the presence of distortions, employment and output dynamics could also reflect misallocation. Examples of such distortions include financial constraints, taxes and regulations, markups, adjustment costs or transportation costs.[19]. Hsieh and Klenow (2009, 2014) propose a method for measuring such misallocation regardless of the underlying source, which the model in section 3 follows. When input and output allocations across firms are efficient, the marginal revenue product of labor and capital – which are proportional to average revenue products in the model – should be equalized across firms. When distortions lead to misallocation, the smaller a firm is relative to its efficient allocation, the higher its average revenue product of inputs – τ in the model – should be. If, for example, firms with less educated entrepreneurs face stronger financial constraints which limit their growth, then these firms should exhibit relatively higher values of τ than firms with more educated entrepreneurs.

Unfortunately, the additional data needed to compute productivity and average revenue products, namely value added and physical capital, are only available starting in 2004, which limits the analysis from tracking a single cohort to a shorter period. To get around this and also to provide evidence that is representative of the broader sample, I next pool firms across all cohorts observed from entry, from the 1995 cohort to the 2015 cohort.

Figure 4a displays employment by age for this pooled sample. Note that sample size declines strongly with age in this sample, as the composition of cohorts changes. Age

[19]For example, several models of firm dynamics interpret growth as the result of loosened financial constraints over the life cycle (Evans and Jovanovic, 1989; Cooley and Quadrini, 2001; Albuquerque and Hopenhayn, 2004; Clementi and Hopenhayn, 2006)

0 includes firms from the 1995-2015 cohorts, age 1 includes the 1995-2014 cohorts and so on until age 20, which includes only the 1995 cohort observed in 2015. Estimates for older ages are therefore noisier and more likely to be affected by cohort effects. Reassuringly, the patterns for employment are exactly the same as for the 1995 cohort. Firms in the top group start out 22% larger than those in the bottom group and are 95% larger by age 20, with growth in the remaining groups increasing monotonically with entrepreneur schooling. The results for value added are shown in figure 4b and are stronger than those for employment, as in the case of gross output. Firms in the top group are 49% larger at entry than those in the bottom group, and 3.2 times larger at age 20.

Figures 4c and 4d plot productivity A and average revenue product τ by age for the same sample, normalized by the respective average values for entrants in the bottom group. Both figures use the same scale for comparison, and the contrast is clear. Productivity dynamics by entrepreneur schooling closely resemble those of employment and value added. Firms in the top group are 19% percent more productive at entry, and 2.1 times more productive at age 20. τ, on the other hand, is essentially flat over the life cycle and very similar across levels of entrepreneur schooling. If anything, firms in the top group have slightly higher levels of τ, which suggests that they would be even larger at their efficient scale, or in other words, that differences in employment and output understate true differences in productivity. This does not of course imply the absence of misallocation in Portugal, only that the differences in firm life cycle dynamics across levels of entrepreneur schooling documented above are largely driven by productivity.

The evidence presented so far comes from firms that are observed from entry. These are the firms whose entrepreneurs, defined as the firm's top managers at entry, can be observed in the data. However, no cohort can be tracked from entry and beyond age 20 given the time span covered by the sample, which does not paint a full picture of the life cycle. In order to do so, I extend the pooled sample to include all firms observed in the data, including those that entered before 1995. In this sample, I define entrepreneurs as the *first* top managers that the firm reports in the data, not necessarily at entry. The implicit assumption is that top manager schooling is a persistent firm characteristic. Figure 5 shows that this is indeed the case by plotting the average years of schooling of top managers over the life cycle for firms in the 1995 cohort.[20] Differences in top manager schooling persist strongly for the 20 year period covered by the sample. More broadly, the one-year autocorrelation of top manager schooling in the full sample is 0.95 and the ten-year autocorrelation is 0.76.

Figures 6a and 6b display employment and value added by age for all firms in the full sample, sorting firms into the same groups by entrepreneur schooling as before. Within these groups, firms are divided into five-year age bins, including a separate bin for entrants and grouping firms over 50 years old into a single bin. For each entrepreneur schooling group, I plot average firm size and average firm age for each age bin. For the first 20 years of the life cycle, the patterns are entirely consistent with the evidence presented above. By age 20, differences across groups are remarkably similar to that observed for the 1995 cohort and for the pooled sample of entrants, suggesting that any bias from extending the sample to older cohorts whose entrepreneurs are not observed at entry is not substantial. Beyond age 20, the figure shows that the divergence appears to continue throughout the life cycle. By age 40, firms in the top group employ 3.1 times more workers and have 5.3 times higher value added than those in the bottom group, and by age 70 those numbers rise to 4.5 and 9.5 respectively.

Figures 6c and 6d replicate the same exercise for productivity and for the average revenue product of inputs and the outcome is the same as for the sample of firms observed from entry. Productivity dynamics follow a pattern similar to employment and value added, while the average revenue product of inputs is again flat over the life cycle and similar across levels of entrepreneur schooling.

Overall, these different samples present a consistent picture of the life cycle. From here onward, unless otherwise noted, all results use the sample of cohorts observed from entry, which includes all firms whose entrepreneurs can be directly identified in the data.

4.2. Functional Form

The evidence presented so far documents a strong relationship between firm dynamics and entrepreneur schooling that is driven by underlying productivity, without imposing

[20]This graph plots simple means, not coefficients from estimating equation (13).

any parametric restrictions on that relationship. But entrepreneurs in the model are compensated out of firm profits, which are linear in value added and, through equation (6), in $A^{\sigma-1}$. The literature on labor market returns to schooling going back to Mincer (1974) has consistently found that the relationship between compensation and schooling in the labor market is approximately log-linear, conditional on experience. This section examines whether the same holds for the relationship between productivity and entrepreneur schooling. Specifically, I estimate the following OLS regression:

$$\ln A_{i,t} = \beta s_i + \phi X_{i,t} + \epsilon_{i,t} \tag{14}$$

where X is the same vector of controls as in equation (13) with an additional quartic in the firm's age.[21] Including firm age is not necessarily the right approach. If schooling affects survival, then part of its effect operates through firm age, which should be omitted. But figure 2b in the previous section and table 3 below show that this mechanism is likely to be weak, and controlling for firm age accounts for the fact that differences in the firm age distribution by schooling level may not be in steady-state, even conditional on experience.

Figure 7a examines how well the log-linear specification fits the conditional expectation function for productivity. Firms are sorted into twenty equal-sized bins by entrepreneur schooling and the log of productivity is plotted against entrepreneur years of schooling for each bin. Both variables are first residualized on X. Along with the binned scatter plot, the graph also displays the corresponding regression line, obtained by estimating (14) on the underlying data. As the figure shows, log productivity is a slightly convex function of schooling, but the log-linear model fits the data closely, with a slope of 0.0544. The emergence of convexity has also been noted in the literature on labor market returns to schooling, and has been associated with an increase in the relative demand for skilled labor not met by a corresponding increase in supply (Lemieux, 2006). The bottom panel in figure 7 adds firm age-specific entrepreneur schooling coefficients to equation (14) and plots these coefficients over the life cycle, along with the respective 95% confidence intervals. Consistent with stronger

[21] Adding more flexible polynomials in the covariates or sector-by-year fixed effects does not change the results. The use of a parsimonious specification eases the computational burden of estimating the quantile analogs of equation (14) below.

growth, the schooling coefficient rises with age, from 0.0241 at entry to 0.0619 at age 20. Nevertheless, the average coefficient over the life cycle offers a convenient summary of the relationship. Given the close fit, I rely on this parsimonious log-linear approximation going forward.

Table 2 presents the results from estimating equation (14), along with a series of robustness checks. Starting with the baseline specification in column one, which corresponds to the top panel of figure 7, the coefficient on entrepreneur schooling equals 0.0554, as reported above. In contrast, non-entrepreneur schooling has a much weaker impact, with a coefficient of 0.0083, in line with the evidence presented in the previous section. Turning to experience, the coefficients on the linear terms for entrepreneurs and non-entrepreneurs are positive while those on the quadratic terms are negative, a pattern which holds across all specifications and is consistent with the evidence on concave labor market returns to experience. Column two presents estimates for the full sample, including firms not observed from entry, and the results are very similar, with a coefficient on entrepreneur schooling of 0.0536. Column three in turn restricts the sample to firms that survived throughout the sample period, and the coefficient increases to 0.0603. In line with the evidence from the 1995 cohort, this indicates that the schooling coefficient is driven by survivor growth, not by selection from less productive firms exiting at higher rates among more educated entrepreneurs.

The next three columns consider variations in the set of covariates X. Column four excludes the firm's age, and the coefficient on entrepreneur schooling drops slightly, to 0.049, while the coefficient on non-entrepreneur schooling becomes negative. As shown in table 3 below, these differences are not driven by a negative effect of schooling on firm survival, and likely reflect the fact that the firm age distribution by schooling level is not in steady-state as discussed above. Column five excludes non-entrepreneur schooling and experience, and the coefficient on entrepreneur schooling is nearly unchanged, at 0.0573, suggesting that assortative matching between entrepreneurial and non-entrepreneurial human capital does not play much of a role in driving firm productivity. In the model, firms have only one entrepreneur, but in the data 35 percent of firms have at least two. Column six adds the log of the number of entrepreneurs to account for these differences, and the coefficient on entrepreneur schooling again remains very similar, at 0.0572.

In addition, productivity itself could be mismeasured. A key parameter here is σ, which determines the elasticity of value added with respect to productivity. The higher σ is, the smaller the productivity differences inferred from the data through equation (7). Columns seven and eight present results for $\sigma = 4$ and $\sigma = 5$, instead of the baseline $\sigma = 3$ from (Hsieh and Klenow, 2014), and the coefficient on entrepreneur schooling drops to 0.0424 and 0.0359, respectively. On the other hand, higher σ increases the relative output shares of more productive firms, and hence their weight on aggregate productivity, which partially offsets the first effect. In section 5, I investigate the aggregate effect of changes in entrepreneur schooling as a function of σ. Other than σ, the values for α could be incorrect, or the human capital of entrepreneurs could enter the production function directly, instead of only through an effect on productivity. Rather than taking a stand on what the true production function is, Appendix table 8 shows that the results are not very sensitive to these choices.

Table 3 presents results for additional outcomes, using the same specification as the baseline estimates for productivity. The first column looks at value added, and the coefficient on entrepreneur schooling equals 0.0778. Based on productivity differences alone, the elasticity of value added with respect to entrepreneur schooling should equal $(\sigma - 1)\beta$, from combining equations (6) and (14). Using $\sigma = 3$ and $\beta = 0.0544$ leads to an elasticity of 0.1088, which is larger than the estimated coefficient. The same holds for the coefficients on the total human capital of workers (.0545) and on physical capital (0.084), in columns two and three. This again suggests that size differences understate productivity differences, as suggested by figures 4d and 6d. Column four confirms this by turning to the average revenue product of inputs τ and finding a positive coefficient on entrepreneur schooling of 0.0165, which implies that firms with more educated entrepreneurs would be larger in the absence of misallocation. Lastly, column five turns to survival, using a linear probability model. The coefficient on entrepreneur schooling equals 0.0003, positive and significant at the 5% level, but too small for survival to be a relevant channel. This supports the assumption of a single exit rate δ across levels of entrepreneur schooling in the model.

To what extent can these findings be extrapolated to other settings? For example, how might the coefficient on entrepreneur schooling change if the distribution of schooling in the population changes, or under a different set of institutions? Answering these questions decisively would require data for other countries, but one way to shed light on the external validity of these findings is to examine how stable the coefficients are over time. During the sample period for which productivity data is available, from 2004 to 2015, average years of schooling in the working population rose by two years in the Quadros de Pessoal data, from 8.17 to 10.18. In addition, Portugal experienced a financial crisis and deep recession from 2011 to 2013, during which access to external finance for firms is likely to have been severely restricted. To examine how these changes affect the relationship between entrepreneur schooling and productivity, table 4 reports estimates of equation (14) with year-specific entrepreneur schooling coefficients, for both productivity and the other outcomes in table 4. I restrict the sample to firms aged nine or less, so that each year includes the same number of cohorts and the age distribution is comparable across years.[22] In all cases, the coefficient is remarkably stable and does not appear to exhibit any trend over time. In the case of productivity and τ, the coefficient rises slightly during the recession, and drops back to the baseline afterwards, such that the coefficients for 2004 and 2015 are nearly identical.

Finally, estimating separate schooling coefficients by sector shows that the relationship with productivity is stronger is manufacturing and services than in agriculture, fishing or mining. In addition, within manufacturing and services it is strongest in more technology intensive sectors. In manufacturing, the highest coefficients are in chemical products (0.1107), basic metals (0.1053), electrical machinery (0.1054) and motor vehicles (0.1527). In services, they are in post and telecommunications (0.1527) and computer and related activities (0.1469). The point estimates and confidence intervals for all sectors are plotted in figure 8. This evidence supports the emphasis placed by Nelson and Phelps (1966) on technology adoption as a key channel for the effect of human capital on productivity.

4.3. The Distribution of Productivity

The evidence presented so far characterizes average outcomes by schooling level. But the model in section 3 allows the entire distribution of productivity, and not just its mean, to depend on entrepreneur schooling. As equation (11) makes clear, this matters because aggregate productivity is not a simple average of firm productivity.

[22] 2004 includes the 1995-2004 cohorts, 2005 includes the 1996-2005 cohorts, and so on.

The exponent $\sigma - 1$ reflects the fact that more productive firms have higher market shares in equilibrium, and therefore receive relatively more weight. This implies that if the effect of entrepreneur schooling varies along the productivity distribution, then its effect on average firm productivity does not fully capture its aggregate impact. This section characterizes the full μ_s productivity distributions, holding the distribution of covariates X constant. I start by estimating a series of quantile regressions, in order to examine how the coefficient on entrepreneur schooling varies along the distribution of productivity, and then use the approach developed by Machado and Mata (2005) to go from the conditional quantiles to the marginal distribution of productivity as a function of entrepreneur schooling.

I estimate quantile analogs of the baseline OLS specification in column one of table 2, based on equation (14), at intervals of one-fifth of a centile, or in other words for quantiles [0.1, 0.3,...,99.7,99.9]. Figure 9 plots the entrepreneur schooling coefficient from these 500 quantile regressions, along with 95% confidence intervals, and a clear pattern emerges. The coefficient is close to zero and insignificant on the lower tail of the productivity distribution, rises steadily with the quantiles of the distribution, and attains its highest levels in the upper tail, particularly in the 99th percentile and above. At the 50.1th percentile, for example, it equals 0.0510, close to the OLS coefficient of 0.554, but at the 99.1th percentile it rises to 0.1009, almost twice as large as the OLS coefficient, and and at the 99.9th percentile it reaches a maximum of 0.1270. To get a sense of magnitude, these coefficients imply that the average firm among entrepreneurs with a college degree is $e^{17\times 0.554} = 2.6$ times more productive than the average firm among those with no schooling. At the 99.1th percentile it is 5.6 times more productive and at the 99.9th percentile it is 8.7 times more productive.

This evidence suggests that entrepreneur schooling is particularly important for the emergence of highly productive firms in the right tail, which weigh disproportionately on aggregate productivity through their large market shares. The OLS results on average firm productivity therefore understate the aggregate impact of entrepreneur schooling. Interestingly, these results are consistent with the evidence from quantile regressions of labor market earnings, in which the schooling coefficient is also stronger in the upper tail (Buchinsky, 1994; Martins and Pereira, 2004), although the differences there are significantly smaller.

The quantile regressions characterize the distribution of productivity as a function of entrepreneur schooling, conditional on the vector of covariates X. To go from conditional to marginal distributions for each level of schooling, I draw a random sample with replacement of 10,000 observations from the data, and then use the quantile regression coefficients to compute the 500 predicted productivity quantiles for each level of schooling using the covariate values for each observation.[23] As Machado and Mata (2005) show, this procedure simulates the marginal distribution of productivity implied by the model in equation (14) for that level of schooling.

To evaluate the simulation procedure, I first check how well the simulated distribution can replicate actual productivity distributions, by applying the model to the actual covariates observed in the data for each level of schooling. I compare the actual and simulated distributions for entrepreneurs with zero, four, six, nine, twelve and seventeen years of schooling. As figure 3 shows, there are relatively few entrepreneurs with zero years of schooling in the sample, 864 firm-year observations to be exact. Despite this, it is important to examine the fit for this group separately in order to assess the validity of the cross-country counterfactuals in section 5, since the fraction of the population with no schooling in some developing countries is significant.

If the functional form for the quantile model holds exactly, then the simulated and actual distributions for every level should overlap perfectly. Figure 10 plots kernel density estimates for the actual and simulated distributions. Starting with the group with no schooling, the fit is remarkably good given the small sample size, though not perfect. The small hump in the left tail of the simulated distribution resembles the one in the actual distribution for entrepreneurs with four years of schooling, which represents a much larger fraction of the data, and the simulated distribution also slightly understates the right tail. Turning to the remaining education levels, the fit is nearly perfect. The right tail for entrepreneurs with seventeen years of schooling is also slightly understated, but the differences are minor. These results not only validate the simulation procedure, but also highlight why the Portuguese context, with its balanced representation of entrepreneurs across education levels, is particularly well suited for this analysis.

[23]This leads to a dataset of 5 million rows for each level of schooling. Increasing sample size beyond this level does not appear to affect the results.

Having validated the method, figure 11 plots the simulated marginal productivity distributions for different levels of entrepreneur schooling, holding the distribution of covariates in X constant and equal to the overall distribution in the sample. For the cross-country counterfactuals, I use the simulated distributions for each schooling level reported in the Barro and Lee (2001) data. For illustration, the figure plots kernel density estimates for the distributions for four of these levels: zero, six, twelve and sixteen years of schooling.[24] As would be expected from the quantile regressions, the higher the schooling level, the more shifted to the right and dilated the productivity distribution is.

4.4. Accounting for Ability

A key challenge in assigning a causal interpretation to the schooling coefficient estimated above is the possibility that it is biased by omitted ability differences that are correlated with schooling. There is a large literature on labor market returns to schooling devoted to this issue and the prevailing view is that ability bias in a simple OLS regression of individual earnings on schooling is small (Card, 1999). Still, this finding may not extend to the context of entrepreneur schooling and firm productivity. This section exploits information on entrepreneurs' labor market earnings in other occupations prior to becoming entrepreneurs as a proxy for omitted ability differences.

Consider a simple extension of equation (14) where productivity is a function of the entrepreneur's natural ability b, in addition to schooling:

$$\ln A_{i,t} = \beta^e s_i + \lambda^e b_i + \phi X_{i,t} + \epsilon_{i,t} \tag{15}$$

Suppose also that the entrepreneur's potential earnings in the labor market, as a worker, take the standard Mincerian form

$$\ln w_{i,t} = \beta^w s_i + \lambda^w b_i + \nu_{i,t} \tag{16}$$

[24] Barro and Lee (2001) assume college corresponds to sixteen years of schooling, which differs from the seventeen for the *licenciatura* in Portugal. I use their assumption to simulate the distributions for the cross-country analysis.

I omit experience in the earnings equation to save on notation but account for it in the results below. Inverting (16) to express b as a function of $\ln w$ and s, equation (15) can be rewritten as

$$\ln A_{i,t} = (\beta^e - \frac{\lambda^e}{\lambda^w}\beta^w)s_i + \frac{\lambda^e}{\lambda^w}\ln w_{i,t} + \phi X_{i,t} + \varepsilon_{i,t} - \frac{\lambda^e}{\lambda^w}\nu_{i,t} \tag{17}$$

This expression shows that the entrepreneur's potential earnings in the labor market, w, can be used as a proxy control for b, but that this introduces an over-controlling bias, since w is also partly determined by schooling s. Intuitively, if w is held constant, higher s implies an offsetting change in b which again biases the coefficient on entrepreneur schooling. However, the new bias is equal to the coefficient on $\ln w$ multiplied by β^w, the labor market return to schooling. I can therefore draw on the extensive literature on returns to schooling (see Card, 1999, for a survey) to obtain estimates of β^w and recover the true coefficient on entrepreneur schooling β^e.

The key assumption underlying this approach is that ability can be represented by a scalar b. If there are multiple dimensions of ability that affect firm productivity and are correlated with schooling, then a single control cannot proxy for those multiple dimensions. In this regard, this approach parallels the widely used Olley and Pakes (1996) method of inverting a firm's investment equation in order to recover its productivity, which also assumes that productivity can be represented by a scalar.

In order to estimate equation (17), data on the entrepreneurs' potential labor market earnings is required. For this purpose I use a sample of switchers – people who worked in other occupations before becoming entrepreneurs within the sample period. In this sample, which comprises just under half of the baseline sample, I observe an entrepreneur's income when working as a non-entrepreneur in a prior employment spell, and I take the entrepreneur's last observed non-entrepreneurial income, residualized on year and experience dummies, as the entrepreneur's potential earnings in the labor market. The results are robust to using the average of all previous observations of non-entrepreneurial income, rather than just the last one.

One concern with this procedure could be measurement error. I do not observe the entrepreneur's actual potential earnings at year t, and instead proxy for it with earnings at a previous job. This might attenuate the coefficient on $\ln w$ and amplify the coefficient on s. But measurement error would attenuate the bias correction for the schooling coefficient as well. As long as measurement error is not correlated with schooling, the bias-corrected estimate of β^e would be minimally affected.[25]

Table 5 presents the results from accounting for ability under this approach. Column one presents the baseline specification, from column one of table 2, estimated on the sample of switchers. The coefficient on entrepreneur schooling in this sample equals 0.0641, somewhat higher than the coefficient in the larger sample. Column two adds the entrepreneur's potential labor market earnings. First, as expected if ability increases both labor market earnings and firm productivity, the coefficient on earnings is positive and significant. Second, the coefficient on entrepreneur schooling falls to 0.0329 and also remains significant. Third, the bias-corrected coefficient β^e equals 0.0604, only marginally lower than the baseline estimate without controlling for ability. As explained above, this is obtained by adding the coefficient on earnings, multiplied by an estimate for the labor market return to schooling, β^w, to the coefficient on entrepreneur schooling. I assume an estimate of 8 percent for the returns to schooling parameter β^w, the midpoint of the 6 to 10 percent range reported in the literature (Card, 1999), but the results are not very sensitive to this choice. With $\beta^w = 0.06$, the bias-corrected

[25] Any effect would depend on the interaction of measurement error and ability bias in a regression of labor market earnings on the entrepreneur's schooling. Formally, let the noisy measure of potential earnings be given by $\ln w^*_{i,t} = \ln w_{i,t} + u_{i,t}$, where $u_{i,t}$ represents classical measurement error. Then the probability limit of the coefficient on $\ln w^*$ in (17) would equal $\frac{\lambda^e}{\lambda^w}\frac{\sigma_e^2}{\sigma_e^2+\sigma_u^2}$, where σ_e^2 is the variance of the residual from a regression of true potential earnings $\ln w$ on the remaining covariates in (17). The probability limit of the coefficient on s would equal $\beta^e - \frac{\lambda^e}{\lambda^w}\beta^w + \beta^{w^*}\frac{\lambda^e}{\lambda^w}(1-\frac{\sigma_e^2}{\sigma_e^2+\sigma_u^2})$, where β^{w^*} is the coefficient from a regression of $\ln w$ on s. β^{w^*} in turn would equal $\beta^w + \lambda^w\zeta$, where ζ is the coefficient from a regression of a on s. Applying the bias correction to the coefficient on s would therefore lead to a consistent estimate of

$$\beta^e - \frac{\lambda^e}{\lambda^w}\beta^w + (\beta^w + \lambda^w\zeta)\frac{\lambda^e}{\lambda^w}(1-\frac{\sigma_e^2}{\sigma_e^2+\sigma_u^2}) + \beta^w\frac{\lambda^e}{\lambda^w}\frac{\sigma_e^2}{\sigma_e^2+\sigma_u^2} = \beta^e + \lambda^w\zeta\frac{\lambda^e}{\lambda^w}(1-\frac{\sigma_e^2}{\sigma_e^2+\sigma_u^2}) \quad (18)$$

The literature on returns to schooling has found the ability bias term $\lambda^w\zeta$ to be small, on the order of 10 percent of β^w (Card, 1999), which implies that the bias term on the right-hand side of (18) will be minimal even if measurement error in the outside option is severe. For example, suppose that measurement error is such that $\frac{\sigma_e^2}{\sigma_e^2+\sigma_u^2} = 0.5$, which implies that the coefficient on the outside option in (17) is attenuated by 50 percent. Assuming a return to schooling of $\beta^w = 8\%$ and using the coefficient on $\ln w^*$ from column two in table 5, the bias on β^e would equal $0.08 \times 0.1 \times 0.3434 = 0.003$. This compares with an estimate for β^e of 0.0604 in column two of table 5.

coefficient drops to 0.0535, while with $\beta^w = 0.10$ it rises to 0.0672.

One limitation of this approach as mentioned above is the assumption of a single dimension of ability, common across occupations. If there is a component of ability that is specific to entrepreneurship, then the entrepreneur's labor market earnings cannot proxy for both general and entrepreneurial ability. Columns three and four repeat the same exercise adding a measure of ability that is specific to entrepreneurship, the number of prior occupations that the entrepreneur has worked in (Lazear, 2005). In Lazear's model, entrepreneurs benefit from being "jacks-of-all-trades" who are competent across a range of skills. As a proxy for a diverse skill set, Lazear uses the number of occupations an entrepreneur has had experience with in previous employment spells, and shows that this variable is a strong predictor of the choice to become an entrepreneur. Following the same method, I use information about each entrepreneur's past employment and the standardized occupational codes in the data to measure each entrepreneur's number of prior occupations. The coefficient on the number of prior occupations in column three is positive (0.0426) and highly significant ($t = 10.52$), which validates the approach. The coefficient on entrepreneur schooling is nearly unchanged at 0.0620, indicating that these are indeed different dimensions of ability. Column four adds potential labor market earnings, and both coefficients drop, which suggests that both components of ability have a positive impact on labor market earnings as well. Most importantly, the bias-corrected coefficient on entrepreneur schooling equals .0596, again only marginally lower than the estimate without controlling for potential earnings.

These OLS results characterize the average effect of entrepreneur schooling on productivity, but as shown in the previous section there is significant heterogeneity in the impact of schooling along the distribution. It is possible that ability bias is a more serious issue at the top of the distribution, where the coefficient on schooling is substantially higher. Columns six and seven perform the same exercise for the 99.1th percentile of the distribution, and columns seven and eight for the 99.9th percentile. In both cases, the results are the same as for the OLS regressions. At the 99.1th percentile, the baseline schooling coefficient equals 0.1225 and the bias-correct coefficient accounting for ability equals 0.1133. At the 99.9th percentile, the coefficients equal 0.1466 and 0.1526, respectively.

Put together these results show that the entrepreneur schooling coefficient is remarkably stable when accounting for ability, both at the mean and at the top of the distribution. This suggests that bias from omitted ability in the baseline estimates is unlikely to be a significant issue.

4.5. Selection into Entrepreneurship

The model in section 3 imposes exogenous entry and exit. In practice, individuals choose whether to select into entrepreneurship as a function of the value of the opportunity they would pursue and the value of their outside option in the labor market. Are differences in observed productivity across schooling levels driven by differences in selection? This section addresses this question by employing an approach developed by Combes et al. (2012) to distinguish agglomeration from selection effects of city size on firm productivity.

Combes et al. (2012) model agglomeration effects as a shift and dilation of an underlying productivity distribution of potential entrants, whereas selection generates an endogenous threshold which truncates that distribution from the left, as in Lucas (1978) or Melitz (2003). They then estimate the combination of a shift, dilation and truncation that can best explain differences in the observed productivity distribution across cities of different sizes, and find that truncation, and therefore selection, plays a very limited role in explaining those differences. I start by presenting an alternative version of the model in section 3 which parallels the Combes et al. (2012) framework. The effect of schooling is parametrized as a shift and dilation of an underlying productivity distribution, and selection into entrepreneurship truncates that distribution from the left. I then use their estimation method to evaluate the role of selection in accounting for differences in productivity across levels of schooling.

In the baseline version of the model there is an infinitely lived mass of agents who switch exogenously between entrepreneurship and employment. Suppose instead that every period a mass M of agents enters the economy, a fraction θ_s of which have schooling level s. Besides schooling, each entering agent is endowed with an entrepreneurial idea of quality q drawn from a cumulative distribution $G(q)$, which is common across schooling levels and constant over time. The effect of schooling on log firm productivity takes the form of a shift C_s and a dilation D_s operating on this common underlying distribution of idea quality

$$\ln A = C_s + D_s q \tag{19}$$

After observing q, agents choose whether to become entrepreneurs or workers in order to maximize expected lifetime income. As before, workers are employed by entrepreneurs and supply their human capital e^{rs} inelastically every period, earning $e^{rs}w$, while entrepreneurs earn the profits $\pi(s, q, \tau(q))$ of the firms they run. In this version, τ is assumed to be a deterministic function of q such that $\pi(s, q, \tau(q))$ is strictly increasing in q. In addition, existing workers and entrepreneurs exit the economy with exogenous probability δ every period, so that the steady-state mass of agents now equals $L = \frac{M}{\delta}$. All else in the model remains unchanged.

In steady-state, both wages and profits are constant, and so entering agents choose to become entrepreneurs if $\pi(s, q, \tau(q)) > e^{rs}w$. Since profits are strictly increasing in q, the optimal decision is to set a threshold q_s^* above which agents sort into entrepreneurship.

The equilibrium distribution of log productivity for each s will be given by

$$\mu_s'(\ln A) = \max\left[0, \frac{G\left(\frac{\ln A - C_s}{D_s}\right) - T_s}{1 - T_s}\right] \tag{20}$$

where $T_s \equiv G(q_s^*)$ is a left-truncation parameter which captures the effect of selection. The higher q_s^* is, the more truncated the distribution will be. Since the q_s^* are endogenous, differences in observed productivity distributions are no longer determined solely by the effect of schooling on productivity. Note, however, that if more educated entrepreneurs have a better outside option in the labor market in absolute terms, that does not necessarily imply stronger selection through a higher q_s^*. What matters is the outside option relative to the profits from entrepreneurship, which also depend on schooling.

The parameters C_s, D_s and T_s cannot be estimated because G is not observed, but as Combes et al. (2012) show it is possible to estimate the *relative* magnitude of these parameters for two observed distributions without making any assumptions on G. More precisely, given two observed distributions indexed by i and j, their method estimates the parameters $D \equiv \frac{D_i}{D_j}$, $C \equiv C_i - DC_j$ and $T \equiv \frac{T_i - T_j}{1 - T_j}$ that best explain the differences between them, in the sense of minimizing mean squared quantile differences.[26] The parameter T, in particular, reveals the extent to which one distribution is more truncated than the other, and in that sense it quantifies differences in the extent of selection between the two distributions. The key assumption, as Combes et al. (2012) emphasize, is that the underlying distribution G is common to all schooling levels.

I implement this method to compare the simulated productivity distributions for different levels of schooling estimated in section 4.3.[27] In particular, I estimate the shift, dilation and truncation that best explain the differences between the distribution for entrepreneurs with no schooling, on one side, and each of the distributions for entrepreneurs with six, nine, twelve and sixteen years of schooling on the other.[28] Visual inspection of the distributions in figure 11 suggests that truncation plays a minor role, if any, in explaining productivity differences across levels of schooling, and the results in table 6 confirm this. Column one says that the distribution of productivity for entrepreneurs with six years of schooling can best be approximated by right-shifting the distribution for entrepreneurs with no schooling by 0.3143 log points and dilating it by a factor of 1.0348, with no role for truncation. In fact the coefficient on truncation is a precisely estimated zero. Moreover, the approximation is highly accurate, with a pseudo-R^2 of 0.9857, the fraction of mean squared quantile differences between the two distributions explained by the transformation. Columns two and three paint a very similar picture for the productivity distributions of entrepreneurs with twelve and

[26]See Lemma 1 in their paper. I implement their method using the STATA package developed by Kondo (2017).

[27]Given the assumption of a common underlying distribution G, it is important to use these counterfactual distributions that hold the distribution of sectors and other firm characteristics constant, rather than the productivity distributions observed in the data.

[28]Following Combes et al. (2012), productivity levels are normalized so that average log productivity for the no schooling group equals zero. This ensures that the shift coefficient can be interpreted as the average difference in log productivity between the two schooling levels, without affecting the dilation or truncation coefficients.

sixteen years of schooling. The coefficients on S and D both increase with schooling, and the coefficient on truncation remains a precise zero, with equally high values for R^2.

These results suggest that selection plays essentially no role in explaining differences in productivity across schooling levels. Another issue is whether the q_s^* *change* meaningfully as other factors change, namely the distribution of schooling in the population. If that is the case, then aggregate counterfactuals based on the productivity distributions estimated above could be misleading. As with the OLS results above, I examine how stable these coefficients are over the sample period, during which Portugal experienced a two-year increase in average years of schooling and a financial crisis. Columns four to nine split the sample into two periods, from 2004 to 2009 and from 2010 to 2015, and repeat the exercise separately for each period.[29] I restrict the sample to firms aged nine or less, so that the age distribution in both periods in comparable. Because of this restriction, the coefficients in these two sub-samples are not comparable with those in the full sample, but they are comparable with each other. Column four compares the distributions for entrepreneurs with no schooling and with six years of schooling in the 2004 to 2009 period, and column seven makes the same comparison in 2010 to 2015. The estimated shift, dilation and truncation are nearly identical in both periods, with a precisely estimated zero in the case of truncation. The same holds for twelve (columns five and eight) and for sixteen years of schooling (columns six and nine).

One final point about the aggregate impact of selection may be made. The quantile coefficients in figure 9 show that entrepreneur schooling mainly affects the right tail of the productivity distribution, while the effect of selection is to truncate the left tail. Coupled with the fact that right tail firms have higher weight on aggregate productivity than left tail firms, due to their market shares, this suggests that changes in selection are unlikely to substantially change the effect of entrepreneur schooling on aggregate productivity.

[29] Specifically, I reestimate quantile regressions separately for each period and then simulate productivity distributions for each level of schooling in each period, following the method described in section 4.3. I exclude year effects from the quantile regressions so that the distributions may be constructed from a common set of covariates. I then implement the Combes et al. (2012) methodology separately for each period.

5. Aggregate Implications

To what extent can differences in entrepreneurial human capital account for differences in aggregate productivity and output across countries? This section uses the findings above and the model in section 3 to perform a simple development accounting exercise. Using equation (12) as a point of departure, I focus on the effect of schooling on TFP through its effect on $\tilde{A}$, and assign any cross-country variation in allocative efficiency E and the remaining terms in (12) to unexplained residual TFP. Denoting this residual by $\tilde{z}$, per capita aggregate output can be expressed as

$$\frac{Y}{L} = \tilde{z}\tilde{A}\left(\frac{K}{L}\right)^{\alpha}\left(\frac{H}{L}\right)^{1-\alpha} \tag{21}$$

which is the expression I focus on for development accounting. The difference relative to the standard expression (see Caselli, 2005, for a survey) is that the effect of schooling on $\tilde{A}$ is part of the contribution of human capital, not assigned to the residual. Under the assumptions made in section 3, this effect can be computed by combining the productivity distributions μ_s estimated in the Portuguese data with cross-country data on θ_s, the distribution of educational attainment in the population.

The literature has proposed two main accounting methods. In Caselli (2005), the fraction of cross-country income differences explained by human and physical capital is given by $\frac{\mathrm{Var}(\ln(\hat{Y}/L))}{\mathrm{Var}(\ln(Y/L))}$, where $\hat{Y}$ is counterfactual output in a factor-only model. Applied to (21), Caselli's approach yields

$$\frac{\mathrm{var}\left[\ln\left(\tilde{A}\left(\frac{K}{L}\right)^{\alpha}\left(\frac{H}{L}\right)^{1-\alpha}\right)\right]}{\mathrm{var}\left(\frac{Y}{L}\right)} \tag{22}$$

The method employed by Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999) (KRHJ) differs in two respects. First, starting from a variance decomposition of $\ln(Y/L)$, they also assign half of the covariance term between $\hat{Y}$ and residual TFP to the factor-only model. Second, they account for endogenous physical capital accumulation as a response to higher levels of residual TFP and human capital. Their method leads to

$$\frac{\text{var}\left[\ln\left(\tilde{A}^{\frac{1}{1-\alpha}}\left(\frac{K}{Y}\right)^{\frac{\alpha}{1-\alpha}}\frac{H}{L}\right)\right]+\text{cov}\left[\ln\left(\tilde{z}^{\frac{1}{1-\alpha}}\right),\ln\left(\tilde{A}^{\frac{1}{1-\alpha}}\left(\frac{K}{Y}\right)^{\frac{\alpha}{1-\alpha}}\frac{H}{L}\right)\right]}{\text{var}\left(\frac{Y}{L}\right)} \tag{23}$$

Both methods require data on cross-country output, physical and human capital per worker. To facilitate comparison, I use the dataset from Caselli (2005), who computes output and physical capital per worker from Penn World Tables data (Heston et al., 2002) and human capital per worker from the educational attainment data in Barro and Lee (2001) for 94 countries in 1996. To compute $\tilde{A}$ for each country, I also use the Barro and Lee (2001) data to obtain θ_s for each level of schooling, combined with the corresponding μ_s distributions estimated in section 4.3.[30] A key parameter in this analysis is σ, which determines the elasticity of value added with respect to productivity. In the firm-level results presented above, I follow Hsieh and Klenow (2014) in setting $\sigma = 3$ as a baseline. Given the central role of this parameter, this section presents results for $\sigma = 3$, $\sigma = 4$ and $\sigma = 5$, in line with the range of values used in similar exercises the literature.[31]

The first row in panel A of table 7 starts by replicating the results in Caselli (2005), when $\tilde{A}$ is assigned to the TFP residual. Physical and human capital account for 39% of per capita income differences in this base case. The remaining columns present the results under Caselli's method, as a function of σ. Accounting for the effect of entrepreneur schooling on firm productivity increases the fraction of income differences explained by human and physical capital to 76% when $\sigma = 3$, 70% when $\sigma = 4$ and 65% when $\sigma = 5$. The increase declines with σ, as expected, but is very substantial in all three cases. The second row in the panel repeats the exercise under KRHJ's method, and finds very similar results. The increase is from 40% in the base case, when $\tilde{A}$ is assigned to the residual, to 74% when $\sigma = 3$, 70% when $\sigma = 4$ and 66% when $\sigma = 5$. These magnitudes suggest that productivity could be the main channel for the effect

[30] Barro and Lee (2001) assume that the schooling levels reported in their data correspond to zero, three, six, nine, twelve, fourteen and sixteen years of schooling, and I follow their assumptions when computing μ_s. In addition, the Barro and Lee (2001) data are available at five-year intervals and I follow Caselli (2005) in using the data for the population 25 and older from 1995. As he notes, education is highly persistent and a one-year difference is unlikely to affect the results.

[31] Hsieh and Klenow (2009, 2014) set $\sigma = 3$, and Hsieh and Klenow (2009) also examine robustness for $\sigma = 5$. Bollard et al. (2016) set $\sigma = 4$. These choices are consistent with the median estimates in the literature(Broda and Weinstein, 2006).

of human capital on output, supporting the hypothesis put forth by Nelson and Phelps (1966).

As pointed out above, the effect of entrepreneur schooling on mean firm productivity does not fully capture its aggregate effect, because aggregate productivity is a power mean of firm productivity with exponent $\sigma - 1$. This reflects the fact that more productive firms have higher market shares in equilibrium and weigh more heavily in aggregate outcomes. The stronger effect of schooling on the upper tail of the productivity distribution documented in section 4.3 suggests that this mechanism could be a particularly important channel for the aggregate effect of schooling. To gauge the relative importance of the mean and upper tail effects of schooling, I decompose $\tilde{A}$ as follows:

$$\tilde{A} = \bar{A} \times \frac{\tilde{A}}{\bar{A}} \tag{24}$$

where

$$\bar{A} \equiv \exp\left[\sum_s \theta_s \int \ln(A)\mu_s(A)dA\right]$$

Here $\bar{A}$ is the geometric mean of firm productivity, which corresponds to aggregate productivity in the limit when $\sigma \to 1$ and market shares are equalized across firms, eliminating the upper tail channel. The results from this decomposition are reported in Panel B of table 7. The effect of entrepreneur schooling on mean productivity accounts for between 42% and 50% of the effect, and the additional effect on the upper tail for the remainder. The effect working through the mean declines with σ, since firm-level productivity differences inferred from equation (7) decline with σ, but this is partly offset by a stronger contribution from the upper tail, since the relative market shares of upper tail firms increase with the power mean exponent $\sigma - 1$. This offsetting effect makes the impact of entrepreneur schooling at the aggregate level less sensitive to the choice of σ than at the firm-level. These results highlight the crucial importance of accounting for the heterogeneous effect of entrepreneur schooling along with productivity distribution, and for the stronger effect in the right tail in particular.

6. Conclusion

The evidence presented suggests that the human capital of entrepreneurs is a key ingredient for the emergence of the fast growing, highly productive firms that are associated with development. Employment, output and productivity all increase substantially with entrepreneur schooling, and the relationship does not appear to be driven by omitted ability or selection bias. It is also strongest in the upper tail of the distribution, among the firms that impact aggregate productivity the most. Non-entrepreneurial human capital, on the other hand, seems to matter much less for firm dynamics.

A simple development accounting exercise shows that accounting for the effect of entrepreneurial human capital on firm productivity can increase the fraction of cross-country income differences explained by human and physical capital from 40% to between 64% and 76%, which strongly supports the Nelson and Phelps (1966) hypothesis that productivity is a key channel for the effect of human capital on output.

On top of highlighting the vital role of education in development, these findings suggest that attracting educated people into entrepreneurship, as opposed to rent seeking, is crucial, as argued by Baumol (1990) and Murphy et al. (1991). In addition, facilitating the allocation of educated entrepreneurs to the most promising entrepreneurial projects, for example through improvements in contract enforcement or the development of financial markets, could also have important implications for growth (Caselli and Gennaioli, 2013).

7. References

Acemoglu, D., P. Antràs, and E. Helpman (2007). Contracts and Technology Adoption. *The American Economic Review 97*(3), 916–943.

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2017, November). Missing Growth from Creative Destruction. Working Paper 24023, National Bureau of Economic Research.

Akcigit, U., H. Alp, and M. Peters (2018). Lack of Selection and Limits to Delegation: Firms Dynamics in Developing Countries. Technical report, National Bureau of Economic Research.

Albuquerque, R. and H. A. Hopenhayn (2004). Optimal lending contracts and firm dynamics. *The Review of Economic Studies 71*(2), 285–315.

Azoulay, P., B. Jones, J. D. Kim, and J. Miranda (2018). Age and High-Growth Entrepreneurship. Technical report, National Bureau of Economic Research.

Barro, R. J. and J.-W. Lee (2001). International Data on Educational Attainment: Updates and Implications. *Oxford Economic Papers 53*(3), 541–563.

Baumol, W. J. (1990). Entrepreneurship: Productive, Unproductive, and Destructive. *Journal of Political Economy 98*(5 Pt 1).

Bento, P. and D. Restuccia (2017, July). Misallocation, Establishment Size, and Productivity. *American Economic Journal: Macroeconomics 9*(3), 267–303.

Bloom, N., R. Sadun, and J. Van Reenen (2012, January). The Organization of Firms Across Countries. *The Quarterly Journal of Economics 127*(4), 1663–1705.

Bloom, N. and J. Van Reenen (2007). Measuring and Explaining Management Practices across Firms and Countries. *The Quarterly Journal of Economics*, 1351–1408.

Bollard, A., P. J. Klenow, and H. Li (2016). Entry costs rise with development. *Stanford Center for International Development Working Paper*.

Braguinsky, S., L. G. Branstetter, and A. Regateiro (2011). The incredible shrinking Portuguese firm. Technical report, National Bureau of Economic Research.

Broda, C. and D. E. Weinstein (2006). Globalization and the Gains From Variety. *The Quarterly journal of economics 121*(2), 541–585.

Buchinsky, M. (1994, March). Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression. *Econometrica 62*(2), 405.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics 3*, 1801–1863.

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of economic growth 1*, 679–741.

Caselli, F. and A. Ciccone (2013). The contribution of schooling in development accounting: Results from a nonparametric upper bound. *Journal of Development Economics 104*, 199–211.

Caselli, F. and N. Gennaioli (2013). Dynastic management. *Economic Inquiry 51*(1), 971–996.

Clementi, G. L. and H. A. Hopenhayn (2006). A theory of financing constraints and firm dynamics. *The Quarterly Journal of Economics*, 229–265.

Cole, H. L., J. Greenwood, and J. M. Sanchez (2016, July). Why Doesn't Technology Flow From Rich to Poor Countries? *Econometrica 84*(4), 1477–1521.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica 80*(6), 2543–2594.

Cooley, T. F. and V. Quadrini (2001). Financial markets and firm dynamics. *American Economic Review*, 1286–1310.

Córdoba, J. C. and M. Ripoll (2008, September). Endogenous TFP and cross-country income differences. *Journal of Monetary Economics 55*(6), 1158–1170.

Erosa, A., T. Koreshkova, and D. Restuccia (2010, October). How Important Is Human Capital? A Quantitative Theory Assessment of World Income Inequality. *The Review of Economic Studies 77*(4), 1421–1449.

Eslava, M. and J. Haltiwanger (2018, May). The Life-Cycle Growth of Plants: The Role of Productivity, Demand and Distortions. SSRN Scholarly Paper ID 3177289, Social Science Research Network, Rochester, NY.

Eslava, M., J. Haltiwanger, and A. Pinzón (2018). Job Creation in Colombia vs the U.S.: "Up or Out Dynamics" Meets "the Life Cycle of Plants". *Working Paper*.

Evans, D. S. and B. Jovanovic (1989). An estimated model of entrepreneurial choice under liquidity constraints. *The Journal of Political Economy*, 808–827.

Foster, L., J. Haltiwanger, and C. Syverson (2008, March). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *The American Economic Review 98*(1), 394–425. ArticleType: research-article / Full publication date: Mar., 2008 / Copyright © 2008 American Economic Association.

Garcia-Macia, D., C.-T. Hsieh, and P. J. Klenow (2016). How destructive is innovation? Technical report, National Bureau of Economic Research.

Gennaioli, N., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer (2013). Human Capital and Regional Development. *The Quarterly Journal of Economics 128*(1), 105–164.

Hall, R. E. and C. I. Jones (1999). Why Do Some Countries Produce So Much More Output Per Worker Than Others? *The Quarterly Journal of Economics 114*(1), 83–116.

Hendricks, L. and T. Schoellman (2018, May). Human Capital and Development Accounting: New Evidence from Wage Gains at Migration. *The Quarterly Journal of Economics 133*(2), 665–700.

Heston, A., R. Summers, and B. Aten (2002). Penn World Table 6.1. *Center for International Comparisons at the University of Pennsylvania*.

Hopenhayn, H. A. (1992). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica 60*(5), 1127–1150. ArticleType: research-article / Full publication date: Sep., 1992 / Copyright © 1992 The Econometric Society.

Hottman, C. J., S. J. Redding, and D. E. Weinstein (2016, August). Quantifying the Sources of Firm Heterogeneity. *The Quarterly Journal of Economics 131*(3), 1291–1364.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics 124*(4), 1403–1448.

Hsieh, C.-T. and P. J. Klenow (2014). The Life Cycle of Plants in India and Mexico. *The Quarterly Journal of Economics 129*(3), 1035–1084.

Hsieh, C.-T. and P. J. Klenow (2017). The Reallocation Myth. *Fostering a Dynamic Global Economy*.

Jaeger, K. (2017). EU KLEMS Growth and Productivity Accounts 2017 Release - Description of Methodology and General Notes. *The Conference Board*.

Jones, B. F. (2014). The Human Capital Stock: A Generalized Approach. *The American Economic Review 104*(11), 3752–3777.

Jovanovic, B. (1982, May). Selection and the Evolution of Industry. *Econometrica 50*(3), 649–670. ArticleType: research-article / Full publication date: May, 1982 / Copyright © 1982 The Econometric Society.

King, R. G. and R. Levine (1993). Finance, entrepreneurship and growth. *Journal of Monetary economics 32*(3), 513–542.

Klenow, P. and A. Rodriguez-Clare (1997). The neoclassical revival in growth economics: Has it gone too far? In *NBER Macroeconomics Annual 1997, Volume 12*, pp. 73–114. MIT Press.

Klenow, P. J. and A. Rodríguez-Clare (2005). Chapter 11 Externalities and Growth. In *Handbook of Economic Growth*, Volume 1, pp. 817–861. Elsevier.

Klette, T. and S. Kortum (2004). Innovating firms and aggregate innovation. *Journal of Political Economy 112*(5), 986–1018.

Kondo, K. (2017, December). ESTQUANT: Stata Module to Implement Quantile Approach by Combes et al.

La Porta, R. and A. Shleifer (2008). The Unofficial Economy and Economic Development. *Brookings Papers on Economic Activity*.

Lagakos, D., B. Moll, T. Porzio, and N. Qian (2012). Experience matters: Human capital and development accounting. Technical report, National Bureau of Economic Research.

Lazear, E. P. (2005). Entrepreneurship. *Journal of Labor Economics 23*(4), 649–680.

Lemieux, T. (2006). The Mincer Equation 30 years after.

Lucas, Jr., R. E. (1978, October). On the Size Distribution of Business Firms. *The Bell Journal of Economics 9*(2), 508–523.

Machado, J. A. F. and J. Mata (2005, May). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics 20*(4), 445–465.

Mankiw, N. G., D. Romer, and D. N. Weil (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics 107*(2), 407–437.

Manuelli, R. E. and A. Seshadri (2014, September). Human Capital and the Wealth of Nations. *American Economic Review 104*(9), 2736–2762.

Martins, P. S. and P. T. Pereira (2004, June). Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour Economics 11*(3), 355–371.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Mincer, J. (1974). *Schooling, Experience, and Earnings.* Columbia University Press.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1991). The Allocation of Talent: Implications for Growth. *The Quarterly Journal of Economics 106*(2), 503–530.

Nelson, R. R. and E. S. Phelps (1966, March). Investment in Humans, Technological Diffusion, and Economic Growth. *The American Economic Review 56*(1/2), 69–75.

Olley, G. and A. Pakes (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica 64*(6), 1263–1297.

Parente, S. L. and E. C. Prescott (1994). Barriers to technology adoption and development. *Journal of political Economy 102*(2), 298–321.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics 11*(4), 707–720.

Schoellman, T. (2012, January). Education Quality and Development Accounting. *The Review of Economic Studies 79*(1), 388–417.

Schultz, T. W. (1975). The value of the ability to deal with disequilibria. *Journal of economic literature*, 827–846.

Welch, F. (1970). Education in production. *The Journal of Political Economy*, 35–59.

A. Appendix

Table 1. **Summary Statistics**

A. All Years (1995-2015)

	Cohorts Observed from Entry					Full Sample				
	Mean	SD	p10	p50	p90	Mean	SD	p10	p50	p90
Employment	7.002	12.04	2	4	14	9.648	24.63	1	4	19
Gross Output	586.8	2,004	33.56	156.6	1,122	1,036	5,318	34.18	187.1	1,781
Number of Entrepreneurs	1.422	1.322	1	1	2	1.526	1.677	1	1	2
Entrepreneur Schooling	9.358	4.314	4	9	17	8.955	4.487	4	9	17
Non-Entrepreneur Schooling	8.175	3.427	4	7.75	12	7.64	3.41	4	6.5	12
Entrepreneur Experience	23.03	11.13	9	22	38	32.13	13.74	14	32	50
Non-Entrepreneur Experience	19.02	10.12	7	18	32.51	26.29	11.6	12	25.5	41.92
Firm Age	4.987	4.318	0	4	11	13.2	12.6	2	10	29
Number of Observations			1,204,484					3,925,817		
Number of Firms			189,425					456,543		

B. Years with Value Added and Productivity Data (2004-2015)

	Cohorts Observed from Entry					Full Sample				
	Mean	SD	p10	p50	p90	Mean	SD	p10	p50	p90
Employment	8.37	13.91	2	5	16	11.46	27.03	2	5	22
Gross Output	691.3	2,221	48.14	192	1,335	1,219	5,805	51.19	233.7	2,102
Value Added	172.1	490.9	16.19	65.14	332.5	292.2	1,186	17.36	76.43	520.3
Physical Capital	147.9	624.9	2.257	29	277.8	316.9	1,712	2.393	38.17	497
Number of Entrepreneurs	1.437	.8739	1	1	2	1.556	1.226	1	1	2
Entrepreneur Schooling	9.558	4.328	4	9	17	9.395	4.471	4	9	17
Non-Entrepreneur Schooling	8.278	3.339	4	8	12	7.914	3.369	4	7.25	12
Entrepreneur Experience	22.99	11.08	9	22	38	33.27	13.86	15	33	52
Non-Entrepreneur Experience	19.07	9.755	7	18.13	32	28.04	11.28	13.75	27.62	42.95
Firm Age	6.058	4.451	1	5	13	14.14	12.81	2	11	30
Number of Observations			691,068					1,714,888		
Number of Firms			133,939					287,922		

Notes: This table presents summary statistics for the samples used in the analysis. Employment is the number of workers reported by the firm, including entrepreneurs and non-entrepreneurs, regardless of employment status and including unpaid workers. Gross output is the firm's reported revenue. Value added is also directly reported and is equal to gross output less intermediate inputs. Physical capital is the book value of the firm's assets, including both tangible and intangible assets. Gross output, value added and physical capital are in thousands of 2011 euros. The number of entrepreneurs includes workers classified as entrepreneurs as described in section 2. Entrepreneur and non-entrepreneur schooling and experience corresponds to average years of schooling and potential experience for each group of workers, where experience is defined as age at entry, minus years of schooling, minus six, plus firm age. Firm age is based on the firm's reported year of incorporation.

Table 2. **Entrepreneur Schooling and Productivity**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Entrepreneur Schooling	0.0554	0.0536	0.0603	0.0490	0.0573	0.0572	0.0424	0.0359
	(0.0012)	(0.0008)	(0.0015)	(0.0013)	(0.0011)	(0.0012)	(0.0011)	(0.0010)
Non-Entrepreneur Schooling	0.0083	0.0141	0.0130	-0.0239		0.0088	0.0049	0.0032
	(0.0016)	(0.0011)	(0.0021)	(0.0016)		(0.0016)	(0.0014)	(0.0013)
Entrepreneur Experience	0.0202	0.0220	0.0198	0.0193	0.0241	0.0143	0.0173	0.0159
	(0.0012)	(0.0007)	(0.0014)	(0.0012)	(0.0012)	(0.0011)	(0.0010)	(0.0009)
Entrepreneur Experience^2	-0.0003	-0.0002	-0.0002	-0.0003	-0.0003	-0.0001	-0.0002	-0.0002
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Non-Entrepreneur Experience	0.0388	0.0456	0.0366	0.0255		0.0394	0.0287	0.0237
	(0.0012)	(0.0008)	(0.0015)	(0.0012)		(0.0012)	(0.0011)	(0.0010)
Non-Entrepreneur Experience^2	-0.0007	-0.0007	-0.0006	-0.0006		-0.0007	-0.0005	-0.0004
	(0.0000)	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
Log Number of Entrepreneurs						0.4367		
						(0.0087)		
Firm Age Quartic	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Observations	691,068	1,714,832	464,818	691,068	691,068	691,068	691,068	691,068
R^2	0.600	0.603	0.633	0.576	0.597	0.607	0.651	0.677

Notes: This table reports OLS regressions of log productivity ($\ln A$) on entrepreneur schooling for the sample of cohorts observed from entry. All regressions include a quartic in firm age (except column three), sector fixed effects and year fixed effects, in addition to the coefficients reported. Column one presents the baseline specification. Column two estimates the same specification on the full sample, including firms not observed from entry and for whom entrepreneurs are defined as the first top managers observed in the data. Column three restricts the sample to firms that survived until 2015. Columns four and five omit controls for firm age and for non-entrepreneur schooling and experience, respectively. Column six adds the log of the number of entrepreneurs. Columns seven and eight set $\sigma = 4$ and $\sigma = 5$, respectively, instead of the baseline $\sigma = 3$. Errors are clustered at the firm level.

Table 3. **Entrepreneur Schooling and Other Outcomes**

	(1)	(2)	(3)	(4)	(5)
Entrepreneur Schooling	0.0778	0.0545	0.0840	0.0165	0.0003
	(0.0014)	(0.0013)	(0.0022)	(0.0008)	(0.0001)
Non-Entrepreneur Schooling	0.0203	0.0284	0.0038	-0.0019	-0.0004
	(0.0017)	(0.0015)	(0.0026)	(0.0011)	(0.0002)
Entrepreneur Experience	0.0172	0.0046	0.0064	0.0116	0.0011
	(0.0013)	(0.0011)	(0.0020)	(0.0007)	(0.0001)
Entrepreneur Experience^2	-0.0002	-0.0000	-0.0002	-0.0001	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Non-Entrepreneur Experience	0.0607	0.0566	0.0438	0.0085	-0.0006
	(0.0013)	(0.0011)	(0.0020)	(0.0008)	(0.0001)
Non-Entrepreneur Experience^2	-0.0013	-0.0013	-0.0011	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	691,068	691,068	690,056	691,068	630,878
R^2	0.182	0.152	0.129	0.745	0.010

Notes: This table reports OLS regressions of additional firm outcomes on entrepreneur schooling for the sample of cohorts observed from entry. The dependent variables are log value added in column one, log total worker human capital in column two, log physical capital in column three, log average revenue product of inputs (τ) in column four and survival in column five. Column five excludes observations from 2015, the last year in the sample, for which survival is not observed. All regressions include a quartic in firm age, sector fixed effects and year fixed effects, in addition to the coefficients reported. Errors are clustered at the firm level.

Table 4. **Entrepreneur Schooling and Firm Outcomes by Year**

	(1)	(2)	(3)	(4)	(5)	(6)
Entrepreneur Schooling × 2004	0.0544	0.0766	0.0528	0.0778	0.0161	-0.0001
	(0.0017)	(0.0017)	(0.0014)	(0.0024)	(0.0011)	(0.0003)
Entrepreneur Schooling × 2005	0.0533	0.0758	0.0516	0.0830	0.0153	0.0001
	(0.0017)	(0.0017)	(0.0014)	(0.0024)	(0.0011)	(0.0003)
Entrepreneur Schooling × 2006	0.0564	0.0784	0.0522	0.0862	0.0171	0.0005
	(0.0017)	(0.0017)	(0.0014)	(0.0024)	(0.0011)	(0.0002)
Entrepreneur Schooling × 2007	0.0541	0.0762	0.0511	0.0866	0.0160	0.0006
	(0.0017)	(0.0017)	(0.0014)	(0.0024)	(0.0011)	(0.0003)
Entrepreneur Schooling × 2008	0.0542	0.0770	0.0526	0.0883	0.0157	0.0009
	(0.0017)	(0.0017)	(0.0014)	(0.0024)	(0.0011)	(0.0003)
Entrepreneur Schooling × 2009	0.0579	0.0807	0.0547	0.0938	0.0176	-0.0003
	(0.0017)	(0.0017)	(0.0014)	(0.0025)	(0.0011)	(0.0003)
Entrepreneur Schooling × 2010	0.0587	0.0826	0.0571	0.0946	0.0175	0.0003
	(0.0017)	(0.0017)	(0.0014)	(0.0026)	(0.0011)	(0.0003)
Entrepreneur Schooling × 2011	0.0579	0.0791	0.0555	0.0829	0.0183	0.0012
	(0.0019)	(0.0018)	(0.0015)	(0.0027)	(0.0012)	(0.0003)
Entrepreneur Schooling × 2012	0.0619	0.0785	0.0528	0.0751	0.0227	0.0009
	(0.0020)	(0.0019)	(0.0016)	(0.0029)	(0.0013)	(0.0003)
Entrepreneur Schooling × 2013	0.0614	0.0775	0.0513	0.0744	0.0226	0.0009
	(0.0020)	(0.0019)	(0.0016)	(0.0029)	(0.0014)	(0.0003)
Entrepreneur Schooling × 2014	0.0609	0.0774	0.0530	0.0710	0.0222	-0.0000
	(0.0020)	(0.0019)	(0.0016)	(0.0029)	(0.0014)	(0.0003)
Entrepreneur Schooling × 2015	0.0538	0.0742	0.0539	0.0730	0.0167	
	(0.0020)	(0.0019)	(0.0016)	(0.0029)	(0.0013)	
Observations	535,957	535,957	535,957	535,125	535,957	499,974
R^2	0.599	0.175	0.147	0.129	0.743	0.010

Notes: This table reports OLS regressions of firm outcomes on entrepreneur schooling interacted with year dummies for the sample of cohorts observed from entry. The dependent variables are log productivity ($\ln A$) in column one, value added in column two, log total worker human capital in column three, log physical capital in column four, log average revenue product of inputs (τ) in column five and survival in column six. The sample is restricted to firms aged 9 or less, so that the firm age distribution by year is comparable. Column six excludes observations from 2015, the last year in the sample, for which survival is not observed. All regressions include a quartic in firm age, sector fixed effects and year fixed effects, in addition to the coefficients reported. Errors are clustered at the firm level.

Table 5. **Accounting for Ability**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Entrepreneur Schooling	0.0641	0.0329	0.0620	0.0325	0.1225	0.0641	0.1466	0.0981
	(0.0018)	(0.0019)	(0.0018)	(0.0019)	(0.0070)	(0.0063)	(0.0105)	(0.0165)
Log last wage		0.3434		0.3388		0.6151		0.6811
		(0.0086)		(0.0087)		(0.0303)		(0.0808)
Bias-corrected Entrep. Sch.		0.0604		0.0596		0.1133		0.1526
($\beta w = 8\%$)		(0.0017)		(0.0017)		(0.0056)		(0.0120)
Number of Prior Occupations			0.0426	0.0189				
			(0.0040)	(0.0040)				
Non-Entrepreneur Schooling	0.0083	0.0010	0.0075	0.0008	0.0670	0.0351	0.1208	0.0680
	(0.0023)	(0.0022)	(0.0023)	(0.0022)	(0.0094)	(0.0081)	(0.0173)	(0.0143)
Entrepreneur Experience	0.0165	0.0055	0.0136	0.0043	0.0068	-0.0189	-0.0044	-0.0179
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0087)	(0.0073)	(0.0139)	(0.0174)
Entrepreneur Experience^2	-0.0001	-0.0000	-0.0001	-0.0000	0.0002	0.0005	0.0004	0.0004
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0001)	(0.0002)	(0.0003)
Non-Entrepreneur Experience	0.0423	0.0402	0.0420	0.0401	0.0558	0.0447	0.0523	0.0484
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0068)	(0.0052)	(0.0115)	(0.0155)
Non-Entrepreneur Experience^2	-0.0008	-0.0007	-0.0007	-0.0007	-0.0008	-0.0006	-0.0008	-0.0008
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Observations	340,083	340,083	340,083	340,083	340,083	340,083	340,083	340,083

Notes: This table presents OLS and quantile regressions of log productivity on entrepreneur schooling for the sample of entrepreneurs observed in other occupations before becoming entrepreneurs. Log last wage is the entrepreneur's income in the last occupation before becoming an entrepreneur. The bias-corrected coefficient equals the coefficient on entrepreneur schooling plus an assumed labor market return to schooling of 8% multiplied by the coefficient on log last wage (see main text for details). Columns one to four are OLS regressions. The number of prior occupations in columns three and four is the number of past occupations the entrepreneur has held before becoming an entrepreneur. Columns five and six are quantile regressions for the 99.1th percentile, and columns seven and eight are quantile regressions for the 99.9th percentile. All regressions include a quartic in firm age, sector fixed effects and year fixed effects. Standard errors are clustered at the firm level.

Table 6. **The Role of Selection**

	2004-2015			2004-2009			2010-2015		
Entrep. Sch:	0 vs 6	0 vs 12	0 vs 16	0 vs 6	0 vs 12	0 vs 16	0 vs 6	0 vs 12	0 vs 16
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Shift S	0.3143	0.6286	0.8380	0.3164	0.6326	0.8436	0.3248	0.6497	0.8662
	(0.0023)	(0.0025)	(0.0024)	(0.0343)	(0.0304)	(0.0386)	(0.0358)	(0.0388)	(0.0590)
Dilation D	1.0348	1.0724	1.0989	1.0089	1.0191	1.0257	1.0060	1.0124	1.0170
	(0.0053)	(0.0054)	(0.0048)	(0.1373)	(0.1221)	(0.1330)	(0.1367)	(0.1391)	(0.1628)
Truncation T	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0004)
R^2	0.9857	0.9859	0.9861	0.9592	0.9571	0.9558	0.9707	0.9688	0.9678

Notes: This table displays estimates of the shift, dilation and truncation parameters that can best explain differences in productivity distributions across levels of entrepreneur schooling, following the approach developed by Combes et al. (2012). Each column compares the log productivity distribution for entrepreneurs with no schooling with the distribution for entrepreneurs with the schooling level indicated in that column. The first three columns perform these comparisons on the entire sample of cohorts observed from entry. Columns four to nine divide this sample into the sub-periods indicated in the table and further restrict the sample to firms aged 9 or less, so that the age distribution is comparable in the two sub-periods. The productivity distributions for each level of schooling in each period are simulated from quantile regressions using the method of Machado and Mata (2005), as described in section 4.3. Productivity levels are normalized such that average log productivity for entrepreneurs with no schooling equals zero. The estimation is performed using the STATA package developed by Kondo (2017). The standard errors are bootstrapped and the R^2 measures the fraction of mean squared quantile differences between the two distributions accounted for by the transformation.

Table 7. **Development Accounting**

A. Overall Results

	Baseline	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$
Caselli	0.3853	0.7594	0.7004	0.6549
KRHJ	0.3968	0.7427	0.6971	0.6599

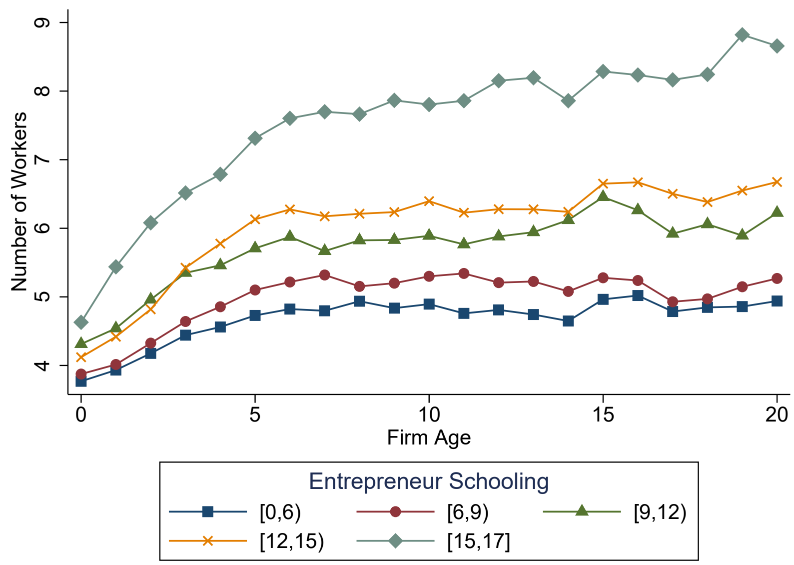
B. Decomposition

	Caselli			KRHJ		
	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$
Effect on Mean Productivity	0.5716	0.5252	0.5018	0.5669	0.5279	0.5084
	(0.4980)	(0.4439)	(0.4321)	(0.4917)	(0.4368)	(0.4243)
Additional Effect on Upper Tail	0.5731	0.5605	0.5384	0.5727	0.5659	0.5483
	(0.5020)	(0.5561)	(0.5679)	(0.5083)	(0.5632)	(0.5757)
Total Effect	0.7594	0.7004	0.6549	0.7427	0.6971	0.6599
	(1.0000)	(1.0000)	(1.0000)	(1.0000)	(1.0000)	(1.0000)

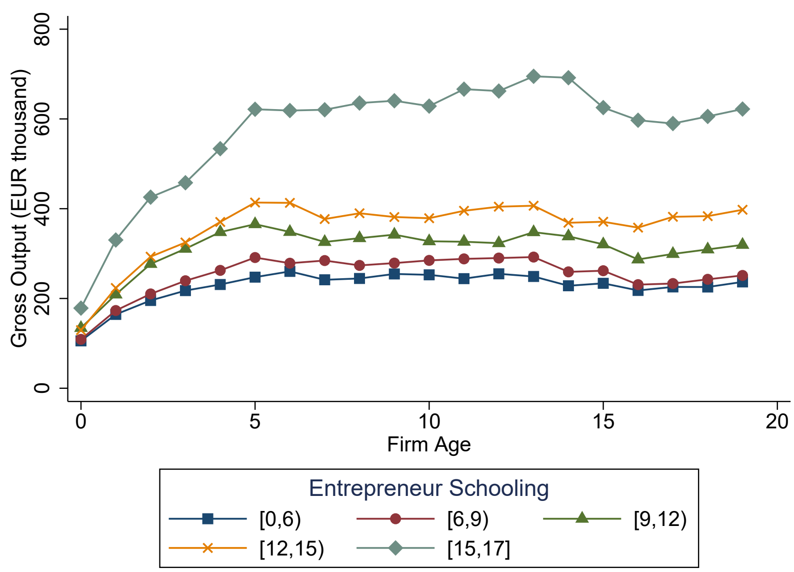
Notes: This table reports results from a development accounting exercise incorporating the effect of entrepreneur schooling on aggregate TFP, under the methods developed by Caselli (2005) and by Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999) (KRHJ). Panel A presents baseline results, excluding the effect of entrepreneur schooling on TFP, along with overall results calculated under different assumptions for σ (see main text for details). Panel B decomposes the overall impact into an effect on mean firm productivity and an additional effect on the upper tail of the productivity distribution. In panel B, the numbers in parentheses indicate the fraction of the overall effect explained by each channel.

Figure 1: **Firm Life Cycle Dynamics for the 1995 Cohort**

(a) **Employment by Entrepreneur Schooling**



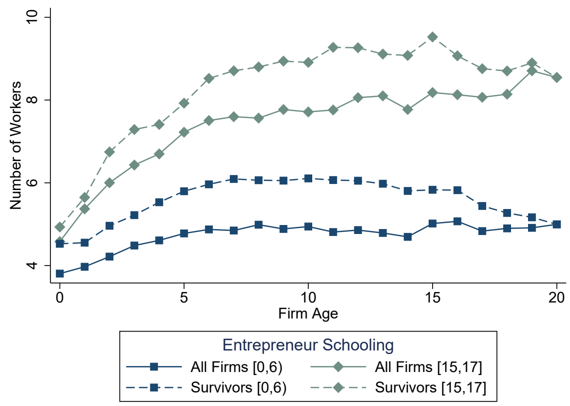
(b) **Gross Output by Entrepreneur Schooling**



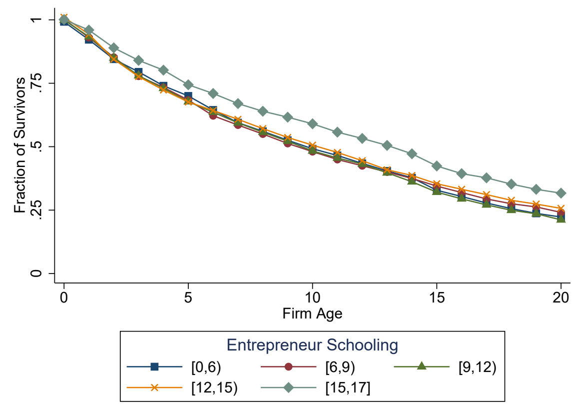
Notes: These graphs plot mean outcomes by firm age for firms in the 1995 cohort, sorting firms into five groups by average entrepreneur years of schooling. The top panel plots employment, and the bottom panel plots gross output. The estimates are conditional on average non-entrepreneur schooling, quadratics in average entrepreneur and non-entrepreneur experience, as well as sector fixed effects, and are evaluated at the sample mean of these covariates.

Figure 2: **Firm Life Cycle Dynamics for the 1995 Cohort (cont.)**

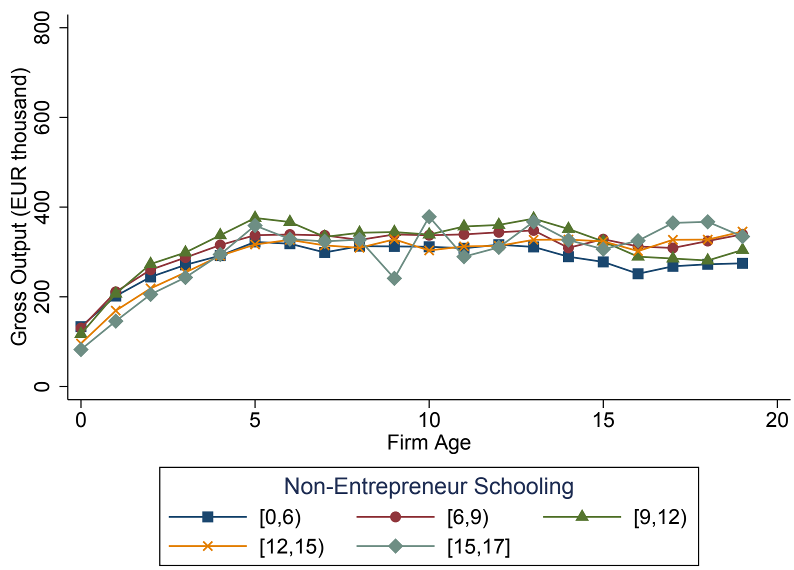
(a) **Employment for Survivors vs All Firms**



(b) **Survival by Entrepreneur Schooling**

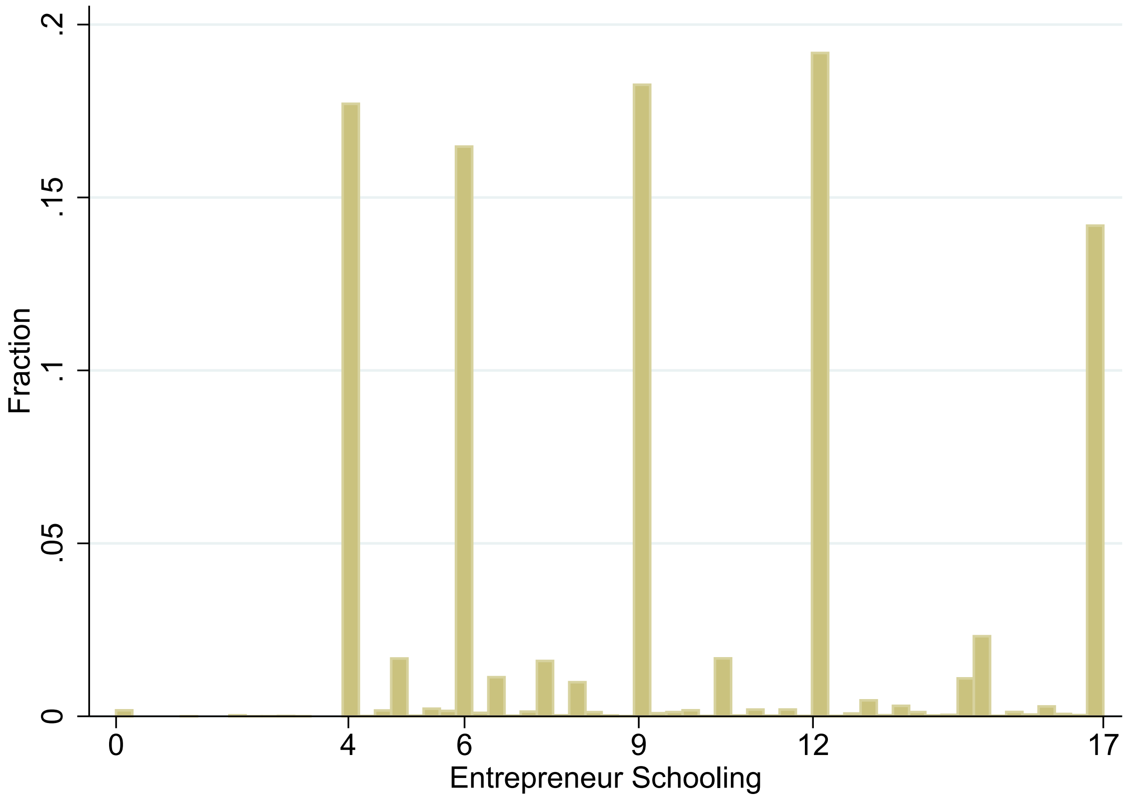


(c) **Gross Output by Non-Entrepreneur Schooling**



Notes: These graphs plot mean outcomes by firm age for firms in the 1995 cohort. The top left panel plots employment for survivors and for all firms, for firms with 0 to 6 years of average entrepreneur schooling and for firms with at least 15 years of average entrepreneur schooling. The top right panel sorts firms into five groups by average entrepreneur years of schooling, and plots cumulative survival rates for each group. The bottom panel sorts firms into five groups by average non-entrepreneur years of schooling, and plots gross output for each group. All estimates are conditional on average non-entrepreneur schooling, quadratics in average entrepreneur and non-entrepreneur experience, as well as sector fixed effects, and are evaluated at the sample mean of these covariates.

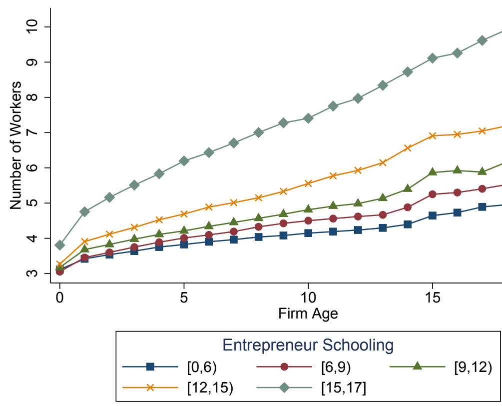
Figure 3: **Histogram of Entrepreneur Education**



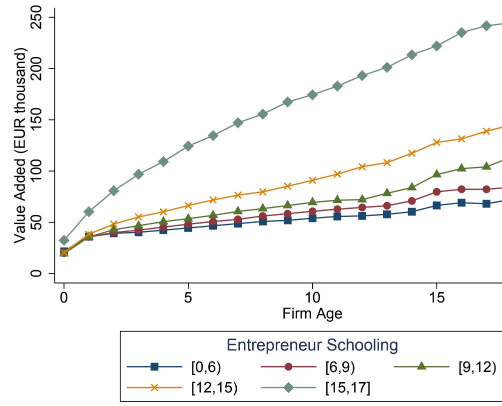
Notes: This graph plots a histogram of average entrepreneur education for firms in the sample of cohorts observed from entry. Observations are unweighted. The five points at which most firms are concentrated correspond to the five main education levels reported in the data: 4th grade, 6th grade, 9th grade, 12th grade and the *licenciatura* higher education degree.

Figure 4: **Firm Life Cycle Dynamics for Cohorts Observed from Entry**

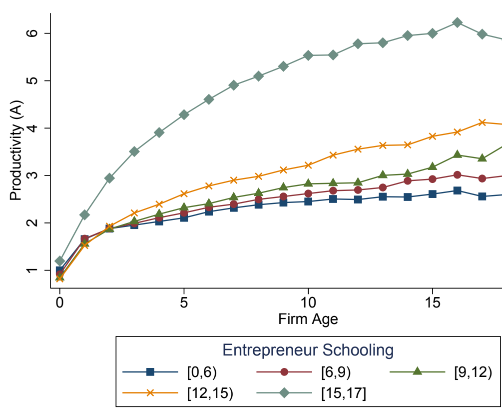
(a) **Employment by Entrepreneur Schooling**



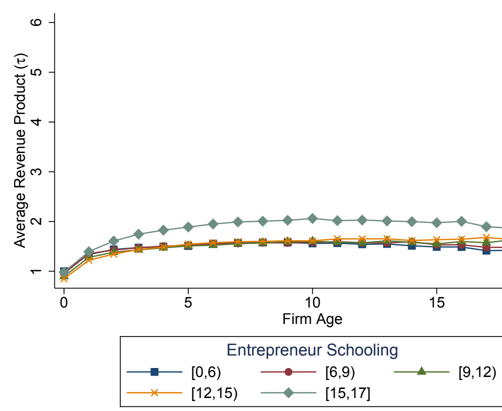
(b) **Value Added by Entrepreneur Schooling**



(c) **Productivity by Entrepreneur Schooling**

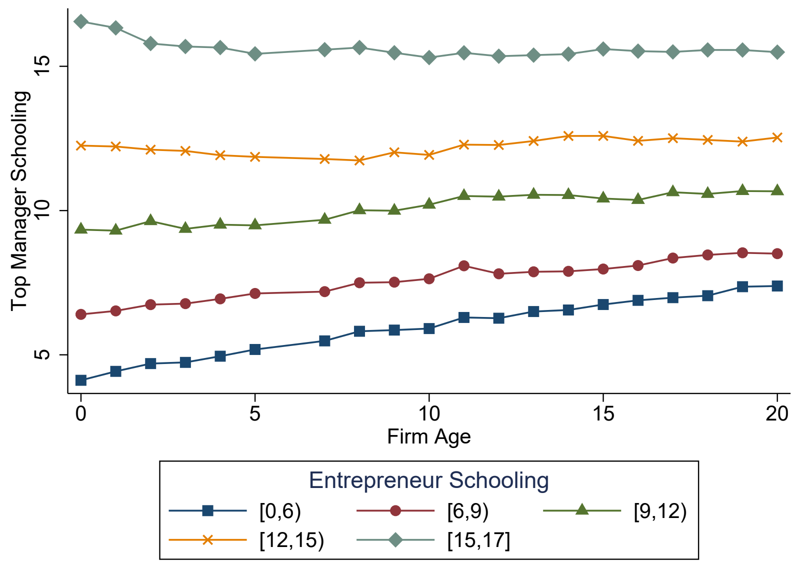


(d) **Avg. Rev. Product by Entrepreneur Schooling**



Notes: These graphs plot mean outcomes by firm age for the pooled sample of cohorts observed from entry, sorting firms into five groups by average entrepreneur years of schooling. The top left panel plots employment, the top right panel plots value added, the bottom left panel plots productivity (A) and the bottom right panel plots the average revenue product of inputs (τ). All estimates are conditional on average non-entrepreneur schooling, quadratics in average entrepreneur and non-entrepreneur experience, sector fixed effects and year fixed effects, and evaluated at the sample mean of these covariates.

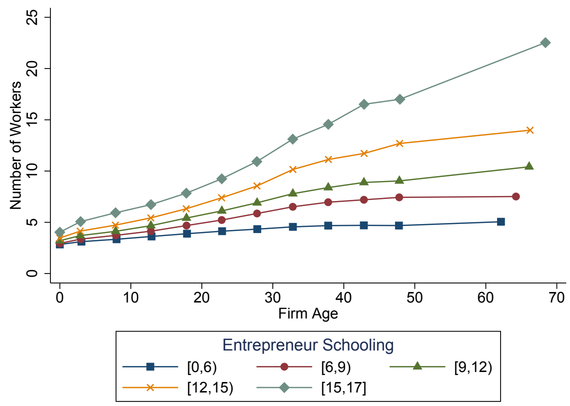
Figure 5: **Persistence of Top Manager Schooling in the 1995 Cohort**



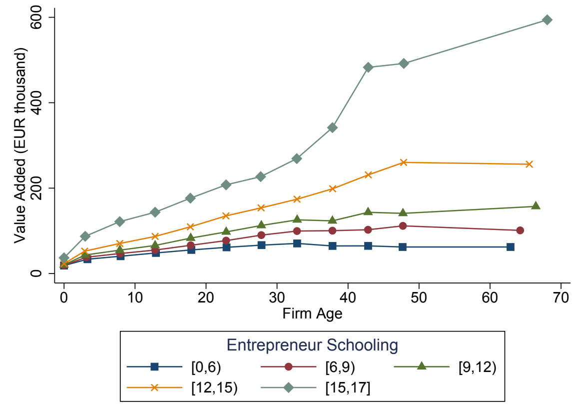
Notes: This graph plots average top manager schooling by firm age for firms in the 1995 cohort who survived until 2015, sorting firms into five groups by average entrepreneur years of schooling.

Figure 6: **Firm Life Cycle Dynamics for the Full Sample**

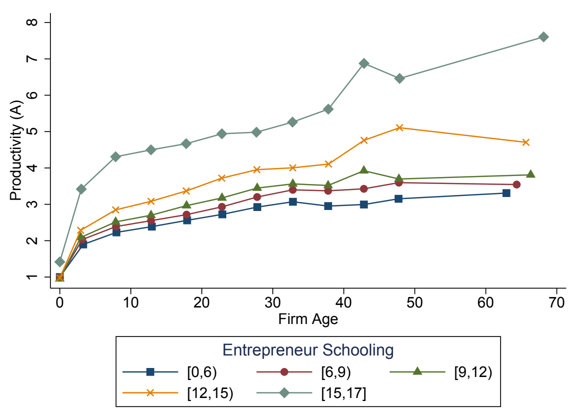
(a) **Employment by Entrepreneur Schooling**



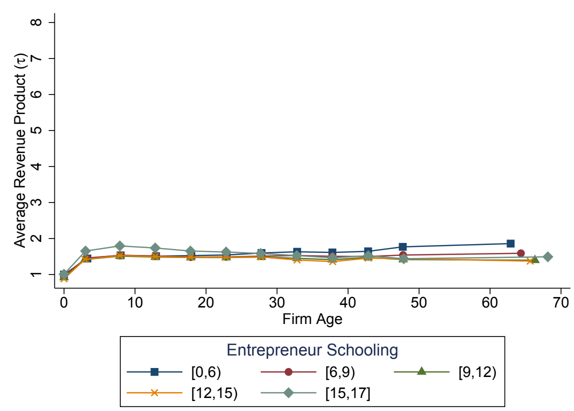
(b) **Value Added by Entrepreneur Schooling**



(c) **Productivity by Entrepreneur Schooling**



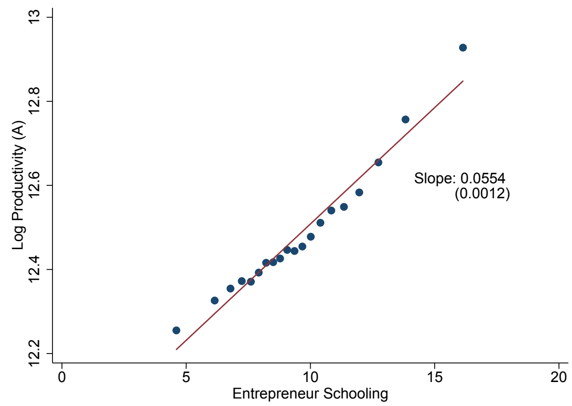
(d) **Avg. Rev. Product by Entrepreneur Schooling**



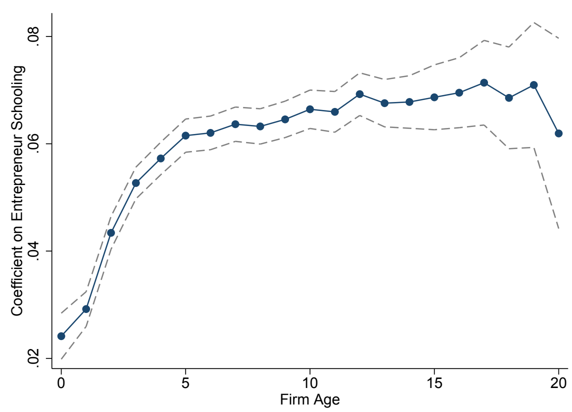
Notes: These graphs plot mean outcomes by firm age for the full sample, sorting firms into five groups by average entrepreneur years of schooling. Firms are grouped into 5-year age bins, plus a separate bin for entrants and one for all firms 50 or older. Entrepreneurs in this sample are defined as the first top managers observed in the data, not necessarily at entry. The top left panel plots employment, the top right panel plots value added, the bottom left panel plots productivity (A) and the bottom right panel plots the average revenue product of inputs (τ). All estimates are conditional on average non-entrepreneur schooling, quadratics in average entrepreneur and non-entrepreneur experience, sector fixed effects and year fixed effects, and evaluated at the sample mean of these covariates.

Figure 7: **Functional Form**

(a) **Log Productivity and Entrepreneur Schooling**

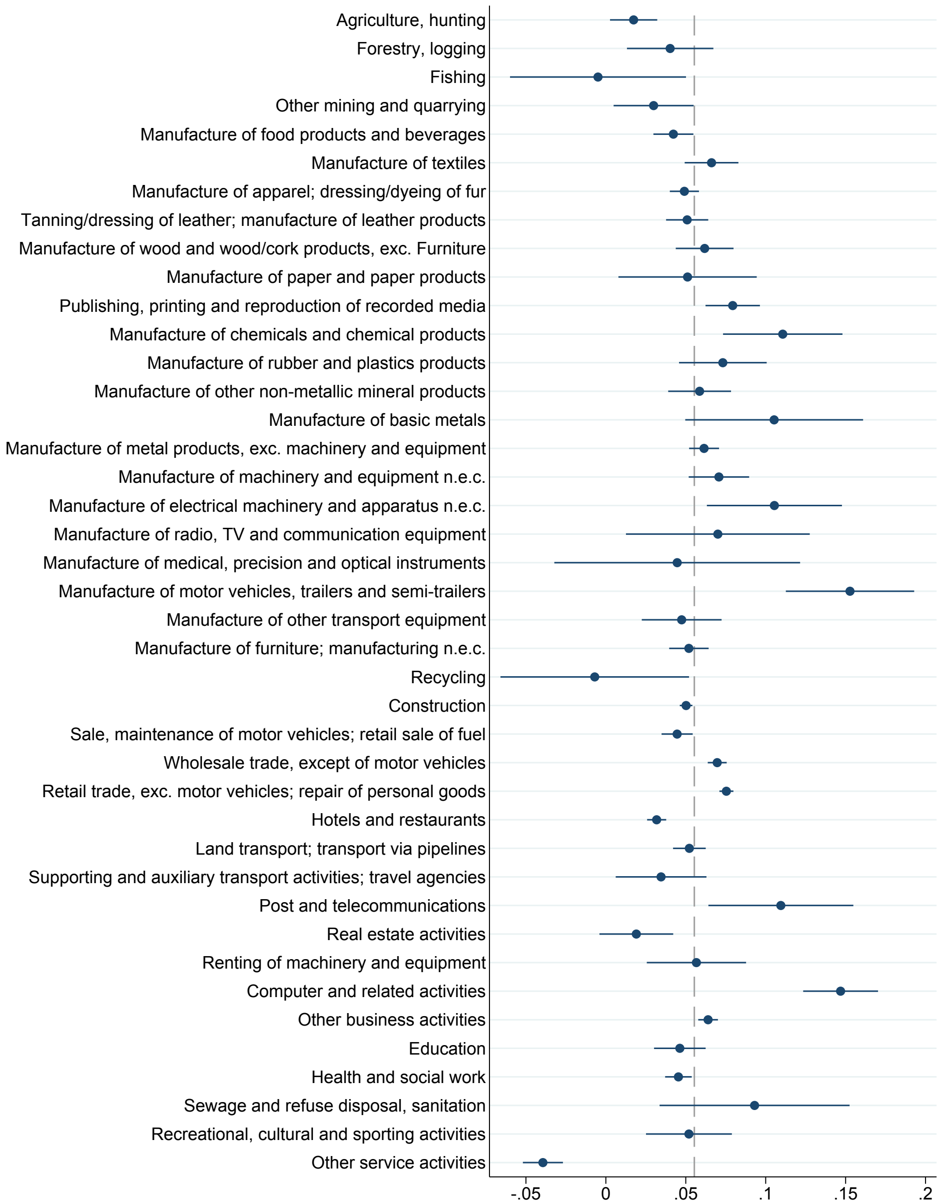


(b) **Age-by-Age Entrepreneur Schooling Coefficients**



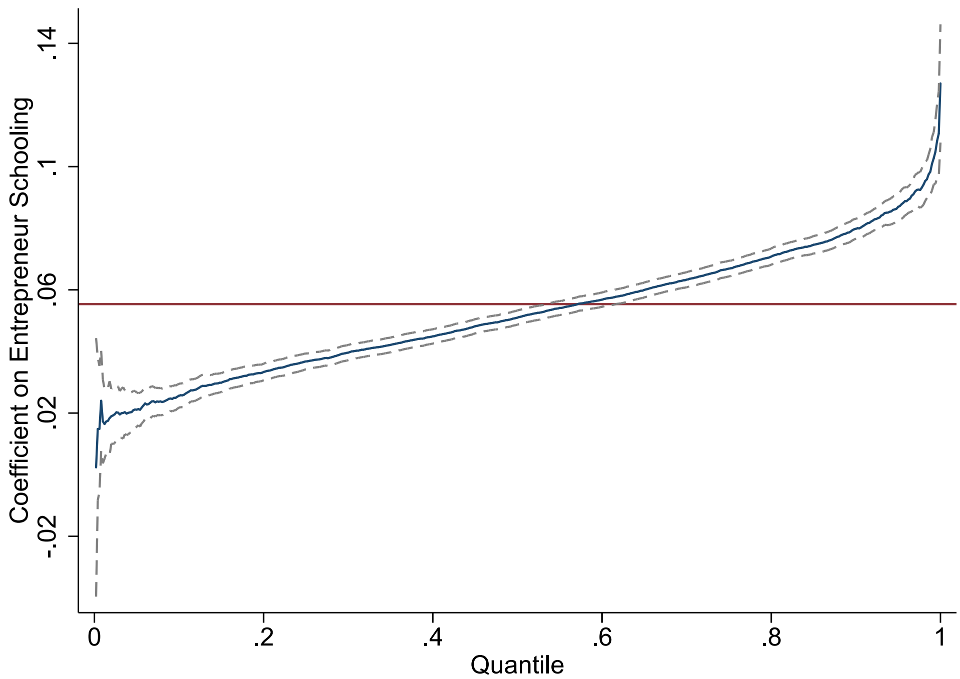
Notes: The top panel presents a binned scatter plot of log productivity ($\ln A$) and entrepreneur schooling. Firms are sorted into twenty equal-sized bins by entrepreneur schooling and the graph plots the log of productivity against entrepreneur years of schooling for each bin. Both variables are first residualized on a baseline set of controls (see main text for details). Along with the scatter plot, the graph also displays the corresponding regression line, obtained by estimating equation (14) on the underlying data. For the bottom panel, I add entrepreneur schooling by firm age interactions to equation (14), and plot these coefficients along with the respective 95% confidence intervals. Both panels use the sample of cohorts observed from entry, and errors are clustered at the firm level.

Figure 8: **Productivity and Entrepreneur Schooling by Sector**



Notes: This graph plots point estimates and 95% confidence intervals from an OLS regression of log productivity ($\ln A$) on entrepreneur schooling interacted with sector dummies and a baseline set of controls (see main text for details), for the sample of cohorts observed from entry. Errors are clustered at the firm level. The dashed line corresponds to the average OLS coefficient on entrepreneur schooling across sectors, from column one of table 2. I exclude a set of small sectors with wide confidence intervals from the graph (but not from the underlying regression) for clarity, namely mining of metal ores, electricity, gas, steam and hot water, collection, purification and distribution of water, water transport, air transport, financial intermediation and auxilliary activities, and activities of membership organizations. Together these sectors represent 0.1% of the sample.

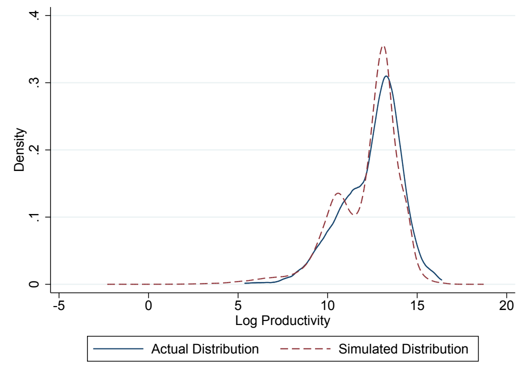
Figure 9: **Entrepreneur Schooling Coefficients by Quantile**



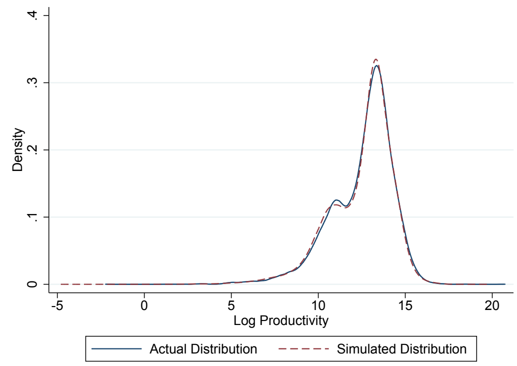
Notes: This graph plots point estimates and 95% confidence intervals for the entrepreneur schooling coefficient from quantile regressions of log productivity ($\ln A$) on entrepreneur schooling and a baseline set of controls (see main text for details). The sample includes all cohorts observed from entry. The quantile regressions are estimated at intervals of one-fifth of a centile, i.e. for the [0.1, 0.3,...,99.7,99.9] quantiles, and the horizontal line corresponds to the baseline OLS coefficient on entrepreneur schooling from column one of table 2. Standard errors are clustered at the firm level.

Figure 10: **Actual versus Simulated Productivity Distributions**

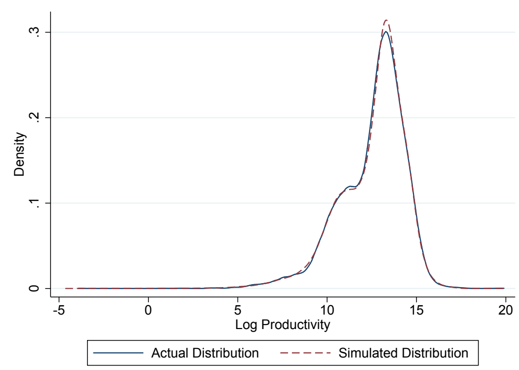
(a) **Zero Years of Entrepreneur Schooling**



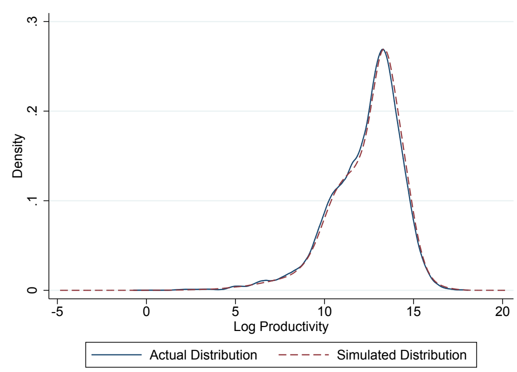
(b) **Four Years of Entrepreneur Schooling**



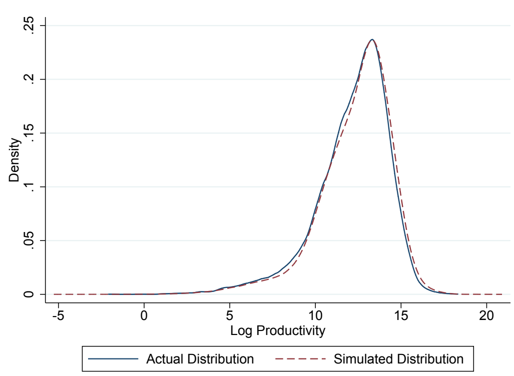
(c) **Six Years of Entrepreneur Schooling**



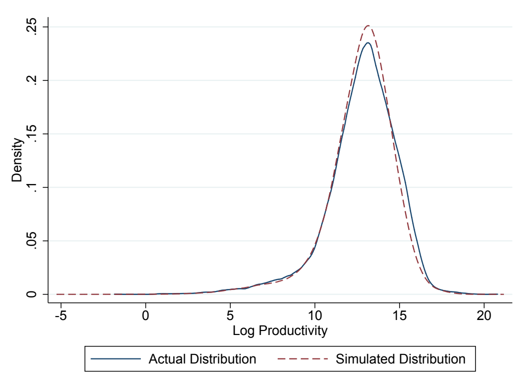
(d) **Nine Years of Entrepreneur Schooling**



(e) **Twelve Years of Entrepreneur Schooling**

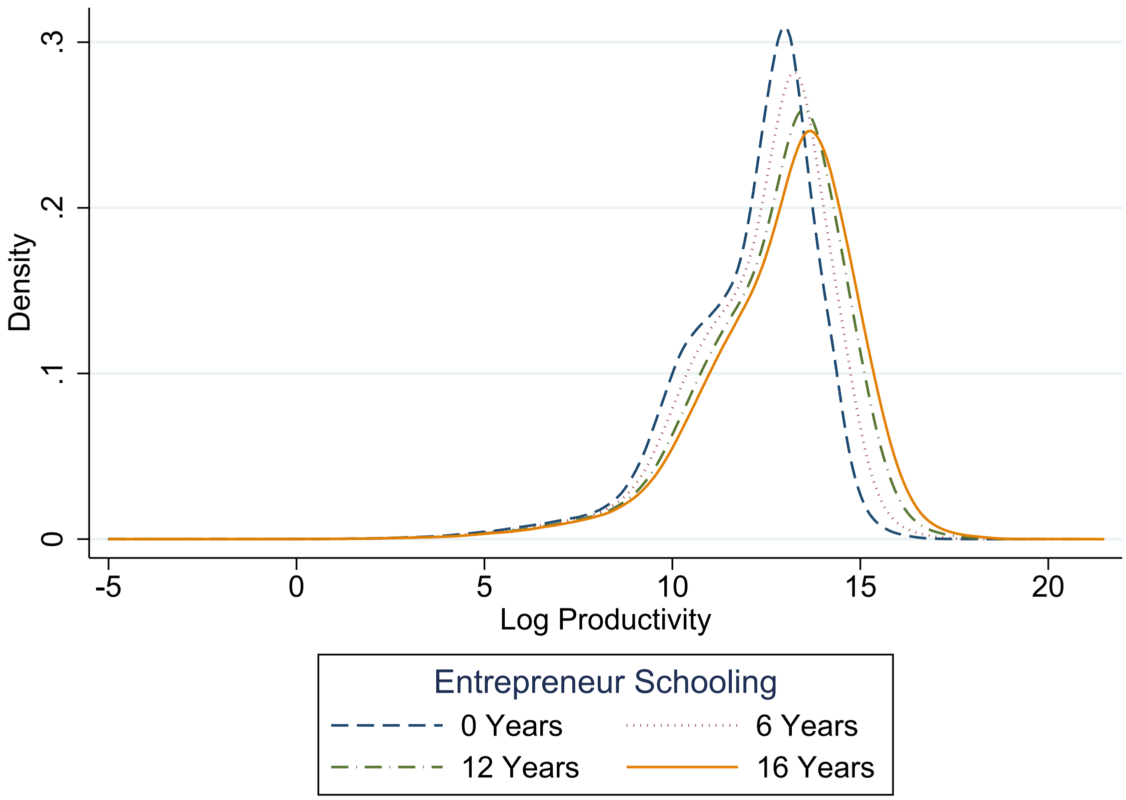


(f) **Seventeen Years of Entrepreneur Schooling**



Notes: These graphs plot kernel density estimates for actual versus simulated productivity distributions under the actual covariate values observed for each level of entrepreneur schooling, The simulated distributions are constructed from quantile regressions using the method of Machado and Mata (2005). See main text for details.

Figure 11: **Productivity Distributions by Entrepreneur Schooling**



Notes: This graph plots kernel density estimates for counterfactual productivity distributions holding the distribution of baseline covariates constant, and equal to the distribution of covariates in the sample of cohorts observed from entry. Each line corresponds to a different level of schooling. The distributions are constructed from quantile regressions using the simulation method of Machado and Mata (2005). See main text for details.

Table 8. **Coefficient on Entrepreneur Schooling - Additional Robustness**

	$\alpha = 0.25$	$\alpha = 0.275$	$\alpha = 0.3$	$\alpha = 0.325$	$\alpha = 0.35$	$\alpha = 0.375$	$\alpha = 0.4$
$\eta = 0$	0.0549	0.0542	0.0534	0.0527	0.0520	0.0512	0.0505
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)
$\eta = 0.025$	0.0544	0.0536	0.0529	0.0522	0.0514	0.0507	0.0499
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)
$\eta = 0.05$	0.0538	0.0531	0.0524	0.0516	0.0509	0.0502	0.0494
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)
$\eta = 0.075$	0.0533	0.0526	0.0518	0.0511	0.0504	0.0496	0.0489
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0013)	(0.0013)
$\eta = 0.1$	0.0528	0.0520	0.0513	0.0506	0.0498	0.0491	0.0483
	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
$\eta = 0.125$	0.0522	0.0515	0.0508	0.0500	0.0493	0.0485	0.0478
	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
$\eta = 0.15$	0.0517	0.0510	0.0502	0.0495	0.0487	0.0480	0.0473
	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
$\eta = 0.175$	0.0512	0.0504	0.0497	0.0490	0.0482	0.0475	0.0467
	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
$\eta = 0.2$	0.0506	0.0499	0.0492	0.0484	0.0477	0.0469	0.0462
	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0014)
Observations				690,056			

Notes: This table presents the coefficient on entrepreneur schooling from estimating equation (14) under different parametrizations of the production function, including the case when entrepreneurial human capital also enters the production function directly, as an additional input. In each cell, the production function is given by $zAh_e^{\eta}k^{\alpha}h^{1-\eta-\alpha}$, where η and α are given by the respective row and column, and $h_e \equiv l_e e^{0.08s}$ denotes the human capital of entrepreneurs (l_e represents the number of entrepreneurs, s their average years of schooling and the assumed return to schooling is 0.08). Adapting equation (7), firm productivity is then inferred from $\frac{(py)^{\frac{\sigma}{\sigma-1}}}{h_e^{\eta}k^{\alpha}h^{1-\eta-\alpha}}$, under the baseline $\sigma = 3$.