Entrepreneurial Human Capital and Firm Dynamics

Francisco Queiró*

June, 2021

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 life cycle growth. Consistent with an effect on growth, the thickness of the right tail of the size distribution increases with entrepreneur schooling. The evidence points to several underlying mechanisms, with technology adoption playing the most important part. I develop and estimate a model of firm dynamics that can parsimoniously account for these findings, and use it to draw aggregate implications. Accounting for the effect of entrepreneurial human capital on firm dynamics can substantially increase aggregate returns to schooling and the fraction of cross-country income differences explained by human and physical capital.

---

*Nova School of Business and Economics. Email: francisco.queiro@novasbe.pt. 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, Adriano Fernandes, Miguel Ferreira, Nicola Gennaioli, Danial Lashkari, Joana Naritomi, Clara Raposo, Kunal Sangani, Carlos Santos, Ariel Stern, Geoffrey Tate, Peter Thompson, Ian Tomb, Tom Zimmerman, Eric Zwick and four anonymous referees 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).

I Introduction

Cross-country regressions imply that human capital plays a major role in explaining output differences across countries (Mankiw, Romer and Weil, 1992), but may be biased by omitted factors such as the quality of institutions, culture or geography, among others. Within-country individual returns to schooling suggest a smaller role (Klenow and Rodriguez-Clare, 1997; Hall and Jones, 1999), but exclude any effect of human capital on total factor productivity (TFP). A recent literature links cross-country variation in TFP to differences in firm dynamics (Hsieh and Klenow, 2014), yet little is known about what underlies these differences.

This paper makes two points. First, it documents a strong relationship between entrepreneurial human capital and firm dynamics using administrative panel data on the universe of firms and workers in Portugal. Second, it shows that accounting for the effect of human capital on TFP through firm dynamics can explain a substantial fraction of the gap between individual-level and cross-country estimates of returns to schooling.

A key challenge in connecting firm dynamics to entrepreneurial characteristics has been the limited availability of comprehensive, high quality data. I combine employer-employee matched data, from which I identify entrepreneurs and their characteristics, with financial statements data, from which I measure firm performance. Portugal is a particularly attractive setting for this study because all schooling levels from primary school to college are well represented among entrepreneurs.

I find that both size at entry and life cycle growth increase with entrepreneur schooling. In my baseline specification, firms started by entrepreneurs with 15 or more years of schooling are 39% larger at entry in terms of sales than those started by entrepreneurs with less than six years of schooling, and 2.7 times larger by age 10. Most of this differential growth occurs within the first five years of the life cycle, with firms across schooling levels following close to parallel average growth paths beyond that. The same pattern holds when size is measured by value added or employment. These differences are almost fully explained by within-sector variation at the 5-digit level, not by selection into particular sectors, and by survivor growth, not by selection from higher exit rates among smaller firms. Moreover, they are specific to entrepreneurs. The average schooling of other workers appears to matter much less for firm dynamics.

One potential source of bias in these findings is omitted ability. To evaluate this concern, I leverage 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. I show that 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 labor market returns to schooling, on which there is a large literature (see Card, 1999, for a survey). When I implement this strategy, the relationship between entrepreneur schooling and firm dynamics is similar to my baseline findings, suggesting that ability bias plays a limited role. These results are in line with the literature on returns to schooling, which also finds that ability bias is small (Card, 1999).

I then examine how these differences in firm dynamics affect the cross-section of firms. I find that the distribution of firm size for higher levels of schooling is right-shifted and dilated relative to the distribution for lower levels, and quantile regressions show that the dilation is particularly strong in the upper tail. As is well-known, exponential growth implies a steady-state Pareto upper tail with an index declining in the growth rate,[1] or perhaps more intuitively with the thickness of the tail increasing with the growth rate. Consistent with an effect of schooling on growth, I find that the upper tail is Pareto for each level of entrepreneur schooling, and that the tail index declines monotonically with schooling, from 1.56 for entrepreneurs with less than six years of schooling to 1.09 for those with 15 or more.[2] In contrast, the upper tail of the wage distribution for all workers is also Pareto, but the tail index does not vary systematically with schooling.

I conclude the empirical exercise by turning to mechanisms. I investigate five possible channels that have been associated with firm heterogeneity in past research: more educated entrepreneurs may be better at innovation and technology adoption (Nelson and Phelps, 1966), better at increasing demand for their products (Foster, Haltiwanger and Syverson, 2016), better managers (Lucas, 1978; Murphy, Shleifer and Vishny, 1991), better at overcoming distortions that lead to misallocation (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009) or simply more ambitious (Hurst and Pugsley, 2011). Combining detailed data from financial statements, employee occupations, entrepreneur college majors, industry technological intensity and a new survey of management practices performed by the National Statistics Office, I find evidence consistent with all channels except misallocation, and with innovation and technology adoption having the most important role.

I introduce a simple model of entrepreneurial human capital and firm dynamics that can account for these findings. Individuals are endowed with a schooling level at birth, and choose to work as entrepreneurs or employees. Schooling affects both the initial level and the growth rate of entrepreneurial productivity, as well as the value of employee human capital. The effect of schooling on productivity growth introduces heterogeneity in the cross-section of returns to schooling among entrepreneurs. The model leads to simple expressions for

[1] Provided such a steady-state exists, which requires the presence of a friction such as random firm exit (see Gabaix, 2009, for a detailed exposition). The role of exponential growth as a mechanism driving the emergence of Pareto tails and their evolution has been explored in several contexts, such as city size (Gabaix, 1999), firm size (Luttmer, 2007), or income (Gabaix et al., 2016; Jones and Kim, 2018).

[2] The value for the higher schooling group is close to Axtell (2001)'s estimate of 1.06 for the tail index for the overall population of firms in the United States

aggregate output and TFP as a function of the distribution of schooling in the population, and it can quantitatively match the differences across schooling levels observed in the data. In particular, it can simultaneously match the differences in life cycle growth and in the cross-sectional size distribution, including the upper tail.

I then use the model to study implications for the aggregate role of human capital. The effect of entrepreneurial human capital on firm productivity translates into an effect on aggregate TFP, as in Lucas (1978) and Murphy, Shleifer and Vishny (1991), which leads to increasing returns. In particular, aggregate returns to schooling are given by the *sum* of individual mean returns in employment and in entrepreneurship, rather than a weighted average of the two. Importantly, this effect on TFP is not captured by estimates of human capital externalities that rely on differences in wage levels, housing costs or the average revenue product of inputs across locations (e.g. Rauch, 1993; Acemoglu and Angrist, 2000; Moretti, 2004a,b; Ciccone and Peri, 2006). Differences in firm productivity in the model are reflected in firm size rather than these factors. In addition, heterogeneous returns to schooling imply that Mincerian regressions understate the contribution of individual returns in each occupation. Accounting for these sources of amplification, I estimate aggregate returns to schooling in the model of 20-26%, substantially higher than the 6-10% typically found in the literature on individual returns (Card, 1999).

Finally, I use the model to perform a development accounting exercise. Using data from Caselli (2005) to facilitate comparison, I implement the variance decompositions proposed by Klenow and Rodriguez-Clare (1997) and Caselli (2005), and find that the fraction of cross-country income variation that can be explained by human and physical capital increases from the 40% reported by Caselli (2005) to between 59% and 78%.

This paper mainly contributes to the large literature on the determinants of TFP. First, it links the long-standing debate on the role of human capital in development[3] with the emerging literature on cross-country differences in firm dynamics, which has mostly focused on misallocation and institutional factors to date.[4] In doing so, the paper proposes and uses micro evidence to quantify a mechanism for the effect of human capital on TFP that amplifies aggregate returns to schooling relative to individual returns, yet is not captured by existing estimates of human capital externalities.

The paper can therefore help reconcile the high aggregate returns to schooling found in cross-country regressions (Mankiw, Romer and Weil, 1992) with the low estimates of human capital externalities found, for example, by Acemoglu and Angrist (2000) and by

[3]See Erosa, Koreshkova and Restuccia (2010), Schoellman (2012), Caselli and Ciccone (2013), Jones (2014), Manuelli and Seshadri (2014), Lagakos et al. (2018) and Hendricks and Schoellman (2018) for recent contributions, and Krueger and Lindahl (2001), Caselli (2005) and Hsieh and Klenow (2010) for overviews of the literature.

[4]Hsieh and Klenow (2014); Cole, Greenwood and Sanchez (2016); Bento and Restuccia (2017); Akcigit, Alp and Peters (2021)

Ciccone and Peri (2006). Other studies have employed macro data and model calibrations to study the effect of human capital on TFP, but have reached conflicting conclusions on its magnitude (Benhabib and Spiegel, 1994; Klenow and Rodriguez-Clare, 1997; Bils and Klenow, 2000; Klenow and Rodríguez-Clare, 2005; Córdoba and Ripoll, 2008). My approach relies on firm-level productivity differences, and in that sense parallels the use of individual returns to schooling in development accounting to infer the aggregate effect of human capital on output conditional on TFP.

Gennaioli et al. (2013) 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 TFP, as in Lucas (1978), Murphy, Shleifer and Vishny (1991) and my model. They find that returns to schooling for entrepreneurs are higher than for other workers, and that accounting for these higher returns can significantly increase average returns in the population relative to existing estimates, which tend to exclude entrepreneurial profits. These higher individual returns amplify the effect of human capital on aggregate output, a different channel than the one emphasized here. In my model, accounting for the effect of entrepreneur schooling on TFP can substantially amplify the aggregate effect of human capital on output even if average Mincerian returns to schooling for entrepreneurs are similar to those for employees, as my estimates suggest.

The paper also relates to the literature on the allocation of talent and TFP. Baumol (1990) and Murphy, Shleifer and Vishny (1991) show that settings where talented individuals select into rent seeking rather than entrepreneurship can be detrimental for growth. Caselli and Gennaioli (2013) find that financial frictions can lead to a failure of meritocracy, increasing the prevalence of dynastic management and reducing TFP. In line with these studies, my results suggest an important interaction between the aggregate supply of human capital and the institutions that govern selection into entrepreneurship.

Finally, at the micro level, the paper adds to a growing literature on the role of education in entrepreneurship. Several studies examine returns to schooling using entrepreneurial earnings, including Parker and Van Praag (2006), Van Praag, van Witteloostuijn and van der Sluis (2013), Levine and Rubinstein (2016) and Michelacci and Schivardi (2020). See Parker (2004) and Van Der Sluis, Van Praag and Vijverberg (2008) for reviews. A smaller set of studies have focused on firm performance, and have found that size (Mata, 1996; Cabral and Mata, 2003), as well as short-run growth rates (Cooper, Gimeno-Gascon and Woo, 1994; Kangasharju and Pekkala, 2002), increase with entrepreneur schooling. Relative to this literature, I study firm life cycle dynamics and the upper tail of the size distribution, and quantify aggregate implications.

The rest of the paper is structured as follows. Section II describes the data. Section III presents the empirical findings. Section IV introduces the model and the estimation. Section V examines aggregate implications, and section VI concludes.

II Data

The data used in the paper come 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, from 1985 to 2017. The survey combines firm-level information, such as total employment and date of incorporation, with a 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 in the non-financial sector, covering the period from 2004 to 2017. I rely on this dataset to obtain measures of firm performance. The two datasets share a firm identifier. Online appendix A provides definitions for variables used in the analysis and not covered in this section.

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.

This paper exploits the rich occupational data reported in QP to define as entrepreneurs the top managers of the firm at entry.[5] This is perhaps closest to the classical notions of Say (1836), who emphasizes the role of the entrepreneur in combining and coordinating factors of production, and Schultz (1975), who sees entrepreneurship as the ability to continuously reallocate resources in response to changes in economic conditions. It also naturally fits into the Lucas (1978) model.

While this definition does not account for risk bearing (Knight, 1921), as I do not observe ownership, there is evidence from several developed countries that the vast majority of businesses are owner-managed.[6] Owner-managers are likely to be even more common in

[5]Some firms do not report data at age zero, and in those cases I use the top managers reported at age one. This includes all firms started in 2001, since QP data on workers was not collected for that year. Firms that do not report a top manager by age one 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.

[6]For example, using data on the universe of S-corporations and partnerships in the United States, which account for the majority of businesses and business taxable income, Smith et al. (2019) show that 89% of firm owners report active income from their businesses (which implies they participate materially in the business), that most owners own just one firm instead of diversified portfolios, and that profits fall by three quarters when an owner dies prematurely or retires.

Portugal, given the prevailing role of family firms (La Porta, Lopez-de-Silanes and Shleifer, 1999), and particularly so at the moment of *entry*. If the owner of a successful firm initially manages the business but then chooses to switch to professional management, I identify the entrepreneur as the owner, not the manager. Still, I cannot reject the possibility that in some cases this definition identifies top managers who are not owners.

I identify top managers using the occupational classification in QP, which is available starting in 1995 and is based on the International Standard Classification of Occupations (ISCO).[7] 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 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% of the entrepreneurs in the 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.[8] The distinction is similar to that between associate and bachelor's degrees in the United States. Entrepreneur schooling is defined as average years of schooling of the firm's entrepreneurs.

Analysis Sample I restrict the sample to firms that entered in 1995 or later, which enables me to identify their entrepreneurs, and to firms that report at least one entrepreneur and one non-entrepreneur, along with their year of incorporation. 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

[7] In particular, it follows ISCO-88 between 1995 and 2009, and ISCO-08 from 2010 onward. I use the ISCOGEN Stata package (Jann, 2020) to match the two.

[8] 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.

of at least 50 percent.[9] I also exclude foreign subsidiaries, which I identify in the data as firms with 100% foreign ownership at entry.

Most of the results in the paper focus on the 2004-2017 period, when both QP and SCIE data are available. Online appendix figure F.1 shows a histogram of entrepreneur schooling for all firm-year observations in this sample. Over 80% of observations cluster at the five main schooling levels reported in the data – four, six, nine, twelve and seventeen years of schooling – with each level accounting for 15-20%. In several results throughout the paper 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. Online appendix table E.1 presents summary statistics at the firm-year level for these five groups in the same sample.

III Entrepreneur Schooling and Firm Dynamics

III.A Life Cycle Dynamics

I start by presenting graphical evidence on firm life cycle dynamics by level of entrepreneur schooling. Throughout this section, firms are sorted by average entrepreneur schooling into the five groups described in section II. For each figure, I estimate an OLS regression of the following form:

$$\ln Q_{i,t} = \sum_s \sum_a \beta_{s,a} c_{s,i} d_{a,i} + \phi X_{i,t} + \epsilon_{i,t} \tag{1}$$

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 dummies for non-entrepreneur schooling, using the same five groups, a quadratic in average entrepreneur and non-entrepreneur experience and year fixed effects.[10] I then plot $\hat{\beta}_{s,a}$ and the corresponding confidence intervals. Standard errors are clustered at the firm level.

[9] I additionally exclude government agencies, which are covered when they employ workers under private sector labor law, and non-profits. A number of large privatizations occurred 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. In some cases the privatized firms were reincorporated and show up as new firms in the data. To identify these cases, I follow Braguinsky, Branstetter and Regateiro (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.

[10] The schooling and experience of non-entrepreneurs may be choice variables for the entrepreneur, in which case they should be omitted from the regression. I exclude this matching channel by controlling for non-entrepreneur characteristics, but online appendix B.3 shows that this has little impact on the results, along with other robustness checks, and abstracting from matching simplifies the model developed in section IV considerably. None of these controls address concerns with omitted ability, which I confront separately below.

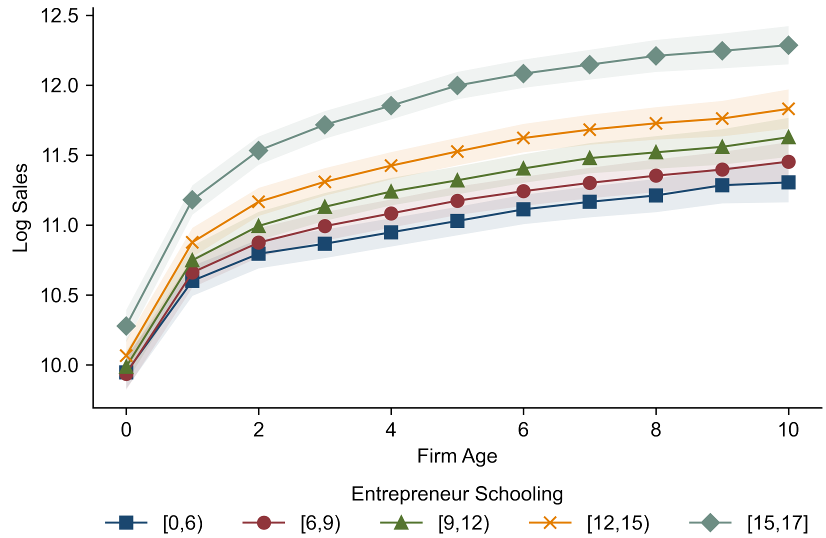

Figure 1: Firm Life Cycle Dynamics and Entrepreneur Schooling

Notes: Entrepreneur schooling by firm age coefficients from estimating (1) on sales data for firms up to age 10 in the 2004-2007 cohorts. The shaded areas represent 95% confidence intervals.

As a baseline sample, I use the 2004 to 2007 cohorts, which I can track from entry up to age 10. Financial statements data are available starting in 2004, including data on sales and value added, as well as several other outcomes that I use to investigate mechanisms below. Pooling these four cohorts together enables me to use the same sample throughout the analysis, and offers a good balance between the precision of estimates and the length of the life cycle I can track. Online appendix B.1 shows that the findings are robust to using older cohorts with more limited data but which I can track for a longer period.

Figure 1 plots coefficients for each entrepreneur schooling group at each age, when output is measured by sales. Both size at entry and life cycle growth increase monotonically with entrepreneur schooling. At entry, firms in the top group, whose entrepreneurs have 15 or more years of schooling, are 39% larger than those in the bottom group, whose entrepreneurs have less than six years of schooling. By age 10, they are 2.7 times larger. The remaining groups fall in between. Most of this differential growth occurs within the first 5 years of the life cycle, with firms across schooling levels following close to parallel average growth paths beyond that. The same pattern holds for value added and employment, as online appendix figure F.2 shows.

These differences could be driven by within-sector variation or by selection into particular sectors with higher growth potential. Figure 2a takes the coefficients for the top and bottom groups by schooling from figure 1, and plots them together with the corresponding coefficients from estimating (1) with 5-digit sector-by-year fixed effects. The differences between the two

(a) Sales: Overall vs Within-sector

(b) Sales: Survivors vs All Firms

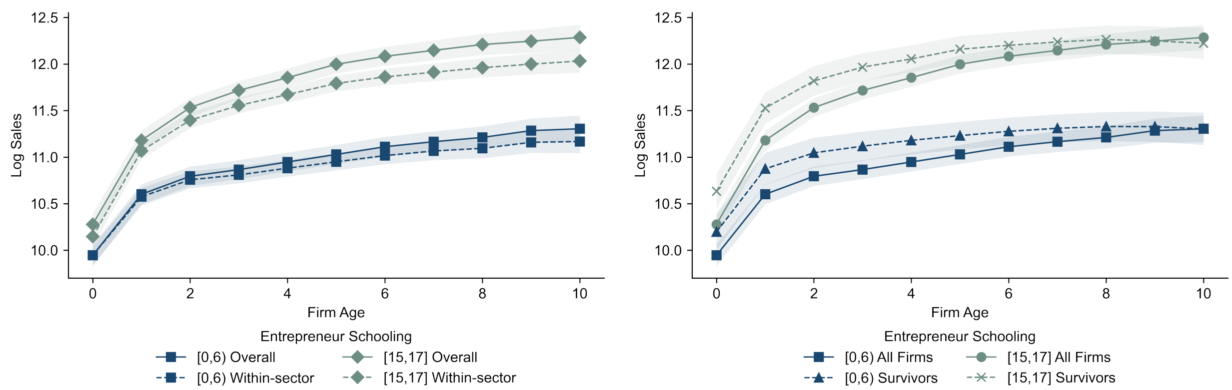

Figure 2: Firm Life Cycle Dynamics for the 2004-2007 Cohorts (cont.)

Notes: Entrepreneur schooling by firm age coefficients from estimating (1) on sales data for firms up to age 10 in the 2004-2007 cohorts. Panel a) compares estimates with and without 5-digit sector-by-year fixed effects. Panel b) compares estimates for the full sample and for survivors only. The shaded areas represent 95% confidence intervals.

groups are only marginally smaller in the within-sector specification, which suggests that they are mostly explained by variation within narrowly defined industries.

The differences could also 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 1 could emerge in the absence of differences in firm growth. To distinguish these channels, figure 2b plots the coefficients for the top and bottom groups from estimating (1) in the sample of firms who survived until age 10, as well as the coefficients estimated in the whole sample. In both groups, survivors are on average larger except at age 10, where they are nearly the same size by construction.[11] This indicates the presence of selection. But the differences between the two groups are clearly driven by differences in survivor growth.[12]

Are these findings specific to entrepreneurs or do they hold for more educated workers in general? Figure 3 is constructed analogously to figure 1 but sorts firms by average non-entrepreneur schooling instead. That is, it plots estimates from replacing the entrepreneur schooling by age terms in (1) with *non*-entrepreneur schooling by age interactions. At entry, schooling is negatively correlated with sales. Although the differences are salient in logs,

[11] They are not exactly identical because the estimated coefficients on controls are slightly different in the two regressions.

[12] Another possibility is that more educated entrepreneurs may simply be pursuing riskier strategies, with lower probabilities of survival but higher growth conditional on survival, but this does not seem to be the case. As online appendix figure F.2c shows, firms in the top group are slightly more likely to survive, while firms in the remaining groups experience similar survival rates.

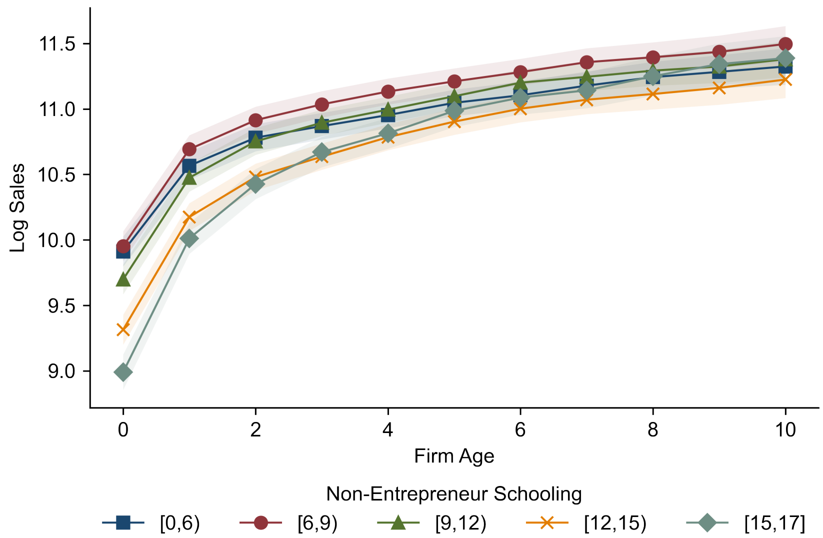

Figure 3: Sales by Non-Entrepreneur Schooling

Notes: Non-entrepreneur schooling by firm age coefficients from estimating (1) with the entrepreneur schooling by age terms replaced with non-entrepreneur schooling by age interactions, using sales data for firms up to age 10 in the 2004-2007 cohorts. The shaded areas represent 95% confidence intervals.

they are small in levels. For example, the difference between the bottom and top groups equals $e^{9.91} - e^{8.99} \approx 12,000$ euros. Post-entry, the groups converge, and by age 5 they are indistinguishable. This evidence suggests that it is the human capital of entrepreneurs, in particular, that matters for firm dynamics.

The previous figures show that the relationship between firm dynamics and entrepreneur schooling is clearly monotonic. Figure 4 shows it is approximately log-linear. To construct it, I first plot the coefficients on sales for each schooling group from figure 1 at ages one, two, five and ten, against average years of schooling in each group. I then estimate the following linear version of (1) and also plot the estimated regression lines for each age:

$$\ln Q_{i,t} = \sum_a \beta_a s_i d_{a,i} + \phi X_{i,t} + \epsilon_{i,t} \tag{2}$$

where X additionally includes age fixed effects and non-entrepreneur schooling also enters linearly. As the figure shows, the relationship between size and schooling at each age is slightly convex but well approximated by the regression lines, with the slope increasing in age as expected.[13] Online appendix B.2 shows that this linear relationship is relatively stable over time. I exploit this linearity in the results on omitted ability in the next section and in

[13]Interestingly, the emergence of convexity has also been noted in the literature on labor market returns to schooling (Lemieux, 2006a).

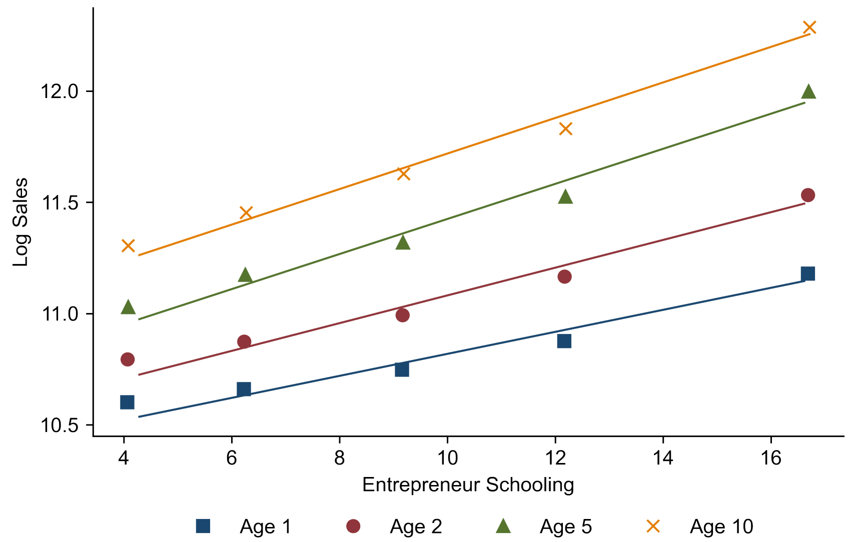

Figure 4: Linearity in Entrepreneur Schooling

Notes: Entrepreneur schooling by firm age coefficients from figure 1 and average years of schooling in each group, along with the corresponding regression lines from estimating (2).

the model developed below.

III.B Accounting for Ability

A key challenge in assigning a causal interpretation to the schooling coefficients estimated above is the possibility that they are 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 an extension of equation (2) where output is a function of the entrepreneur's natural ability b, in addition to schooling:

$$\ln Q_{i,t} = \beta^e s_i + \lambda^e b_i + \phi X_{i,t} + \epsilon_{i,t} \tag{3}$$

I omit firm age interactions to simplify the exposition, but the approach is easily extended to the case where the schooling and ability coefficients are age-specific, and I do this in the results reported below.

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

$$E(\ln w_i) = \beta^w s_i + \lambda^w b_i \tag{4}$$

I also omit experience in the earnings equation but account for it below. Correlation between s and b among entrepreneurs might arise either because ability and schooling are correlated in the overall population, or through selection into entrepreneurship as a function of the relative returns to s and b in (3) and (4).

Inverting (4) to express b as a function of $E(\ln w)$ and s, equation (3) can be rewritten as

$$\ln Q_{i,t} = \left(\beta^e - \frac{\lambda^e}{\lambda^w}\beta^w\right) s_i + \frac{\lambda^e}{\lambda^w} E(\ln w_i) + \phi X_{i,t} + \epsilon_{i,t} \tag{5}$$

This expression shows that the entrepreneur's expected earnings in the labor market 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 $E(\ln w)$ 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 $E(\ln w)$ 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.

To estimate equation (5), data on the entrepreneurs' expected 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. For comparison with my baseline findings, I use switchers from the 2004-2007 cohorts up to age 10, but the results are robust to using other cohorts. In this sample of switchers, which comprises about 60% of the baseline sample, I observe an entrepreneur's earnings when working as a non-entrepreneur in a prior employment spell, and I take the entrepreneur's last observed non-entrepreneurial earnings, residualized on year and experience dummies, as the entrepreneur's expected earnings in the labor market. The results are robust to using the average of all previous observations of non-entrepreneurial earnings, rather than just the last one.

One concern with this procedure could be measurement error. I do not observe the entrepreneur's expected earnings, and instead proxy for them with actual earnings in a previous job that the entrepreneur held at some time t'. Let these actual earnings be given by $\ln w_{i,t'} = E(\ln w_i) + \nu_{i,t'}$, where ν is a random shock. Then (5) becomes

$$\ln Q_{i,t} = \left(\beta^e - \frac{\lambda^e}{\lambda^w}\beta^w\right) s_i + \frac{\lambda^e}{\lambda^w} \ln w_{i,t'} - \frac{\lambda^e}{\lambda^w}\nu_{i,t'} + \phi X_{i,t} + \epsilon_{i,t} \tag{6}$$

The presence of ν in (6) in the error term attenuates the coefficient on $\ln w$, which possibly amplifies the coefficient on s if schooling and ability are positively correlated. But note that this attenuates the bias correction for the schooling coefficient as well. As long as ν is not correlated with s, the bias-corrected estimate of β^e will be minimally affected, as shown in online appendix B.4.

Table 1 presents the results from accounting for ability under this approach. I report estimates for the case where s and b in (3) are interacted with firm age indicators. This implies that $\ln w$ is also interacted with age in (5) and (6) and the bias-correction becomes age-specific. To avoid cluttering the table, I report only the coefficients at ages 0 and 10.

Column one presents the baseline specification from equation (2) estimated on the sample of switchers, with output measured by sales. The coefficients on entrepreneur schooling at ages 0 and 10 in this sample equal 0.0359 and 0.0861. Column two adds the entrepreneur's labor market earnings interacted with age. First, as expected if ability increases both labor market earnings and firm output, the coefficients on earnings are positive and significant. Second, the coefficient on entrepreneur schooling falls to -0.0047 at age 0 and 0.0260 at age 10. Third, the bias-corrected coefficients equal 0.0308 and 0.0778. These are lower than the baseline estimates without controlling for ability, but the differences are small. As explained above, the bias-corrected coefficients are obtained by adding the coefficients on earnings at each age, multiplied by an estimate for the labor market return to schooling, β^w, to the corresponding coefficients on entrepreneur schooling. I assume an estimate of 8% for the returns to schooling parameter β^w, the midpoint of the 6%-10% range reported in the literature (Card, 1999). With $\beta^w = 6\%$, the bias-corrected coefficients drop to 0.0219 and 0.0648, while with $\beta^w = 10\%$ they rise to 0.0397 and 0.0907.

One limitation of this approach, as just discussed, 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),[14] and the bias-corrected coefficients are

[14] 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

Table 1: Accounting for Ability

	Sales				Value Added	
	(1)	(2)	(3)	(4)	(5)	(6)
Entrepreneur Schooling ×						
Firm Age = 0	0.0359	-0.0047	0.0321	-0.0057	0.0361	-0.0094
	(0.0041)	(0.0044)	(0.0041)	(0.0044)	(0.0041)	(0.0044)
Firm Age = 10	0.0861	0.0260	0.0807	0.0252	0.0878	0.0300
	(0.0049)	(0.0054)	(0.0049)	(0.0054)	(0.0046)	(0.0051)
Log Last Wage ×						
Firm Age = 0		0.4437		0.4304		0.5034
		(0.0258)		(0.0261)		(0.0268)
Firm Age = 10		0.6474		0.6246		0.5907
		(0.0342)		(0.0346)		(0.0338)
Bias-corrected Entrep. Sch. ×						
Firm Age = 0		0.0308		0.0287		0.0309
		(0.0040)		(0.0041)		(0.0040)
Firm Age = 10		0.0778		0.0751		0.0773
		(0.0047)		(0.0047)		(0.0044)
Number of Prior Occupations ×						
Firm Age = 0			0.0756	0.0478		0.0126
			(0.0116)	(0.0116)		(0.0120)
Firm Age = 10			0.1144	0.0612		0.0508
			(0.0161)	(0.0158)		(0.0152)
Entrepreneur Experience	0.0165	0.0011	0.0086	-0.0030	0.0091	-0.0093
	(0.0042)	(0.0040)	(0.0042)	(0.0040)	(0.0039)	(0.0038)
Entrepreneur Experience^2	-0.0002	-0.0001	-0.0000	-0.0000	-0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Non-Entrepreneur Schooling	0.0019	-0.0101	-0.0003	-0.0110	0.0034	-0.0089
	(0.0043)	(0.0041)	(0.0043)	(0.0042)	(0.0040)	(0.0038)
Non-Entrepreneur Experience	0.0734	0.0695	0.0730	0.0694	0.0790	0.0756
	(0.0034)	(0.0033)	(0.0034)	(0.0033)	(0.0032)	(0.0031)
Non-Entrepreneur Experience^2	-0.0015	-0.0015	-0.0015	-0.0015	-0.0017	-0.0016
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
N	128,102	128,102	128,102	128,102	114,920	114,920
R^2	0.116	0.161	0.123	0.163	0.114	0.162

Notes: Estimates of (2) on the sample of entrepreneurs observed in other occupations before becoming entrepreneurs. Only coefficients at ages 0 and 10 are reported. Columns two, four and six add the log of the entrepreneur's wage in the last occupation before becoming an entrepreneur, as in (6). The bias-corrected coefficients at each age equal the coefficients on entrepreneur schooling plus an assumed labor market return to schooling of 8% multiplied by the coefficient on log last wage. Output is measured by sales in columns one to four and by value added in columns five and six. The number of prior occupations in columns three, four and six is the number of past occupations the entrepreneur has held before becoming an entrepreneur. All regressions include firm age and year fixed effects. Standard errors are clustered at the firm level.

past employment and the standardized occupational codes in the data to measure each entrepreneur's number of prior occupations.

close to the estimates in column two. Columns five and six report estimates for value added instead of sales, using the two ability controls, and the results are again similar.

Put together these results suggest that bias from omitted ability in the baseline estimates is unlikely to be a significant issue, in line with the literature on labor market returns to schooling (Card, 1999).

III.C The Cross-Section of Firm Size

Next, I explore how these differences in dynamics translate into heterogeneity in the cross-section of firms. In this section, my baseline sample pools all firms aged 10 or less in the 2005 to 2017 cross-sections. This maximizes power and ensures that the age distribution is comparable to my baseline sample for firm dynamics.

Figure 5a plots the cross-sectional distribution of sales by entrepreneur schooling. To parallel the results on firm dynamics, I first estimate (1) in this sample, and then plot the resulting residuals added to the estimated $\hat{\beta}_{s,a}$ coefficients. The figure shows that the distributions of sales for higher levels of schooling are right-shifted and dilated relative to those for lower levels. The left tails are largely indistinguishable, but as the quantiles increase the differences become more prominent.[15] To probe further, I estimate quantile regressions of the form

$$\ln Q_{i,t} = \beta_\theta s_i + \phi_\theta X_{i,t} + \epsilon_{i,t}$$

for every percentile θ of the distribution of log sales, where X denotes the same set of controls as in equation (2). To zoom in on the top of the distribution, I estimate an additional quantile regression for the 99.9th percentile. Figure 5b plots the estimated β_θ coefficients on entrepreneur schooling, along with the corresponding estimate from an OLS regression, and a clear pattern emerges. Consistent with a right-shift and dilation, the coefficient is close to zero in the left tail, rises steadily with the quantiles of the distribution, equaling the OLS coefficient of 0.057 at around the 60th percentile, and attains its highest levels in the upper tail. The rise at the top of the distribution is striking, from 0.096 in the 90th percentile to 0.165 in the 99.99th percentile, almost three times the OLS coefficient. These quantile coefficients suggest that returns to schooling among entrepreneurs exhibit substantial heterogeneity, and are particularly high in the upper tail of the firm size distribution.

What might account for the sharp divergence in the tail? As is well-known, the upper tail of the firm size distribution tends to follow a Pareto distribution (e.g. Axtell, 2001), and a natural mechanism that can account for this shape is exponential growth (Gabaix, 1999; Luttmer, 2007). Coupled with a friction such as random exit, exponential growth yields a

[15] The patterns are the same when including five-digit sector-by-year fixed effects in the estimation. Accounting for sectors makes the distributions smoother, and in particular shows that the hump in the right shoulder of the distribution for the top group in figure 5a is driven by sector heterogeneity.

(a) Overall Distribution of Sales (b) Quantile Regression Coefficients

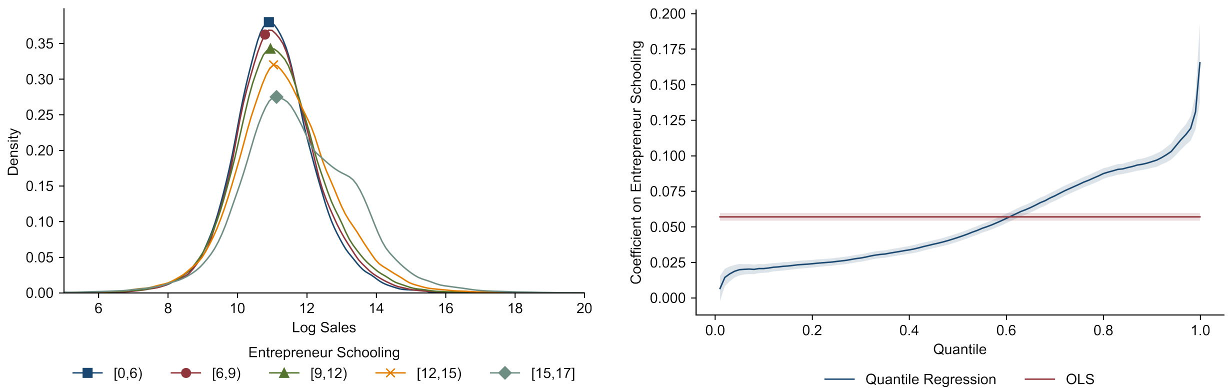

Figure 5: Entrepreneur Schooling and the Cross Section of Firms

Notes: Panel a) plots the distribution of log sales by level of entrepreneur schooling in the pooled 2005-2017 cross-sections of firms aged 10 or less. I first estimate (1) in this sample, and then plot the resulting residuals added to the estimated $\hat{\beta}_{s,a}$ coefficients. Panel b) plots coefficients from quantile regressions of log sales on entrepreneur schooling for every percentile of the distribution plus for the 99.9th percentile (i.e. for the [1,2,...,99,99.9] quantiles). The horizontal line represents the corresponding OLS coefficient. The shaded areas represent 95% confidence intervals.

steady-state cross-sectional Pareto tail with an index declining in the growth rate, or more intuitively with the thickness of the tail increasing in the growth rate (see e.g. Jones and Kim, 2018, for a simple derivation). An effect of entrepreneur schooling on firm growth in such a model would therefore imply that the upper tail of the size distribution should be Pareto, and that the thickness of the tail should increase with schooling.

Figure 6a shows that this is indeed the case, suggesting a link between the findings on growth presented above and the cross-section of size. I explore this link in the model developed below. The figure displays a binned scatterplot of the log CCDF (i.e. 1-CDF) of sales in the top decile of the distribution for each entrepreneur schooling group. To construct it, I take the top decile of the sales residuals plotted in figure 5a for each schooling group, sort firms by these residuals into 20 equal-sized bins, and then plot the mean of the log CCDF versus the mean of log sales residuals in each bin, along with the corresponding regression line estimated on the underlying data. As the figure shows, the relationship is almost perfectly linear for each group, consistent with a Pareto-shaped tail, and the thickness of the tail increases monotonically with entrepreneur schooling. Column one of table 2 reports the regression slopes for each group, which imply tail indices ranging from 1.09 for the top group to 1.56 for the bottom group (the tail indices are given by the negative of the regression slopes). These results are not specific to sales or this particular sample, and columns two to six present alternative specifications. In every case, the indices fall with schooling and

(a) Top Decile of Sales (b) Top Decile of Wages

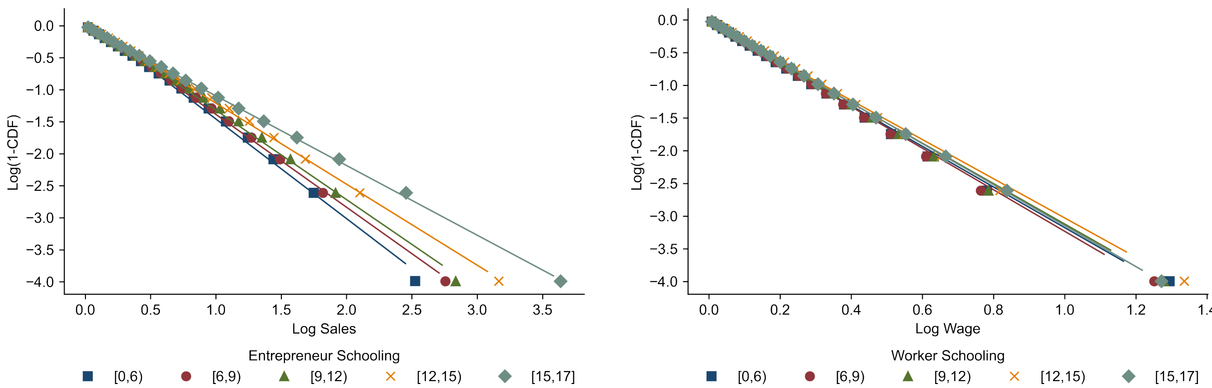

Figure 6: Upper Tails of the Firm Size and Wage Distributions

Notes: Panel a) displays a binned scatterplot of the log CCDF (i.e. 1-CDF) of sales in the top decile of the distribution by schooling level. Panel b) is a binned scatterplot of the log CCDF of wages in the top decile of the distribution by schooling level, for the overall population of workers in the 1995 to 2017 cross-sections.

the linear slopes are an almost perfect description of the tail data, as indicated by the high values of R^2.

In contrast, figure 6b shows that the tail of the *wage* distribution for all workers is also Pareto, but the tail index does not seem to vary systematically with schooling. I first estimate a standard Mincerian regression of log earnings on schooling, a quadratic in experience and year fixed effects for all workers in the QP data, pooling the 1995 to 2017 cross-sections, and then construct the figure using the top decile of the resulting residuals for each schooling group. Column seven of table 2 reports the corresponding estimates. This suggests the relationship is specific to firm size and entrepreneur schooling.

III.D Mechanisms

In this section I turn to possible mechanisms that might underlie an effect of entrepreneur schooling on firm dynamics. I investigate five possible channels that have been associated with firm heterogeneity in past research: more educated entrepreneurs may be better at innovation and technology adoption (Nelson and Phelps, 1966), better at increasing demand for their products (Foster, Haltiwanger and Syverson, 2016), better managers (Lucas, 1978; Murphy, Shleifer and Vishny, 1991), better at overcoming distortions that lead to misallocation (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009) or simply more ambitious (Hurst and Pugsley, 2011).

I present four sets of results. I use the baseline 2004-2007 cohorts sample throughout

Table 2: Upper Tail Index Estimates

	Sales				Value Added		Wages
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Entrepreneur Schooling							
[0,6)	-1.5559	-1.3829	-1.6719	-1.8988	-1.7946	-2.0047	-3.1071
	(0.0009)	(0.0029)	(0.0014)	(0.0010)	(0.0010)	(0.0008)	(0.0003)
[6,9)	-1.4499	-1.2758	-1.4952	-1.7227	-1.6696	-1.8141	-3.1749
	(0.0008)	(0.0023)	(0.0012)	(0.0008)	(0.0008)	(0.0007)	(0.0003)
[9,12)	-1.3977	-1.3024	-1.4774	-1.6566	-1.6221	-1.7659	-3.0886
	(0.0007)	(0.0020)	(0.0010)	(0.0007)	(0.0007)	(0.0006)	(0.0003)
[12,15)	-1.2647	-1.2049	-1.3009	-1.4128	-1.4033	-1.5451	-2.9966
	(0.0006)	(0.0018)	(0.0009)	(0.0006)	(0.0006)	(0.0005)	(0.0003)
15+	-1.0882	-1.0710	-1.0971	-1.1776	-1.0965	-1.2228	-3.1325
	(0.0006)	(0.0017)	(0.0008)	(0.0005)	(0.0005)	(0.0004)	(0.0004)
N	73,400	7,849	36,700	73,400	66,137	66,137	5,395,971
R^2	0.996	0.996	0.996	0.997	0.997	0.998	0.990

Notes: Columns one to six in this table present results from regressions of the log CCDF of output (i.e. 1-CDF) on indicators for entrepreneur schooling groups and on log output interacted with these indicators, estimated on the upper tail of the cross-section of firms. The table only reports the interactions. Output is first residualized on the set of controls in equation (1) and standard errors are clustered at the firm level. In column one output is measured by sales, the sample includes firms up to age 10 in the 2005 to 2017 cross-sections, and the upper tail is defined as the top decile of output for each schooling group. Relative to this baseline, column two includes all firms observed from entry in the 2017 cross-section, up to age 22, column three defines the upper tail as the top 5% and column four adds 5-digit sector-by-year fixed effects to the set of controls used to calculate residuals. Columns five and six replicate columns one and four for value added instead of sales. Finally, column seven presents analogous results for the top decile of wages and schooling in the entire workforce.

the analysis, except for the second set of results, which employs a separate dataset. In the first set, I combine data from financial statements and worker occupations to construct a series of outcomes that point to specific channels. For each of these outcomes, I estimate a regression of the form given by (2), including five-digit sector-by-year fixed effects so that the results are driven by within-sector variation only. All outcomes are standardized to make the coefficients across outcomes easier to compare. The results are summarized in figure 7, which plots entrepreneur schooling-by-age coefficients for each outcome, along with 95% confidence intervals.

Figure 7a focuses on six outcomes related to innovation and technology adoption. The first two are whether the firm invests in R&D and whether it employs workers in STEM occupations, such as engineers and scientists.[16] These are direct measures of investment

[16] I define positive R&D investment as the firm reporting any R&D expense or having any "workers allocated to R&D" (the latter is reported by firms in SCIE). STEM stands for Science, Technology, Engineering and Math. I define STEM workers as those classified with the following ISCO 2008 occupation codes: 1223 (R&D managers), all codes in sub-major group 21 (Science and Engineering Professionals) and all codes in sub-major group 25 (Information and Communications Technology Professionals).

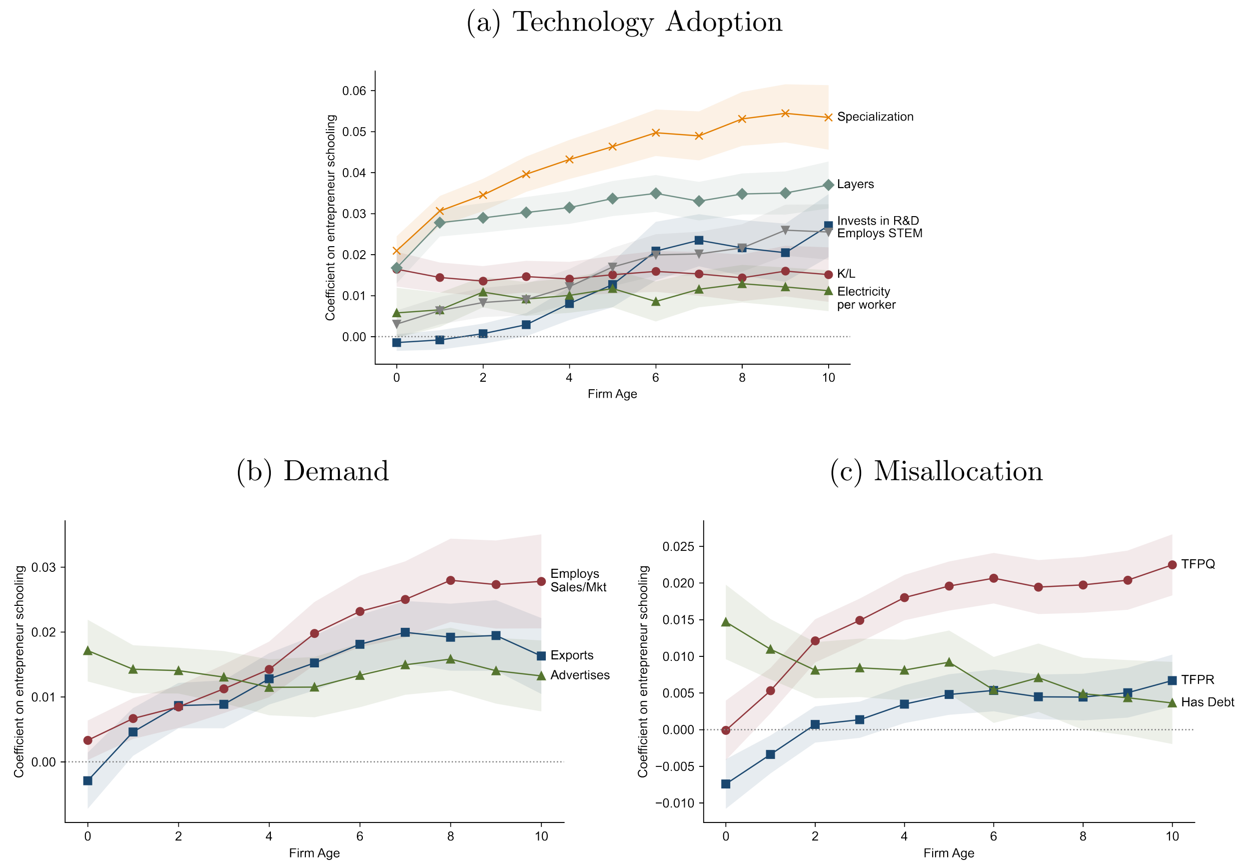

Figure 7: Mechanisms: Evidence From Financial Statements and Worker Occupations

Notes: Entrepreneur schooling by firm age coefficients from estimating regressions of the form in (2) for a set of outcomes linked to particular mechanisms. Outcomes are described in the text. All regressions include five-digit sector-by-year fixed effects. The shaded areas represent 95% confidence intervals.

in technology in a strict sense. The third and fourth are the log of the capital-labor ratio and the log of electricity expense per worker.[17] These are less direct measures, but if new technologies are at least to some extent embodied in capital (Solow, 1960), then higher capital-labor ratios within narrowly-defined sectors are consistent with a more intensive use of technology. The same goes for electricity usage, which has been used as a proxy for automation (Aghion et al., 2020).[18]

The last two outcomes are the degree of specialization, measured by the number of unique 4-digit ISCO occupations the firm employs, and the number of organizational layers, measured as in Caliendo et al. (2020). The idea that specialization and the division of

[17]Capital is measured using the book value of fixed assets in SCIE, including both tangible and intangible assets. Electricity expense is also reported in SCIE.

[18]Aghion et al. (2020) are able to distinguish between electricity consumed by motors and other uses, such as heating, and only use the former. The financial statements I use only report total electricity consumption.

labor are crucial drivers of productivity growth goes back to Adam Smith. Becker and Murphy (1994) develop a model where the extent of specialization is determined by the cost of coordinating specialized workers and the amount of knowledge employed in production. Both coordination costs and knowledge are plausibly affected by entrepreneurial human capital, for instance through better management and faster technology adoption. In a similar vein, Garicano (2000) models the organization of firms into hierarchical layers to maximize efficiency in the use of knowledge, and Caliendo, Monte and Rossi-Hansberg (2015) show that firms tend to add layers as they grow. I interpret these two outcomes as measures of technology, but they might also be interpreted as managerial quality.

As figure 7a shows, all six outcomes increase with entrepreneur schooling and, except for the capital-labor ratio and electricity per worker, the coefficients rise with age. Out of all outcomes, schooling is most strongly associated with specialization. At age 10, for example, the estimates imply that the degree of specialization employed by the average entrepreneur in the top schooling group in figure 1 is 0.68 standard deviations higher than the one employed by the average entrepreneur in the bottom group. This compares with 0.47 for layers, 0.34 for R&D, 0.32 for STEM workers, 0.19 for the capital labor ratio and 0.14 for electricity per worker.

Figure 7b presents outcomes related to demand. First, I examine whether the firm advertises, and second whether it employs sales or marketing professionals.[19] These first two outcomes are directly related to demand stimulation. Third, I examine whether a firm exports. Regardless of the underlying drivers of entry into export markets, exporting firms face increased demand for their products, and are able to expand as a result.[20] All three outcomes are positively related to schooling. The coefficients on sales and marketing professionals and on exports grow with age, while the ones on advertising are stable. The results imply that, at age 10, the average entrepreneur in the top group is 0.35 standard deviations more likely to employ a sales or marketing professional, 0.21 standard deviations more likely to export and 0.17 standard deviations more likely to advertise than the average entrepreneur in the bottom group.

The last figure in this first set of results, figure 7c, plots outcomes related to misallocation. In the presence of distortions, differences in firm size may not reflect differences in productivity and demand fundamentals alone, but also misallocation. Examples of such distortions include financial constraints, taxes and regulations. If more educated entrepreneurs are better able to overcome these distortions, then misallocation could partly account for

[19] These are defined as workers classified with ISCO 2008 occupation codes 1221 (Sales and Marketing Managers) and 1222 (Advertising and Public Relations Managers), and all codes in minor group 243 (Sales, Marketing and Public Relations Professionals).

[20] In Melitz (2003), for example, entry into exports is driven by productivity, since the gains from exporting are larger for more productive firms. Another possibility is that the cost of exporting is lower for more educated entrepreneurs, perhaps because they are more likely to speak multiple languages.

the findings documented in the previous sections. Following Hsieh and Klenow (2009), I use the firm's average revenue product of inputs (TFPR), often referred to as revenue-based productivity, as a measure of misallocation. In their widely adopted framework, the smaller a firm is relative to its efficient size, the higher its TFPR will be, regardless of the underlying source of misallocation. If, for example, firms with less educated entrepreneurs face stronger financial constraints which limit their growth (Parker and Van Praag, 2006), then these firms should exhibit relatively higher values of TFPR than firms with more educated entrepreneurs. For comparison, I also plot TFPQ, the measure of firm productivity in Hsieh and Klenow (2009).[21] In addition to TFPR and TFPQ, I examine a measure directly related to access to finance: whether the firm reports any debt on its balance sheet. This does not distinguish between credit supply and demand, and should be seen as suggestive.

Figure 7c shows that entrepreneur schooling is negatively associated with log TFPR at entry. This suggests that less educated entrepreneurs start inefficiently small, and that part of the size differences at entry documented above may be explained by misallocation. However, in contrast to the results on technology and demand, the relationship quickly reverses. By age two the coefficient on schooling is close to zero, and beyond that it is positive. This implies that, if anything, size differences at older ages might understate differences in productivity and demand fundamentals. The coefficients for debt suggest this pattern might be linked with credit constraints. At age zero, schooling is positively associated with having debt, but the coefficient declines substantially by age two. By age 10 it is close to zero and insignificant. The same holds when I use the ratio of debt to assets, instead of whether the firm has debt. The coefficients for TFPQ, on the other hand, are close to zero at entry but grow strongly until age five or so and remain stable beyond that, very much in line with the evidence on firm growth from figure 1. These results indicate that misallocation might play a role at entry, but beyond that the relationship between entrepreneur schooling and firm dynamics is driven by fundamentals.[22]

For the second set of results, I employ data from a new survey of management practices performed by the National Statistics Office in Portugal in 2016 for a sample of firms, and which includes information on whether the top manager of the firm has a college degree. I focus on the respondents to the survey that report being both founder-managed and not controlled by another firm. In this sample, one-third of entrepreneurs have a college degree, which compares with 21% in the QP data in the same year. The survey asks a detailed set

[21] I measure TFPR as $\frac{PY}{K^\gamma H^{1-\gamma}}$ and TFPQ as $\frac{(PY)^{\frac{\sigma}{\sigma-1}}}{K^\gamma H^{1-\gamma}}$, where PY is value added, K is the value of fixed assets reported by the firm and $H = \sum_i L_i e^{rs_i}$, where L_i is the number of workers with schooling s_i and $r = 0.08$. I follow Hsieh and Klenow (2009) in setting $\sigma = 3$, and γ is computed from sector labor shares in Portugal from the EU KLEMS database (Stehrer et al., 2019).

[22] This does not imply that misallocation is low in Portugal. In fact, Reis (2013) and Gopinath et al. (2017) find the opposite. It only suggests that it cannot account for the systematic differences in firm dynamics across schooling levels that I document here.

of questions about the firm and its management practices. I present results for a subset of questions related to three areas: (1) targets and monitoring, (2) incentives and (3) the extent of decentralization in decision-making. The first two correspond to the main areas covered by Bloom and Van Reenen (2007)'s survey of management practices, with the exception of operations. The third is analyzed by Bloom, Sadun and Van Reenen (2012).

Table 3 reports the fraction of firms that engage in each practice among entrepreneurs with and without college degrees. In every case, college-educated entrepreneurs are more likely to adopt better management practices, as characterized by Bloom and Van Reenen (2007) and Bloom, Sadun and Van Reenen (2012), and the differences are significant at the 1% level. Starting with targets and monitoring, college-educated entrepreneurs are more likely to set both short and long term goals, to monitor key performance indicators at least monthly, and to disseminate these indicators throughout the organization. In addition, they are more likely to conduct individual performance reviews at least annually. In terms of incentives, they are more likely to offer training and development, to offer stock awards, profit-sharing or bonuses, to promote exclusively based on performance and not on seniority or family connections, to have promoted workers in the past year and to have fired low performing workers in the past year. Finally, in terms of decentralization, they are more likely to report believing in collective intelligence, to allow their teams to make decisions within predefined limits, and to report that non-managers are highly involved in decision-making.

In addition, the survey also asks questions related to the entrepreneur's desire to innovate and grow, which Hurst and Pugsley (2011) show is a key determinant of actual firm growth. I report results on these questions at the bottom of table 3. More educated entrepreneurs are more likely to report having a strategy geared towards growth, to characterize their goals as ambitious, to be focused on new goods and services and on new management practices, and to have learned about new management practices in the past year.

The evidence presented so far shows that several mechanisms may play a part in accounting for the relationship between entrepreneurial human capital and firm dynamics. But it does not tell us about their relative importance. The third set of results examines differences in firm dynamics among college-educated entrepreneurs as a function of their field of study. Fields of study are reported for all college-educated workers in QP. I group them into five areas: STEM (Science, Technology, Engineering and Math), which comprises 24% of entrepreneurs in the sample, HASS (Humanities, Arts and Social Sciences, 19%), Business (18%), Health (17%) and Other (23%).[23] I then calculate the fraction of entrepreneurs with degrees in each field at each firm, and estimate (2) with the entrepreneur schooling-by-age

[23]Other includes degrees classified as Architecture, Agriculture and Fishing, Veterinary Medicine, Social Services, Personal Services, Transport Services, Environmental Protection, Security Services and Unknown. Most entrepreneurs in this group are classified as unknown.

Table 3: Mechanisms: Evidence from Survey of Management Practices

	College No	College Yes	Δ	*p*
Targets and Monitoring				
The firm has both short and long term goals	0.51	0.61	0.10	0.00
Top management monitors key performance indicators at least monthly	0.40	0.55	0.15	0.00
Key performance indicators are disseminated to the organization	0.69	0.84	0.16	0.00
The firm conducts individual performance reviews at least annually	0.13	0.25	0.13	0.00
Incentives				
The firm offers training and development opportunities to its workers	0.21	0.35	0.14	0.00
The firm provides incentives through stock awards, profit-sharing or bonuses	0.10	0.20	0.11	0.00
The firm promotes its workers exclusively based on ability and performance	0.27	0.42	0.15	0.00
The firm promotes its workers mostly based seniority or family connections	0.04	0.01	-0.03	0.00
The firm has promoted workers in the past year	0.40	0.49	0.08	0.01
The firm has fired low performing workers in the past year	0.08	0.15	0.07	0.00
Decentralization				
The entrepreneur believes in collective intelligence	0.12	0.21	0.08	0.00
The entrepreneur tends to allow others to make decisions	0.14	0.23	0.09	0.00
Non-managers are highly involved in decision-making	0.23	0.34	0.11	0.00
Desire to Innovate and Grow				
The firm's strategy is geared towards growth	0.44	0.56	0.12	0.00
The firm has ambitious goals	0.23	0.30	0.06	0.02
In terms of products, the firm is focused on new goods/services	0.09	0.12	0.04	0.06
In terms of organization, the firm is focused on new management practices	0.28	0.38	0.10	0.00
Managers learned about new management practices in the past year	0.85	0.92	0.07	0.00

Notes: Answers from a survey of management practices performed by the National Statistics Office in 2016 on a sample of firms in Portugal. The first column displays the fraction of entrepreneurs without college degrees that report engaging in each practice. The second column does the same for those with college degrees. The third and fourth columns report the difference between the two groups and the p-value from a two-sided test that the difference is zero.

terms replaced by field-by-age interactions. These coefficients therefore capture differences in output at each age between entrepreneurs in each field and those without a college degree, which is the omitted category. I measure output using sales but the results are similar for value added.

Figure 8 plots these coefficients for each field of study. By age 10, they are all positive. Regardless of field, firms started by college-educated entrepreneurs are larger than those without college degrees. But the graph shows there are substantial differences across fields. Those started by entrepreneurs with STEM degrees are the largest, followed by Health, Business, Other and HASS. To get a sense of magnitude, a firm started by a HASS entrepreneur at age 10 is 25% larger on average than one started by an entrepreneur without a college degree, while one started by a STEM entrepreneur is 2.4 times larger. These coefficients do

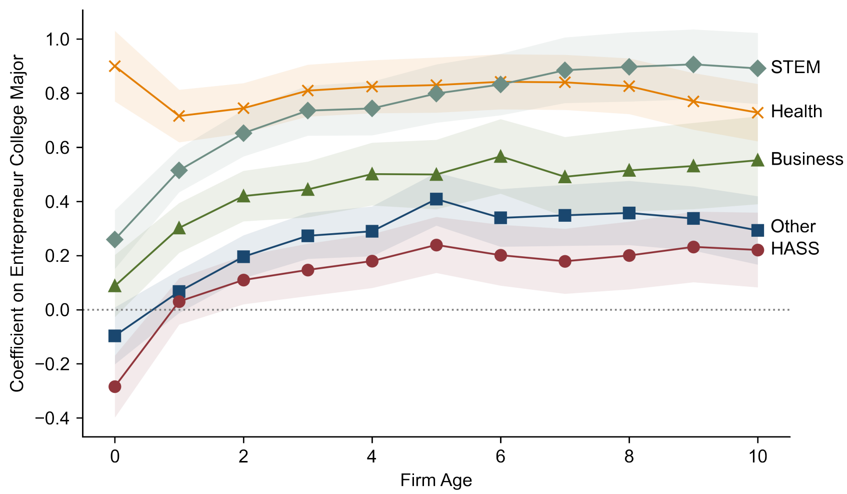

Figure 8: Mechanisms: Evidence from Entrepreneur College Degree Fields

Notes: Entrepreneur college field of study by firm age coefficients from a regression of the form in equation (2), where the entrepreneur schooling by age terms are replaced by field of study by age interactions, estimated on sales data. STEM stands for Science, Technology, Engineering and Math, and HASS stands for Humanities, Arts and Social Sciences. Other includes degrees classified as Architecture, Agriculture and Fishing, Veterinary Medicine, Social Services, Personal Services, Transport Services, Environmental Protection, Security Services and Unknown. The shaded areas represent 95% confidence intervals.

not control for sectors, which are likely to be partly determined by fields of study. Adding five-digit sector-by-year fixed effects to the estimation shows that about half of the gap between STEM and HASS can be explained by sector effects.

The large coefficients for STEM degrees suggest innovation and technology adoption may play an important role. The fourth and last set of results examines heterogeneity across sectors, and points in the same direction. I first sort sectors by technological intensity into five groups, using Eurostat's "High-tech industry and knowledge-intensive services" classification:[24] high-tech manufacturing, which includes Eurostat's high and medium-high-tech manufacturing sectors (e.g. electronics, motor vehicles), high-tech services, which corresponds to knowledge-intensive high-tech services (e.g. computer programming), other manufacturing (e.g. rubber and plastic, textiles), other services (e.g. legal and accounting, health) and finally other sectors, which groups sectors not included in the Eurostat classification, namely agriculture and fishing, mining, utilities and construction. I then estimate equation (2) on sales data separately for each sector group, including 5-digit sector-by-year fixed effects.

Figure 9 plots the schooling-by-age coefficients for each sector group, and reveals stark differences. The largest coefficients are in high-tech manufacturing, followed by high-tech

[24] The classification can be accessed at: https://ec.europa.eu/eurostat/cache/metadata/en/htec_esms.htm

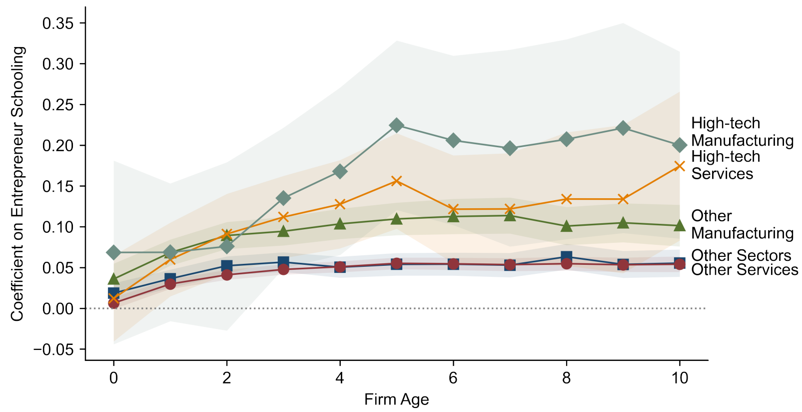

Figure 9: Mechanisms: Evidence from Sector Heterogeneity

Notes: Entrepreneur schooling by firm age coefficients from estimating (2) separately for five sector groups, using Eurostat's "High-tech industry and knowledge intensive services" classification. The regressions are estimated on sales data and include five-digit sector-by-year fixed effects. The shaded areas represent 95% confidence intervals.

services, other manufacturing, other sectors and other services, with the latter two practically indistinguishable. The relationship between entrepreneur schooling and firm dynamics is clearly stronger in technology-intensive sectors. The point estimates imply, for example, that in high-tech manufacturing a firm started by the average entrepreneur in the top schooling group is on average 13 times larger by age 10 than one started by the average entrepreneur in the bottom group. Given the small number of firms in these sectors, the 95% confidence interval for this estimate is wide, ranging from to 3 to 54. The corresponding difference in high-tech services is 9 times. In other manufacturing and other services, it is 4 and 2 times respectively. As the figure shows, most of these differences across sector groups arise from growth, rather than size at entry. Online appendix B.5 presents additional results on sector heterogeneity.

To sum up across the four sets of results, the evidence is consistent with all the mechanisms analyzed except misallocation playing a role in accounting for the relationship between entrepreneur schooling and firm dynamics. The results on college fields and sectors indicate that innovation and technology adoption may be the most important one. This is in line with the pioneering ideas of Nelson and Phelps (1966) and Schultz (1975), as well as the evidence provided by Welch (1970) and, more recently, by Ciccone and Papaioannou (2009).

IV Model

This section lays out a model of firm dynamics motivated by the findings presented above. I then estimate the model and evaluate its ability to match the empirical findings quantitatively. Time is continuous, and for simplicity time subscripts t are omitted throughout, since I focus on the steady-state equilibrium of the model.

IV.A Setup

Final Goods Sector The final consumption good is produced competitively by a representative firm. This good is a CES aggregate of intermediate goods indexed by ω

$$Y = \left(\int y(\omega)^{\frac{\sigma-1}{\sigma}} d(\omega) \right)^{\frac{\sigma}{\sigma-1}} \tag{7}$$

where $\sigma > 1$ denotes the elasticity of substitution across varieties. Intermediate goods are produced by a continuum of firms under monopolistic competition, and each firm faces a demand function given by

$$y(\omega) = Y \left(\frac{p(\omega)}{P} \right)^{-\sigma} \tag{8}$$

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.

Households A measure one of agents in the economy are born and die at exogenous rate δ. At birth, each agent is endowed with a schooling level, and the probability of drawing schooling level $s \in S$ is denoted by θ_s. Schooling in turn determines an agent's human capital e^h when working as an employee, which is constant throughout the agent's life, and productivity e^z when working as an entrepreneur, which evolves over time. I parametrize h and the initial value of z as follows:

$$h = rs + \epsilon$$
$$z = vs + \eta$$

The parameter v is constant across agents, while returns to schooling in the labor market r are heterogeneous and normally distributed with mean $\bar{r}$ and variance σ_r^2. As I show below, heterogeneity in r helps account for the rising variance of firm size with schooling in figure 5.[25] ϵ and η represent residual ability in employment and entrepreneurship, and are normally distributed with mean zero and variances σ_ϵ^2 and σ_η^2. The variables r, ϵ and η are independent from each other and from s. The assumption of normality for these variables keeps the model tractable and turns out to provide a close fit to the data, but it is not an

[25] Card (2001)and Lemieux (2006b) show that heterogeneity in labor market returns to schooling can also help account for several findings in the literature on returns to schooling and on wage inequality.

essential feature of the model. To simplify expressions, let $\sigma_h^2(s) \equiv \sigma_\epsilon^2 + \sigma_r^2 s^2$ denote the variance of h given s.

During an initial growth phase, entrepreneurial productivity grows at a constant rate that also depends on s, given by

$$\dot{z} = \mu_0 + \mu_1 s \tag{9}$$

In the data, differences in firm growth as a function of entrepreneur schooling are concentrated in the early years of the life cycle, as shown in figure 1. To account for this, I follow Luttmer (2011) and assume agents transition randomly from their initial growth phase into a mature phase at rate m. After transition into the mature phase, z remains constant for the rest of the agent's life. This simple setup can reconcile declining average growth with age, as the fraction of mature agents rises, with a Pareto tail in the cross-section, driven by the agents who remain in the growth phase.[26]

At each time t, agents maximize income by choosing to work as entrepreneurs or employees. Each entrepreneur operates an intermediate goods firm with productivity e^z, and receives the firm's profits. Employees inelastically supply their human capital to intermediate goods firms, and are paid a wage w per unit of human capital.

Firms A firm with productivity e^z and employing physical capital K_z and human capital H_z produces output $y = e^z K_z^\gamma H_z^{1-\gamma}$.[27] Physical capital is rented on the world market at a fixed rate ρ. Given ρ and w, entrepreneurs choose prices and inputs to maximize profits

$$\pi(z) = \max_{p, K_z, H_z} \left(p e^z K_z^\gamma H_z^{1-\gamma} - \rho K_z - w H_z \right) \tag{10}$$

subject to (8).

IV.B Equilibrium

Firm optimization implies that profit, sales and labor allocations are given by

$$\pi(z) \propto p(z)y(z) \propto H_z \propto e^{(\sigma-1)z} \tag{11}$$

[26] As Luttmer (2011) shows, it also has the advantage of accounting both for the relatively young age of large firms and for the inverse relationship between the variance of growth rates and firm size observed in the data. Gabaix et al. (2016) and Jones and Kim (2018) employ the same approach to explain the fast rise in income inequality in the United States.

[27] Productivity here measures process efficiency but the model can be generalized to include differences in product quality and demand as well, with equivalent observational implications when z is appropriately interpreted.

and that the capital-labor ratio $\kappa = \frac{K_z}{H_z}$ is constant across firms and equal to $\frac{\gamma}{1-\gamma}\frac{w}{\rho}$. Since entrepreneurial profits increase with z, there is a threshold z^* above which agents choose to become entrepreneurs, and below which they choose to become employed workers. This threshold depends on the agent's employee human capital, and must satisfy

$$\pi(z^*) = we^h \tag{12}$$

Since agents can switch between entrepreneurship and employment at every moment at no cost, occupational choice is a static decision, as in Lucas (1978). The cost of switching into entrepreneurship is simply the foregone wage for that period. Given (11), taking the ratio of this indifference condition for any two agents i and j yields $z_i^* = z_j^* + \frac{h_i - h_j}{\sigma - 1}$. Using this relationship, it is convenient to express the threshold for all schooling and ability combinations as a function of a single threshold, which captures the extent of selection into entrepreneurship in equilibrium. Choosing the threshold for an agent with $h = 0$ as the normalizing threshold, and denoting it by z_0^*, leads to

$$z^* = z_0^* + \frac{h}{\sigma - 1} \tag{13}$$

Labor market clearing in turn pins down w and implies that z_0^* will be given by

$$z_0^* = \frac{\ln\left[(\sigma-1)(1-\gamma)\frac{Z^*}{H^*}\right]}{\sigma - 1} \tag{14}$$

where $Z^* \equiv E[e^{(\sigma-1)z}|z \geq z^*]Pr(z \geq z^*)$ and $H^* \equiv E[e^h|z < z^*]Pr(z < z^*)$.

Given a stationary distribution of productivity, a steady-state equilibrium consists of a threshold z_0^*, a capital-labor ratio κ, profit allocations π and wage w such that occupational choices maximize income, firms maximize profits and the labor market clears.

IV.C Steady-state Productivity

The Cross-Section of Firms The distribution of z at birth for a given s is normal with mean υs and variance σ_η^2, by assumption. Combined with exit at rate δ and transition to maturity at rate m, the constant growth rate in (9) implies that the stationary distribution of cumulative growth in z is exponential with rate:[28]

$$\alpha_s = \frac{\delta + m}{\mu_0 + \mu_1 s} \tag{15}$$

The stationary distribution of z is then given by the sum of independent normal and ex-

[28]See online appendix C.1 for a derivation.

ponential variables, which implies that z follows an exponentially modified Gaussian (EMG) distribution with density $f(z;vs,\sigma_\eta^2,\alpha_s)$ and CDF $F(z;vs,\sigma_\eta^2,\alpha_s)$, where vs, σ_η^2 and α_s are the mean, variance and rate parameters. The distribution of the *level* of productivity e^z, in turn, is given by the product of independent log-normal and Pareto variables, and its right tail is Pareto with index α_s.[29]

Omitting the dependence on parameters, $f(z)$ characterizes the distribution of productivity for all agents with schooling s, including those employed as workers. The distribution of firm productivity, in turn, only includes those working as entrepreneurs, that is those with productivity $z \geq z^*$. Its density $f^*(z)$ is given by multiplying $f(z)$ by the fraction of active entrepreneurs for each z, and dividing by the overall entrepreneurship rate. Given (13), this can be expressed as[30]

$$f^*(z) = \frac{f(z;vs,\sigma_\eta^2,\alpha_s)\Phi\left(\frac{z-z_0^*-\frac{\bar{r}}{\sigma-1}s}{\sigma_h(s)/(\sigma-1)}\right)}{1-F\left(z_0^*;\left(v-\frac{\bar{r}}{\sigma-1}\right)s,\sigma_\xi^2(s),\alpha_s\right)} \tag{16}$$

where $\sigma_\xi^2(s) \equiv \sigma_\eta^2 + \left(\frac{\sigma_h(s)}{\sigma-1}\right)^2$ and Φ denotes the CDF of the standard normal distribution.

Relative to $f(z)$, the left tail of $f^*(z)$ is thinner, since agents with lower z tend to avoid entrepreneurship. This is the effect of selection. For large z, the fraction of active entrepreneurs converges to one, and the shape of $f^*(z)$ converges to that of $f(z)$. In particular, the right tail of the distribution of e^z for active entrepreneurs is also Pareto with index α_s. From (11), firm size is proportional to $e^{(\sigma-1)z}$, which implies that the right tail of the size distribution is Pareto with index $\frac{\alpha_s}{\sigma-1}$. Note that the impact of schooling on the tail is governed by μ_1, the effect of schooling on productivity growth, through (15). If $\mu_1 > 0$, then higher s lowers the index, yielding a thicker right tail in line with the data in figure 6a.

Likewise, the stationary density of employee human capital is given by multiplying the density of h in the population by the fraction of active employees for each h, and dividing by the employment rate:

$$g^*(h) = \frac{\frac{1}{\sigma_h(s)}\phi\left(\frac{h-\bar{r}s}{\sigma_h(s)}\right)F\left(\frac{h}{\sigma-1};vs-z_0^*,\sigma_\eta^2,\alpha_s\right)}{F\left(z_0^*;\left(v-\frac{\bar{r}}{\sigma-1}\right)s,\sigma_\xi^2(s),\alpha_s\right)} \tag{17}$$

where ϕ denotes the density of the standard normal distribution.

Firm Dynamics Next, consider steady-state dynamics for a given cohort. For any entrant, growth conditional on survival will be governed by (9) and by transition to maturity at rate

[29] Since the log-normal tail is thinner than the Pareto, the product inherits the Pareto tail (Gabaix, 2009).

[30] Online appendix C.2 presents a derivation along with the expressions for $f(z;vs,\sigma_\eta^2,\alpha_s)$ and $F(z;vs,\sigma_\eta^2,\alpha_s)$.

m. Since growth stops after maturity and age at maturity is exponentially distributed, expected cumulative growth for a given cohort as its age increases will converge to $\frac{\mu_0+\mu_1 s}{m}$, and convergence will be exponential at rate m. This implies that expected z at age a for a given s, conditional on survival, will be given by

$$E\left(z|s,a\right) = E\left(z|s,a=0\right) + \frac{\mu_0+\mu_1 s}{m}(1-e^{-ma}) \tag{18}$$

Each cohort will include two types of entrants: firms created by consumers born with $z \geq z^*$, who start firms immediately at birth, and firms created by consumers born with $z < z^*$, who start their careers as employees and then reach z^* during their growth phase, at which point they switch into entrepreneurship. $E\left(z|s,a=0\right)$ will be a weighted average of the two types. Selection effects make the derivation of this expectation and its expression somewhat cumbersome, and these are given in online appendix C.3.

IV.D Aggregate Output and TFP

Firm optimization and labor market clearing imply that aggregate output takes the familiar Cobb-Douglas form, which can be expressed as:

$$Y = \left(\frac{K}{Y}\right)^{\frac{\gamma}{1-\gamma}} A^{\frac{1}{1-\gamma}} H \tag{19}$$

where $\frac{K}{Y} = \frac{\gamma(\sigma-1)}{\rho\sigma}$ is the equilibrium capital-output ratio, $H \equiv E(e^h)$ represents employee human capital in the population and TFP A is given by

$$A = Z^{*\frac{1}{\sigma-1}} \left(\frac{H^*}{H}\right)^{1-\gamma} \tag{20}$$

This is the expression for output in the development accounting framework developed by Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999), except that here schooling affects A. Using (16) and (17) to evaluate Z^* and H^* leads to the following expression for A:[31]

$$A = \left(\sum_S \theta_s e^{(\sigma-1)vs+\frac{(\sigma-1)^2\sigma_\eta^2}{2}} \frac{\alpha_s}{\alpha_s-\sigma+1}\Gamma_s^z(z_0^*)\right)^{\frac{1}{\sigma-1}} \left(\frac{\sum_S \theta_s e^{\bar{r}s+\frac{\sigma_r^2}{2}s^2}\Gamma_s^h(z_0^*)}{\sum_S \theta_s e^{\bar{r}s+\frac{\sigma_r^2}{2}s^2}}\right)^{1-\gamma} \tag{21}$$

The first term corresponds to Z^*, and represents average firm productivity as well as the effect of variety. More precisely, the average is a power mean with exponent $\sigma-1$, reflecting the fact that more productive firms are also larger in equilibrium. The effect of schooling on

[31]See online appendix C.4 for a derivation.

productivity at birth is given by vs, and the term in α_s captures the effect of schooling on growth. $\Gamma_s^z(z_0^*)$, in turn, is given by

$$\Gamma_s^z(z_0^*) = 1 - F\left(z_0^*; \left(v - \frac{\bar{r}}{\sigma - 1}\right)s + (\sigma - 1)\sigma_\eta^2, \sigma_\xi^2(s), \alpha_s - \sigma + 1\right)$$

and accounts for the fraction of agents working as entrepreneurs, and hence the extent of variety,[32] as well as a selection effect arising from endogenous occupational choice.

The second term represents $\frac{H^*}{H}$, the fraction of human capital employed in production. The standard effect of schooling on human capital is given by $\bar{r}s$. The additional effect through $\frac{\sigma_r^2}{2}s^2$ is driven by heterogeneity in returns to schooling, and takes this particular form from the assumption of normally distributed returns. The difference between H^* and H lies in the $\Gamma_s^h(z_0^*)$ term, which is given by

$$\Gamma_s^h(z_0^*) = F\left(z_0^*; \left(v - \frac{\bar{r}}{\sigma - 1}\right)s - \frac{\sigma_h^2(s)}{(\sigma - 1)}, \sigma_\xi^2(s), \alpha_s\right)$$

Analogously to $\Gamma_s^z(z_0^*)$, $\Gamma_s^h(z_0^*)$ captures the fraction of agents working as employees and the effect of selection into employment.

IV.E Estimation and Model Fit

To fit the model to the Portuguese data, I employ a combination of calibration, using parameters reported in the literature, and estimation. I base my estimation on cross-sectional data, without targeting firm dynamics directly with the exception of parameter m, as described below. I then evaluate the model's ability to reproduce the patterns documented in section III for both the cross-section and for firm dynamics.

I set three parameters a priori. First, I set $\sigma = 3$, following Hsieh and Klenow (2009, 2014). This choice has little impact on the model's ability to match the empirical findings on firm size and growth described above, but does affect the magnitude of the productivity differences inferred from the data, and hence the implied effect of schooling on productivity. I examine the sensitivity of the cross-country results presented below to setting $\sigma = 4$ and $\sigma = 5$. Second, I set $\gamma = 1/3$, approximately equal to one minus the labor share of income in Portugal (Stehrer et al., 2019) and a standard value in the literature. Third, I set $\bar{r} = 8\%$, the mid-point of the 6%-10% range estimated in the literature on returns to schooling (Card, 1999).[33]

[32] Note that there are no scale effects from variety because population size is normalized to one. These can be thought of as part of the TFP residual in the development accounting exercise below.

[33] Note that I do not set a value for the capital rental rate ρ, which simply acts as an aggregate output shifter through its effect on the capital-output ratio in (19), and plays no role in the analysis that follows.

Turning to the estimation, I start with α_s, the tail index of the productivity distribution for each level of schooling. Recall that table 2 presents estimates of the tail index for the size distribution, which correspond to $\frac{\alpha_s}{\sigma-1}$, for each of the five schooling groups in the data. Given σ, estimates of α_s can be recovered for each group. Taking estimates of α_{s_i} and α_{s_j} for any two values of schooling s_i and s_j, equation (15) implies that

$$\mu_0 = (\delta + m)\frac{s_i\alpha_{s_i} - s_j\alpha_{s_j}}{\alpha_{s_i}\alpha_{s_j}(s_i - s_j)} \tag{22}$$

$$\mu_1 = (\delta + m)\frac{\alpha_{s_j} - \alpha_{s_i}}{\alpha_{s_i}\alpha_{s_j}(s_i - s_j)} \tag{23}$$

Plugging back into (15) shows that α_s for any s can be expressed as a function of s_i, s_j, α_{s_i} and α_{s_j}:

$$\alpha_s = \frac{\alpha_{s_i}\alpha_{s_j}(s_j - s_i)}{s(\alpha_{s_i} - \alpha_{s_j}) - s_i\alpha_{s_i} + s_j\alpha_{s_j}} \tag{24}$$

Combining (24) with the empirical estimates of $\frac{\alpha_s}{\sigma-1}$ for the top and bottom schooling groups in table 2, along with average years of schooling in each of the two groups, I recover α_s for any level of schooling.

Since (24) is independent of δ and m, these two parameters are not pinned down by the cross-sectional estimates of α_s and must be determined separately. Note, however, that α_s is a sufficient statistic for the effect of schooling on aggregate output through productivity growth in (21), so any combination of δ, m, μ_0 and μ_1 that is consistent with a given value of α_s will lead to the same aggregate implications.

I set $\delta = \ln(1 - 0.107) = 0.113$, where 0.107 is the average annual firm exit rate in the data. The parameter m decreases the long run value of expected productivity growth in equation (18), while increasing the rate of convergence towards that value. Following the procedure developed by Chamberlain (1984), I estimate m by minimizing the distance between expected firm growth as implied by the model, conditional on the parameters determined above, and the unrestricted $\beta_{s,a}$ schooling-by-age coefficients estimated using equation (1) and plotted in figure 1. I assign equal weights to all coefficients in the estimation. While the $\beta_{s,a}$ coefficients for older ages represent fewer firms, they contain valuable information on long run growth and on the rate of convergence, and therefore play an important role in identifying m. In addition, I exclude growth between ages zero and one from the estimation, since it is biased upward by the fact that firms at age zero only report sales for part of the year. The details of the estimation are explained in online appendix D.

Panel A of table 4 reports estimates for both m and the implied values of μ_0 and μ_1 as given by (22) and (23). Standard errors for μ_0 and μ_1 are obtained via the delta method. Column one reports my baseline estimates when firm output is given by sales. This corresponds

Table 4: Model Estimation Results

A. Minimum distance estimates for components of α_s

	Sales		Value Added	
	Overall	Within-sector	Overall	Within-sector
m	0.1925	0.2060	0.2375	0.1953
	(0.0542)	(0.0590)	(0.0694)	(0.0573)
μ_0	0.0846	0.0674	0.0776	0.0610
	(0.0150)	(0.0125)	(0.0154)	(0.0114)
μ_1	0.0033	0.0041	0.0049	0.0039
	(0.0006)	(0.0008)	(0.0010)	(0.0007)
N	50	50	50	50

B. Maximum likelihood estimates for remaining parameters

	Sales		Value Added	
	Overall	Within-sector	Overall	Within-sector
v	0.0436	0.0446	0.0409	0.0463
	(0.0001)	(0.0000)	(0.0001)	(0.0000)
σ_η	0.9337	0.9505	0.9292	1.0043
	(0.0012)	(0.0008)	(0.0010)	(0.0008)
σ_ϵ	0.7312	0.6120	0.7276	0.6384
	(0.0023)	(0.0019)	(0.0022)	(0.0020)
σ_r	0.0713	0.0553	0.0646	0.0558
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
N	656 771	656 771	598 304	598 304

Notes: Prior to the estimation, output is residualized on the set of controls in (1). In columns one and two output is measured by sales, and in columns three and four by value added. In columns two and four the set of controls used to calculate residuals includes 5-digit sector-by-year fixed effects.

to the values of $\frac{\alpha_s}{\sigma-1}$ reported in column one of table 2 and to the coefficients plotted in figure 1. The estimated value of m is 0.1925, which implies that entrepreneurs mature relatively fast. Five years after entry, for example, only $e^{-5\times0.1925} = 38\%$ of surviving entrepreneurs will remain in the growth phase. This is consistent with the evidence showing that average firm growth is driven by a small subset of high growth businesses (Decker et al., 2014). The values for μ_0 and μ_1 equal 0.0846 and 0.0033 respectively, which implies growth rates ranging from 0.0846 for an entrepreneur with no schooling to 0.1412 for an entrepreneur with a college degree. The remaining columns report estimates for sales using within-sector estimates of α_s and schooling-by-age coefficients, as well as estimates for value-added, both using overall and within-sector α_s and coefficients. The results are similar across specifications.

I then estimate v, σ_η, σ_ϵ and σ_r by maximum likelihood on the cross-sectional data plotted in figure 5a, using the density of firm productivity given by (16) and with the equilibrium threshold z_0^* determined by (14). To recover log productivity z, I divide log sales by $\sigma - 1$, as implied by (11).[34] In addition, I constrain the estimation such that the model-implied entrepreneurship rates for the top and bottom schooling groups match those observed in the 2017 cross-section, which equal 18.5% and 11.2% respectively.[35] This restriction is important in identifying v, which is a key determinant of differences in entrepreneurship rates across schooling levels in the model.

Panel B of table 4 presents the results. Column one reports my baseline estimates when firm output is given by sales. I find that $v = 0.0436$, $\sigma_\eta = 0.9337$, $\sigma_\epsilon = 0.7312$ and $\sigma_r = 0.0713$, all precisely estimated. Columns two to four report similar estimates for the same alternative samples as panel A.

Figure 10 offers a visual summary of the estimated model for the cross-section. The z line represents the cross-sectional distribution of productivity in the population, given the distribution of schooling in the data. This is not directly observed, since it includes the potential productivity of individuals who choose to work as employees rather than entrepreneurs. The z^* line represents the distribution of productivity thresholds for entrepreneurship in the population, also unobserved. The $z \geq z^*$ line, in turn, is the distribution of productivity for active entrepreneurs, which includes only those individuals who select into entrepreneurship. This is what is directly observed in the data and targeted in the estimation.

Since z^* is a function of heterogeneous h, the distribution of z for active entrepreneurs is not simply a truncated version of the distribution for the entire population, as in Lucas (1978) or Melitz (2003). Instead, it is compressed and shifted to the right. The left tail, in particular, closely tracks the left tail of the distribution of z^*, highlighting the role that the variance of h plays in determining the effect of selection on the observed distribution of productivity. When the variance is low, selection resembles a truncation at the mean of z^*. The higher the variance, the higher the chance that an individual with low z will still select into entrepreneurship, and the lower the chance that an individual with high z will do the same, weakening the effect of selection.

How well does this estimated productivity distribution fit the cross-sectional data on firm size? Figure 11 plots theoretical densities from the estimated model and histograms of log

[34] I drop the bottom 1% of the sample by level of schooling in the estimation, which includes very small firms that are unlikely to be fully active. In the baseline sample of sales, for example, this consists of firms with less than 3,000 euros in sales. I account for this truncation in the estimation. I also exclude age zero firms, since these only report sales and value added for part of the year. In addition, I add a constant to z in the estimation to match the units in which output is expressed in the data (i.e. euros). I don't report its value since it just acts as an aggregate TFP shifter in the model, but the values of model-implied productivity and firm size plotted in the figures below include it, to facilitate comparison with the data.

[35] I implement this by constraining the values of v and σ_η to match these moments given the remaining parameter values. Standard errors for v and σ_η are then obtained by the delta method.

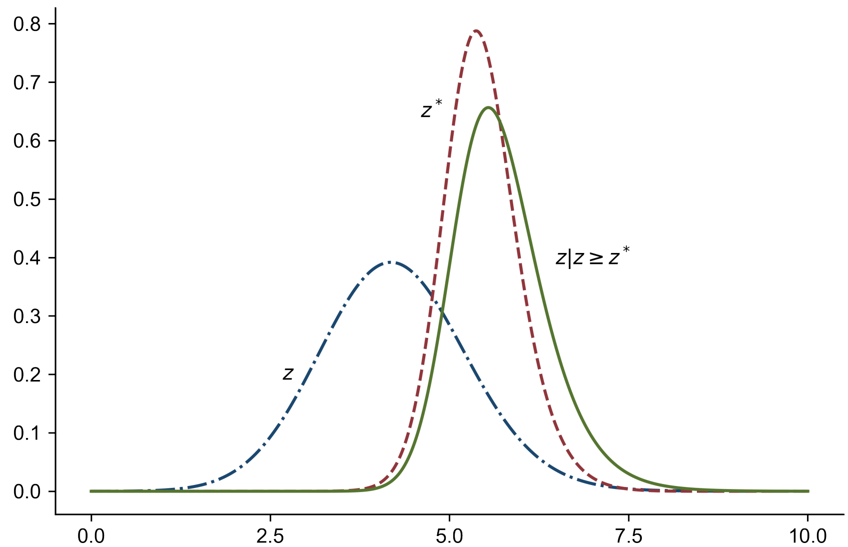

Figure 10: Model-implied Distribution of Productivity

Notes: Distributions from the baseline specification of the estimated model. z is entrepreneurial productivity in the overall population, z^* is the productivity threshold for entry into entrepreneurship and $z|z \geq z^*$ is productivity among entrepreneurs.

sales in the data separately for each level of schooling. In each case, the fit is remarkably accurate. In particular, the model can account for the right-shift and dilation of the distribution as schooling increases, as well as for the noticeably thicker right tail for higher levels of schooling. The results are similar using value added or within-sector data.

To shed light on the roles of σ_r, σ_ϵ and σ_η, I estimate restricted versions of the model when each of these parameters is constrained to equal zero. Online appendix figure F.9 plots densities for the top and bottom groups by schooling from these restricted models. Under homogeneous returns to schooling in the labor market, when $\sigma_r = 0$, the model cannot account for the increasing dilation of the distribution with schooling. Without residual employee ability, when $\sigma_\epsilon = 0$, it overpredicts differences in dilation. And without residual entrepreneurial ability, when $\sigma_\eta = 0$, it yields distributions that are truncated and compressed relative to the data.[36]

Figure 11 shows that the model fits the cross-sectional data it was estimated on. Figure 12 turns to life cycle dynamics, and plots average sales by age in the estimated model, using (18), as well as the schooling-by-age coefficients estimated from the data and displayed in figure 1. Model-implied size is calculated using average years of schooling within each group. As in the estimation, I exclude age zero firms, which report incomplete sales data. Overall,

[36]The truncation in this case arises because z follows an exponential distribution with rate α_s and location vs, instead of an EMG distribution. Its minimum value vs is the value of productivity at birth for agents with schooling s.

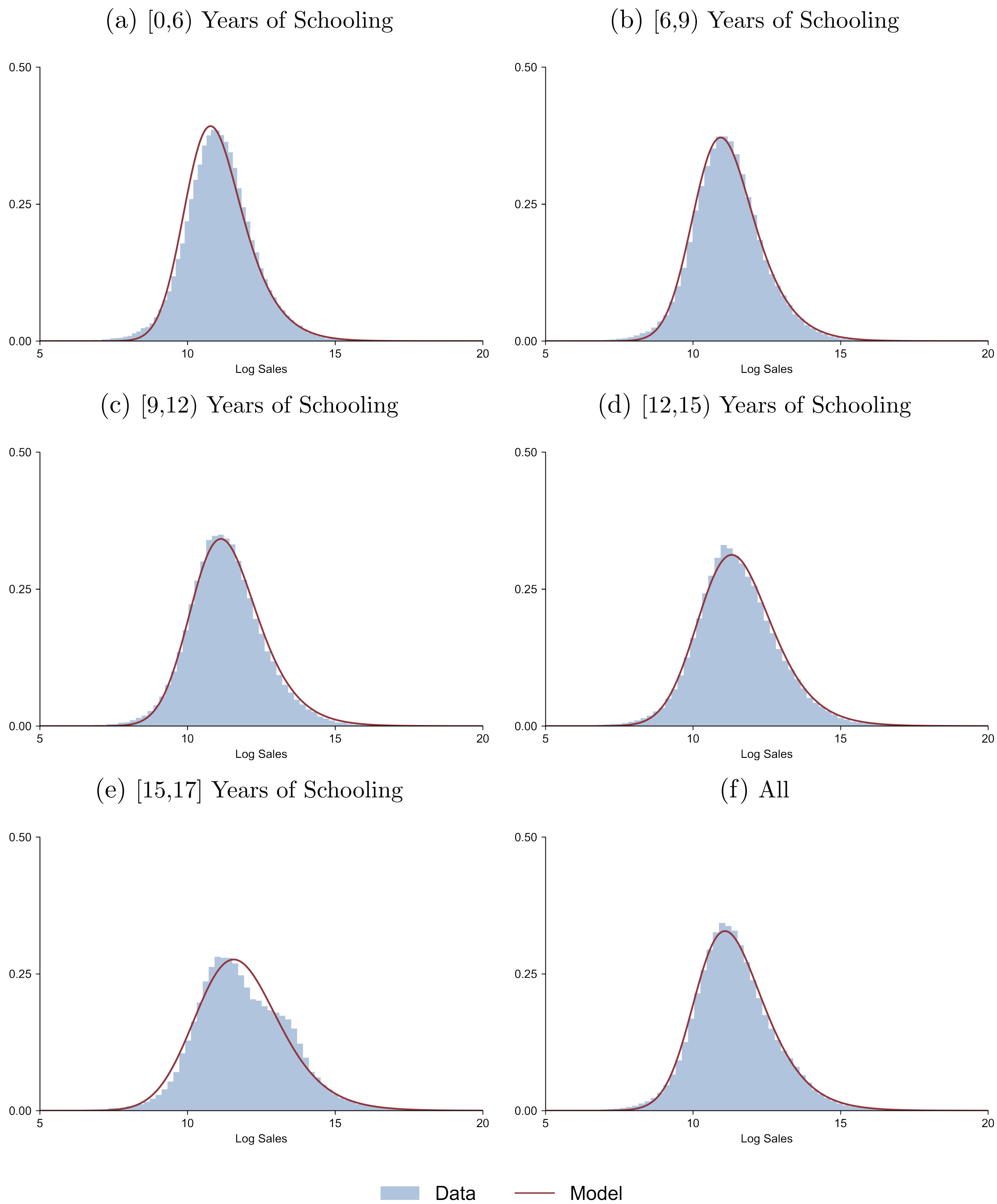

Figure 11: Sales Distributions: Data and Model

Notes: Histograms of log sales in the sample and densities of log sales in the baseline specification of the estimated model. Panels a) to e) present histograms and model densities by level of entrepreneur schooling, where model densities are evaluated at the mean of entrepreneur schooling in the sample within each group. Panel f) aggregates the model densities in panels a) to e), weighted by population shares and model-implied entrepreneurship rates for each level, and plots the resulting density along with log sales in the whole sample.

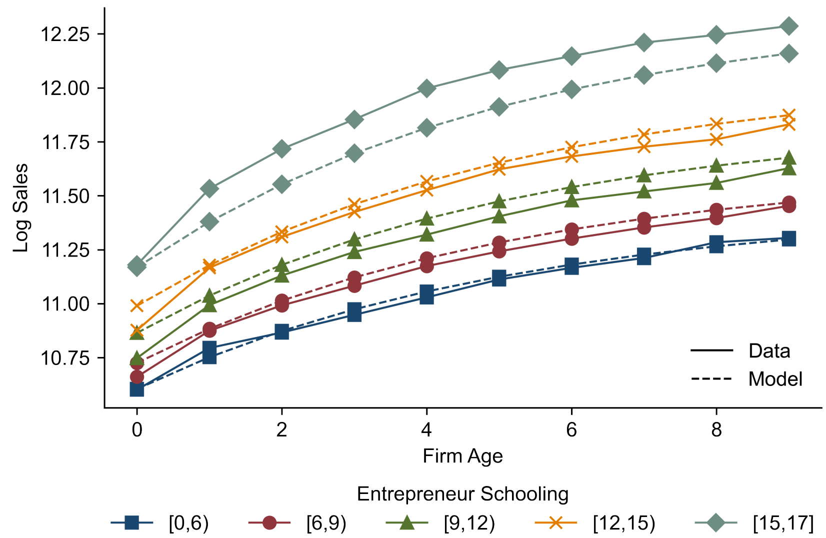

Figure 12: Firm Dynamics: Data and Model

Notes: Entrepreneur schooling group by firm age coefficients from figure 1 and average log sales by age in the baseline specification of the estimated model for the same entrepreneur schooling groups. Model-implied sales are evaluated at the mean of entrepreneur schooling in the sample within each group.

both size at entry and growth over the life cycle are close to the actual values in the data. The same holds for value added and for within-sector estimates.

This shows that even though the estimated model targets the cross-sectional data only, with the exception of m, it can also accurately reproduce the data on firm dynamics. In particular, the values of α_s estimated from the cross-sectional upper tail are quantitatively consistent with the levels and differences in life cycle growth across schooling levels observed in the firm dynamics data.

V Cross-country Implications

Next, I explore the model's implications for the aggregate effect of human capital on output. I first show how aggregate returns to schooling differ from individual returns, and then perform a development accounting exercise to evaluate the ability of human capital to account for cross-country differences in output. To facilitate comparison, I use the dataset from Caselli (2005), who computes output and physical capital per worker from Penn World Tables data (Heston, Summers and Aten, 2002) and human capital per worker from the educational attainment data in Barro and Lee (2001) for 94 countries in 1996. I also use the Barro and Lee (2001) data directly to obtain attainment shares and the duration of each level of schooling across countries. 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.

V.A Aggregate returns to schooling

In the standard development accounting framework proposed by Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999), output per capita takes the form in (19), with $H = e^{\bar{r}s}$, s representing the schooling level of a representative agent and A independent of s.[37] In this framework, $\frac{\partial \ln Y}{\partial s} = \frac{\partial \ln H}{\partial s} = \bar{r}$. This implies that aggregate and individual returns to schooling are the same, which is what underpins the use of individual returns to infer aggregate returns.

When s is homogeneous, aggregate returns to schooling in my model are given by differentiating the log of output in (19) with respect to s:

$$\frac{\partial \ln Y}{\partial s} = \frac{1}{1-\gamma}\frac{\partial \ln A}{\partial s} + \frac{\partial \ln H}{\partial s} \tag{25}$$

The key difference is that, as in Lucas (1978) and Murphy, Shleifer and Vishny (1991), an effect of entrepreneurial human capital on productivity at the firm level translates into an effect on TFP at the aggregate level.[38] This is captured by the $\frac{\partial \ln A}{\partial s}$ term, with $\frac{1}{1-\gamma}$ representing the effect of TFP on the capital-labor ratio. Ignoring selection effects through Γ_s^z and Γ_s^h for the moment, (11) and (20) imply that the effect of human capital on TFP is proportional to the return to schooling in entrepreneurship:[39]

$$\frac{\partial \ln A}{\partial s} = \frac{1}{\sigma-1}\frac{\partial \ln E\left[e^{(\sigma-1)z}\right]}{\partial s} = \frac{1}{\sigma-1}\frac{\partial \ln E\left[\pi(z)\right]}{\partial s}$$

If output exhibits constant returns to scale in human and physical capital conditional on TFP, a standard assumption in development accounting, then an effect of human capital on TFP leads to increasing returns: aggregate returns to schooling in (25) are given by the *sum* of individual returns in employment and entrepreneurship.

Increasing returns imply that the model features a pecuniary human capital externality, and some studies find that human capital externalities are too small to account for a meaningful difference between individual and aggregate returns to schooling (Acemoglu and

[37]Hall and Jones (1999) adopt a piecewise linear formulation that allows $\bar{r}$ to vary with s, which I abstract from here for simplicity.

[38]A growing literature explores the connection between firm-level productivity and aggregate TFP in models of heterogeneous firms. See Hopenhayn (2014) for a review.

[39]The $\frac{1}{\sigma-1}$ factor multiplying entrepreneurial returns to schooling accounts for a market share effect that influences individual but not aggregate returns. Increasing the productivity of an individual entrepreneur raises that entrepreneur's profits both by increasing output holding inputs fixed and by increasing that entrepreneur's share of inputs, at the expense of other firms. The effect operating through a higher input share is driven by $\sigma - 1$, the elasticity of firm size with respect to z in (11).

Angrist, 2000; Ciccone and Peri, 2006), even if other studies report larger values (Moretti, 2004a; Iranzo and Peri, 2009). However, the externality in this model is not captured by the estimates in these studies, which rely on differences in wage levels across locations. The same holds for estimates that rely on differences in housing costs (Rauch, 1993) or in the average revenue product of inputs (Moretti, 2004b).

To see this, consider an extension of the model with two cities, each endowed with a fixed supply of housing. Agents must pay housing rent ζ_i per unit of housing to live and work in city i. Let utility be given by $u = c^{1-\psi}d^{\psi}$, where c and d denote consumption and housing respectively, so that agents maximize utility by devoting a share ψ of income to housing and the remainder to consumption. This implies that utility in city i can be expressed as $u_i = \frac{w_i e^h}{\zeta_i^{\psi}}$ for an employee and as $u_i = \frac{\pi_i(z)}{\zeta_i^{\psi}}$ for an entrepreneur. Cobb-Douglas utility simplifies expressions but the argument is more general.

As in standard models of externalities, output can be traded costlessly across cities, entrepreneurs and employees are fully mobile and rents adjust to clear the housing market. The final consumption good combines intermediate goods from both cities according to (7), regardless of where they are produced. Within cities, wages are equalized across firms and differences in firm productivity are reflected in firm size. If there are no advantages to locating in one city over the other, full mobility implies that wages and rents must also be equalized across cities in equilibrium. The spatial distribution of human capital is determined by housing supply, and all else remains the same as in the baseline model.

More generally, suppose that each city offers entrepreneurs an amenity $B_i(z)$, so that utility for entrepreneurs located in city i equals $B_i(z)u$. $B_i(z)$ can also be interpreted as a factor that increases firm productivity, with similar implications. Entrepreneurs are attracted to the city with the higher amenity given their productivity, causing labor demand and wages to rise there. This in turn attracts employees into that city, lowering the marginal revenue product of labor and wages. What prevents full wage equalization when cities differ in amenities is the fixed housing supply, which implies migration pushes up rents. Higher rents also limit the extent of migration even if $B_1(z) > B_2(z)$ for all z. In equilibrium, rents offset differences in wages so that employees are indifferent between the two cities:

$$\frac{w_1}{w_2} = \left(\frac{\zeta_1}{\zeta_2}\right)^{\psi}$$

To generate variation in average entrepreneurial human capital across cities, let $B(z) = \frac{B_1(z)}{B_2(z)}$ and assume $B'(z) > 0$. Since the relative value of the amenity in city 1 increases with z but rents do not, entrepreneurs with higher human capital select into city 1: those above a threshold z' choose to locate in city 1, and those below that threshold choose to locate in city 2. Then profit maximization in (10) and the indifference condition for employees imply that the marginal entrepreneur will be indifferent between cities when wages satisfy

$$\frac{w_1}{w_2} = B(z')^{\frac{1}{\sigma-\gamma(\sigma-1)}}$$

Wages in city 1 may be higher or lower than in city 2, depending on the relative value of the amenity for the marginal entrepreneur. The same holds for differences in rents and in the average revenue product of inputs, both of which are a function of $B(z')$. The key implication, as pointed out above, is that estimates of human capital externalities based on cross-city variation in these factors fail to capture the effect of entrepreneurial human capital on TFP through firm productivity.

An additional source of amplification in (25) is heterogeneity in returns to schooling, which raises the contribution of individual returns in both occupations. Recall that $H \equiv E(e^h)$. The $\frac{\partial \ln H}{\partial s}$ term in (25) thus corresponds to $\frac{\partial \ln E(e^h)}{\partial s}$, the arithmetic mean of returns to schooling in employment. Yet what is estimated from Mincerian regressions of log individual earnings on years of schooling is $\frac{\partial E(\ln e^h)}{\partial s}$, the geometric mean return. Under homogeneous returns to schooling, this distinction is irrelevant. But when returns are heterogeneous, $\frac{\partial E(\ln e^h)}{\partial s} < \frac{\partial \ln E(e^h)}{\partial s}$, by Jensen's inequality. This implies that the Mincerian estimates of individual returns typically used in development accounting understate $\frac{\partial \ln H}{\partial s}$ in the presence of heterogeneity.

In the model, geometric and arithmetic mean returns in employment are respectively equal to

$$\frac{\partial E(\ln e^h)}{\partial s} = \bar{r}$$
$$\frac{\partial \ln E(e^h)}{\partial s} = \bar{r} + \sigma_r^2 s$$

The difference between the two means is increasing in s and in σ_r^2, the variance of returns. Given $\bar{r} = 8\%$ and my baseline estimate of $\sigma_r = 0.0713$, the arithmetic mean return for an employee with six years of schooling, which corresponds to the mean of average years of schooling across countries in Caselli (2005), equals 11.0%.

The same applies to returns for entrepreneurs. Geometric and arithmetic mean returns in this case are given by

$$\frac{\partial \ln \pi[E(z)]}{\partial s} = (\sigma - 1)\left(v + \frac{\mu_1}{m + \delta}\right)$$
$$\frac{\partial \ln E[\pi(z)]}{\partial s} = (\sigma - 1)\left(v + \frac{\mu_1}{\delta + m - (\sigma - 1)(\mu_0 + \mu_1 s)}\right)$$

The difference between the two means lies in the term involving μ_1, which shows how heterogeneity in returns in the model is linked to the effect of schooling on growth. As firms age, differences in growth rates cause productivity differences as a function of schooling to widen. My baseline estimate for the geometric mean return equals 10.9%.[40] The arithmetic mean, which determines $\frac{\partial \ln A}{\partial s}$, equals 15.6% for an entrepreneur with six years of schooling.

With these expressions in hand, aggregate returns can be written as

$$\frac{\partial \ln Y}{\partial s} = \frac{1}{1-\gamma}\left(v + \frac{\mu_1}{m + \delta - (\sigma - 1)(\mu_0 + \mu_1 s)}\right) + \bar{r} + \sigma_r^2 s \tag{26}$$

The calculations above imply that aggregate returns to schooling for a population with $s = 6$ equal 22.8%, with 11.7% coming from the effect of entrepreneurial human capital on TFP. To incorporate the effects of heterogeneity in s and selection, I proceed numerically. I first compute aggregate output, given by (19), for the 94 countries in Caselli (2005) as a function of the distribution of educational attainment in the population, θ_s, using my baseline parameters and ensuring that z_0^* for each country solves (14). I then regress this predicted log output on average years of schooling across countries. Figure 13 plots the data and regression line, and shows that the linear slope provides an accurate summary of aggregate returns to schooling. Incorporating heterogeneity in s through this procedure yields an aggregate return of 25.0%, and further adding the effect of selection increases it to 25.9%. The latter case corresponds to the slope in the figure.

The estimated return to schooling in entrepreneurship is sensitive to the value of σ. As equation (11) shows, σ governs the magnitude of productivity differences inferred from data on firm size, and hence the implied effect of schooling on productivity. I follow Hsieh and Klenow (2009) in setting $\sigma = 3$ as a baseline, which yields the aggregate return of 25.9% just described. For $\sigma = 4$ and $\sigma = 5$, in line with the range of values examined by Hsieh and Klenow (2009), aggregate returns equal 21.6 and 19.5 percent respectively.

In short, I find that aggregate returns to schooling are between two and three times larger than $\bar{r}$. For comparison, Acemoglu and Angrist (2000) report a slope of 0.29 from a univariate regression of log aggregate output on average years of schooling. Accounting for the effect of human capital on TFP through firm productivity and for heterogeneity can therefore bridge most of the gap between micro and macro returns to schooling.

[40] Gennaioli et al. (2013) estimate significantly larger Mincerian returns for entrepreneurs. One possible reason for this is that the World Bank Enterprise Survey data they use focuses on large firms: the average firm in their sample employs 110 employees, which corresponds to the 99.5th percentile of employment in my baseline sample. Their estimates are therefore likely to be representative of returns for firms in the upper tail of the size distribution, which figure 5b suggests are much higher than average returns. In my model, returns in the tail are higher because those are the firms that have experienced differences in growth rates as a function of schooling over a longer period of time.

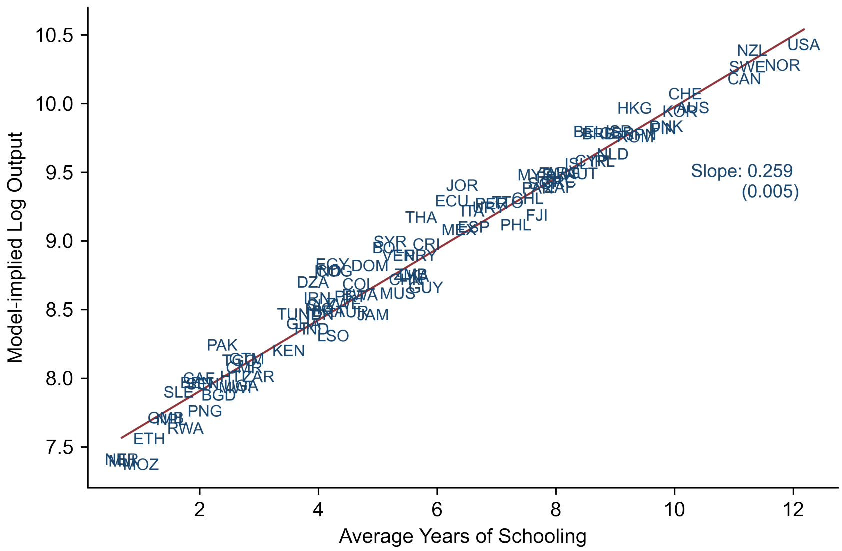

Figure 13: Aggregate Return to Schooling in the Model

Notes: Model-implied log aggregate output as a function of the distribution of educational attainment in the population, under the baseline specification of the estimated model, and average years of schooling for the 94 countries in Caselli (2005), along with the corresponding regression line. Attainment shares by level of schooling and average years of schooling are from Barro and Lee (2001).

V.B Development Accounting

The previous section shows that the model implies aggregate returns to schooling substantially larger than individual returns. I next evaluate the extent to which these larger aggregate returns can account for differences in aggregate output across countries.

The literature has proposed two main accounting methods for this. In Klenow and Rodriguez-Clare (1997), the fraction of cross-country income differences explained by human and physical capital is given by $\frac{\text{cov}[\ln(Y/L),\ln(\hat{Y}/L)]}{\text{var}[\ln(Y/L)]}$, where Y represents actual output in the data and $\hat{Y}$ is counterfactual output in a factor-only model. This corresponds to a variance decomposition of $\ln(Y/L)$, where half of the covariance term between $\hat{Y}$ and residual TFP is assigned to the factor-only model. Letting $\hat{Y}$ be given by (19) and denoting residual TFP by $\tilde{A}$, their approach yields

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

The method developed by Caselli (2005) differs in two respects. First, the fraction of cross-country income differences explained by human and physical capital is simply given by $\frac{\text{var}[\ln(\hat{Y}/L)]}{\text{var}[\ln(Y/L)]}$, which assigns the covariance between $\hat{Y}$ and residual TFP entirely to the contribution of residual TFP. Second, it does not account for the endogenous response of physical capital to higher levels of residual TFP and human capital, and instead starts from output expressed as $Y = AK^{\gamma}H^{1-\gamma}$. Applied to the model developed above, this method leads to

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

To facilitate comparison, I implement these two methods using the the same dataset as Caselli (2005), who summarizes the literature's findings. I start by replicating Caselli's baseline results, which correspond to the case where A is assigned to the TFP residual, s is homogeneous and equal to average years of schooling in the population, and $\sigma_r = 0$. In this benchmark case, physical and human capital account for 40% of per capita income differences in the first method, and 39% in the second. Following Hall and Jones (1999), Caselli uses a piecewise linear formulation where the return to schooling decreases with s. Setting $\bar{r} = 8\%$, as I do in my model, decreases these fractions to 35% and 34% respectively.

Table 5 summarizes the results from implementing the two methods under my estimated model. I use educational attainment shares from Barro and Lee (2001) to measure θ_s, $\frac{K}{L}$ as reported by Caselli (2005), and solve for z_0^* in each country using (14). I report results for $\sigma = 3$, $\sigma = 4$ and $\sigma = 5$.

Panel A presents estimates under the first method, and the first row presents the overall results. Under my baseline estimates, the fraction of income differences explained by human and physical capital increases to 75% when $\sigma = 3$, 66% when $\sigma = 4$ and 61% when $\sigma = 5$, as reported in columns one to three. Columns four to six report very similar results when I estimate the model using value added data instead of sales. As with aggregate returns, these estimates can substantially narrow the gap between the 40% benchmark and the 78% reported by Mankiw, Romer and Weil (1992) based on cross-country regressions.

The remaining rows shed light on the sources of this increase. I focus on the first column here, where $\sigma = 3$ and the model is estimated on sales data, but the remaining columns yield the same conclusions.

The two key channels that amplify the effect of human capital in the model are increasing returns through the effect of schooling on TFP and heterogeneity. The second row shuts down the effect of schooling on TFP by assuming constant A across countries. Relative to the base case in Caselli (2005), this only adds the effect of heterogeneity in labor market returns to schooling r, driven by σ_r. In this case, the fraction of income differences explained by the model drops from 75% to 44%, implying that heterogeneity in r by itself has a small impact. The bulk of the difference is driven by the effect of schooling on TFP.

The third, fourth and fifth rows highlight the effects of schooling on firm productivity at birth, on productivity growth and on selection separately, by respectively setting $\upsilon = 0$,

Table 5: Development Accounting

A. Klenow and Rodriguez-Clare (1997) method

	Sales			Value Added		
	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$
Overall	0.75	0.66	0.61	0.74	0.64	0.60
No effect on productivity	0.44	0.45	0.45	0.42	0.43	0.44
No effect on productivity at birth	0.60	0.56	0.54	0.60	0.56	0.54
No effect on productivity growth	0.59	0.54	0.52	0.57	0.52	0.51
No selection	0.73	0.64	0.59	0.72	0.62	0.58
No heterogeneity	0.61	0.56	0.53	0.60	0.54	0.52
Within-sector	0.69	0.60	0.56	0.68	0.59	0.56

B. Caselli (2005) method

	Sales			Value Added		
	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$
Overall	0.78	0.66	0.60	0.76	0.64	0.59
No effect on productivity	0.42	0.43	0.44	0.41	0.42	0.43
No effect on productivity at birth	0.59	0.55	0.53	0.59	0.54	0.52
No effect on productivity growth	0.58	0.53	0.51	0.55	0.51	0.49
No selection	0.75	0.63	0.58	0.74	0.62	0.57
No heterogeneity	0.61	0.54	0.52	0.59	0.53	0.50
Within-sector	0.70	0.59	0.55	0.69	0.58	0.54

Notes: Development accounting calculations using the estimated model and cross-country data from Caselli (2005), under the methods developed by Klenow and Rodriguez-Clare (1997) in panel A and by Caselli (2005) in panel B. Results are reported for different levels of σ. In the first three columns the model is estimated on sales data, and in the remaining three columns it is estimated on value added. The first row in each panel presents the overall result, and the remaining rows report results when specific channels in the model are shut down.

$\mu_1 = 0$ and $\Gamma_s^z = \Gamma_h^z = 1$. The latter is equivalent to assuming that agents sort randomly into entrepreneurship and employment, regardless of their human capital. The fraction drops to 60% in the case of no effect at birth, 59% when there is no effect on growth and 73% when selection is shut down. The effects at birth and on growth both play significant roles, while the effect through selection is minimal.

The sixth row removes the effect of heterogeneity by setting s equal to average years of schooling in each country, $\sigma_r = 0$ and letting z be constant and equal to its expected value in the population, $vs + \frac{\mu_0 + \mu_1 s}{m + \delta}$. The fraction drops to 61%, which still represents an increase of 21 percentage points over the base case of 40%. Heterogeneity amplifies the effect of schooling on TFP considerably, but the effect is large even under homogeneity.

Finally, the last row uses within-sector parameter estimates, and the fraction drops to 69%. In line with the evidence on firm dynamics, most of the effect is driven by within-sector differences at the 5-digit level, not by selection into more productive sectors.

I find very similar results under Caselli's method, which are reported in panel B. Under my baseline estimates for sales, the fraction of income differences explained by human and physical capital equals 78% when $\sigma = 3$, 66% when $\sigma = 4$ and 60% when $\sigma = 5$.

VI Conclusion

I find 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. Size at entry and growth increase strongly with entrepreneur schooling, as does the thickness of the right tail of the cross-sectional size distribution. The evidence is consistent with several mechanisms, with innovation and technology adoption having a prominent role. Non-entrepreneurial human capital, on the other hand, seems to matter much less for firm dynamics.

A simple model of firm dynamics where schooling affects both the initial level and the growth rate of entrepreneurial productivity, as well as the value of employee human capital, can quantitatively match the empirical findings. Entrepreneurial human capital affects TFP through firm productivity, which leads to increasing returns: aggregate returns to schooling are given by the sum of individual returns in employment and in entrepreneurship. In addition, heterogeneous returns to schooling imply that Mincerian regressions understate the contribution of individual returns in each occupation. These effects can substantially narrow the gap between existing estimates of returns to schooling at the micro and macro levels, and amplify the role of human capital in an otherwise standard development accounting exercise.

Data Availability Statement The administrative data on firms and workers used in this article were provided by the National Statistics Office in Portugal. Details on how to access the data and supplementary material are available at https://doi.org/10.5281/zenodo.5464854.

References

Acemoglu, Daron, and Joshua Angrist. 2000. "How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws." *NBER Macroeconomics Annual*, 15: 9.

Aghion, Philippe, Celine Antonin, Simon Bunel, and Xavier Jaravel. 2020. "What Are the Labor and Product Market Effects of Automation? New Evidence from France." *Working Paper*.

Akcigit, Ufuk, Harun Alp, and Michael Peters. 2021. "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries." *American Economic Review*, 111(1): 231–275.

Axtell, R. L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science*, 293(5536): 1818–1820.

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

Baumol, William J. 1990. "Entrepreneurship: Productive, Unproductive, and Destructive." *Journal of Political Economy*, 98(5): 893–921.

Becker, Gary S., and Kevin M. Murphy. 1994. "The Division of Labor, Coordination Costs, and Knowledge." In *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*. 299–322. The University of Chicago Press.

Benhabib, Jess, and Mark M. Spiegel. 1994. "The Role of Human Capital in Economic Development Evidence from Aggregate Cross-Country Data." *Journal of Monetary economics*, 34(2): 143–173.

Bento, Pedro, and Diego Restuccia. 2017. "Misallocation, Establishment Size, and Productivity." *American Economic Journal: Macroeconomics*, 9(3): 267–303.

Bils, Mark, and Peter J. Klenow. 2000. "Does Schooling Cause Growth?" *The American Economic Review*, 90(5): 1160–1183.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *The Quarterly Journal of Economics*, 1351–1408.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "The Organization of Firms Across Countries." *The Quarterly Journal of Economics*, 127(4): 1663–1705.

Braguinsky, S., L. G Branstetter, and A. Regateiro. 2011. "The Incredible Shrinking Portuguese Firm." *National Bureau of Economic Research*, Working Paper 17265.

Cabral, Luis MB, and Jose Mata. 2003. "On the Evolution of the Firm Size Distribution: Facts and Theory." *American economic review*, 1075–1090.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. 2015. "The Anatomy of French Production Hierarchies." *Journal of Political Economy*, 123(4): 809–852.

Caliendo, Lorenzo, Giordano Mion, Luca David Opromolla, and Esteban Rossi-Hansberg. 2020. "Productivity and Organization in Portuguese Firms." *Journal of Political Economy*, 128(11): 4211–4257.

Card, David. 1999. "The Causal Effect of Education on Earnings." *Handbook of labor economics*, 3: 1801–1863.

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica*, 69(5): 1127–1160.

Caselli, Francesco. 2005. "Accounting for Cross-Country Income Differences." *Handbook of economic growth*, 1: 679–741.

Caselli, Francesco, and Antonio Ciccone. 2013. "The Contribution of Schooling in Development Accounting: Results from a Nonparametric Upper Bound." *Journal of Development Economics*, 104: 199–211.

Caselli, Francesco, and Nicola Gennaioli. 2013. "Dynastic Management." *Economic Inquiry*, 51(1): 971–996.

Chamberlain, Gary. 1984. "Panel Data." *Handbook of econometrics*, 2: 1247–1318.

Ciccone, Antonio, and Elias Papaioannou. 2009. "Human Capital, the Structure of Production, and Growth." *The Review of Economics and Statistics*, 91(1): 66–82.

Ciccone, Antonio, and Giovanni Peri. 2006. "Identifying Human-Capital Externalities: Theory with Applications." *The Review of Economic Studies*, 73(2): 381–412.

Cole, Harold L., Jeremy Greenwood, and Juan M. Sanchez. 2016. "Why Doesn't Technology Flow From Rich to Poor Countries?" *Econometrica*, 84(4): 1477–1521.

Cooper, Arnold C., F. Javier Gimeno-Gascon, and Carolyn Y. Woo. 1994. "Initial Human and Financial Capital as Predictors of New Venture Performance." *Journal of business venturing*, 9(5): 371–395.

Córdoba, Juan Carlos, and Marla Ripoll. 2008. "Endogenous TFP and Cross-Country Income Differences." *Journal of Monetary Economics*, 55(6): 1158–1170.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." *Journal of Economic Perspectives*, 28(3): 3–24.

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

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2016. "The Slow Growth of New Plants: Learning about Demand?" *Economica*, 83(329): 91–129.

Gabaix, X. 1999. "Zipf's Law for Cities: An Explanation." *The Quarterly Journal of Economics*, 114(3): 739–767.

Gabaix, Xavier. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics*, 1(1): 255–294.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The Dynamics of Inequality." *Econometrica*, 84(6): 2071–2111.

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

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

Gopinath, Gita, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. 2017. "Capital Allocation and Productivity in South Europe*." *The Quarterly Journal of Economics*, 132(4): 1915–1967.

Hall, Robert E., and Charles 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, Lutz, and Todd Schoellman. 2018. "Human Capital and Development Accounting: New Evidence from Wage Gains at Migration." *The Quarterly Journal of Economics*, 133(2): 665–700.

Heston, Alan, Robert Summers, and Bettina Aten. 2002. "Penn World Table 6.1." *Center for International Comparisons at the University of Pennsylvania.*

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5): 1127–1150.

Hopenhayn, Hugo A. 2014. "Firms, Misallocation, and Aggregate Productivity: A Review." *Annual Review of Economics*, 6(1): 735–770.

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

Hsieh, Chang-Tai, and Peter J. Klenow. 2010. "Development Accounting." *American Economic Journal: Macroeconomics*, 2(1): 207–223.

Hsieh, Chang-Tai, and Peter J. Klenow. 2014. "The Life Cycle of Plants in India and Mexico." *The Quarterly Journal of Economics*, 129(3): 1035–1084.

Hurst, Erik, and Benjamin Wild Pugsley. 2011. "What Do Small Businesses Do?" *Brookings Papers on Economic Activity*, 2011(2): 73–118.

Iranzo, Susana, and Giovanni Peri. 2009. "Schooling Externalities, Technology, and Productivity: Theory and Evidence from US States." *The Review of Economics and Statistics*, 91(2): 420–431.

Jann, Ben. 2020. "ISCOGEN: Stata Module to Translate ISCO Codes." *Boston College Department of Economics.*

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

Jones, Charles I., and Jihee Kim. 2018. "A Schumpeterian Model of Top Income Inequality." *Journal of Political Economy*, 126(5): 1785–1826.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica*, 50(3): 649–670.

Kangasharju, Aki, and Sari Pekkala. 2002. "The Role of Education in Self–Employment Success in Finland." *Growth and change*, 33(2): 216–237.

Klenow, Peter, and Andres Rodriguez-Clare. 1997. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" In *NBER Macroeconomics Annual 1997, Volume 12.* 73–114. MIT Press.

Klenow, Peter J., and Andrés Rodríguez-Clare. 2005. "Chapter 11 Externalities and Growth." In *Handbook of Economic Growth.* Vol. 1, 817–861. Elsevier.

Knight, Frank Hyneman. 1921. *Risk, Uncertainty and Profit.* Vol. 31, Houghton Mifflin.

Krueger, Alan B., and Mikael Lindahl. 2001. "Education for Growth: Why and for Whom?" *Journal of Economic Literature*, 39(4): 1101–1136.

Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman. 2018. "Life Cycle Wage Growth across Countries." *Journal of Political Economy*, 126(2): 797–849.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 1999. "Corporate Ownership around the World." *The Journal of Finance*, 54(2): 471–517.

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

Lemieux, Thomas. 2006*a*. "The "Mincer Equation" Thirty Years After Schooling, Experience, and Earnings." In *Jacob Mincer A Pioneer of Modern Labor Economics.* , ed. Shoshana Grossbard, 127–145. Boston, MA:Springer US.

Lemieux, Thomas. 2006*b*. "Postsecondary Education and Increasing Wage Inequality." *The American Economic Review*, 96(2): 195–199.

Levine, Ross, and Yona Rubinstein. 2016. "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?" *The Quarterly Journal of Economics*, qjw044.

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

Luttmer, Erzo GJ. 2007. "Selection, Growth, and the Size Distribution of Firms." *The Quarterly Journal of Economics*, 122(3): 1103–1144.

Luttmer, Erzo G. J. 2011. "On the Mechanics of Firm Growth." *The Review of Economic Studies*, 78(3): 1042–1068.

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

Manuelli, Rodolfo E., and Ananth Seshadri. 2014. "Human Capital and the Wealth of Nations." *American Economic Review*, 104(9): 2736–2762.

Mata, Jose. 1996. "Markets, Entrepreneurs and the Size of New Firms." *Economics Letters*, 52(1): 89–94.

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

Michelacci, Claudio, and Fabiano Schivardi. 2020. "Are They All like Bill, Mark, and Steve? The Education Premium for Entrepreneurs." *Labour Economics*, 67: 101933.

Moretti, Enrico. 2004*a*. "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics*, 121(1): 175–212.

Moretti, Enrico. 2004*b*. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *The American Economic Review*, 94(3): 656–690.

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

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

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

Parker, Simon C. 2004. *The Economics of Self-Employment and Entrepreneurship.* Cambridge University Press.

Parker, Simon C., and C. Mirjam Van Praag. 2006. "Schooling, Capital Constraints, and Entrepreneurial Performance: The Endogenous Triangle." *Journal of Business & Economic Statistics*, 24(4): 416–431.

Rauch, James. 1993. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics*, 34(3): 380–400.

Reis, Ricardo. 2013. "The Portuguese Slump and Crash and the Euro Crisis." *Brookings Papers on Economic Activity*, 2013(1): 143–210.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments." *Review of Economic dynamics*, 11(4): 707–720.

Say, Jean Baptiste. 1836. *A Treatise on Political Economy.* Grigg & Elliot.

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

Schultz, Theodore W. 1975. "The Value of the Ability to Deal with Disequilibria." *Journal of economic literature*, 827–846.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the Twenty-First Century*." *The Quarterly Journal of Economics*, 134(4): 1675–1745.

Solow, Robert M. 1960. "Investment and Technical Progress." *Mathematical methods in the social sciences*, 1: 48–93.

Stehrer, Robert, Alexandra Bykova, Kirsten Jäger, Oliver Reiter, and Monika Schwarzhappel. 2019. "Industry Level Growth and Productivity Data with Special Focus on Intangible Assets." *Vienna Institute for International Economic Studies Statistical Report*, , (8).

Van Der Sluis, Justin, Mirjam Van Praag, and Wim Vijverberg. 2008. "Education and Entrepreneurship Selection and Performance: A Review of the Empirical Literature." *Journal of Economic Surveys*, 22(5): 795–841.

Van Praag, Mirjam, Arjen van Witteloostuijn, and Justin van der Sluis. 2013. "The Higher Returns to Formal Education for Entrepreneurs versus Employees." *Small Business Economics*, 40(2): 375–396.

Welch, Finis. 1970. "Education in Production." *The Journal of Political Economy*, 35–59.