Sorting into Entrepreneurial Teams *

Edoardo Maria Acabbi[1], Andrea Alati[2], Luca Mazzone[3], and Marta Morazzoni[4]

[1] *Universidad Carlos III de Madrid*
[2] *Bank of England*
[3] *CERGE-EI*
[4] *University College London, IFS and CEPR*

February 2025

[Preliminary. Click here for the latest version.]

Abstract

This paper studies how entrepreneurs sort into teams and how team entrepreneurship affects the equilibrium distribution of firms. Leveraging employer-employee administrative records matched with privately-held firms' balance sheet data for Portugal, we show that firms of entrepreneurial teams have higher sales, productivity and survival rates than those owned by single entrepreneurs. We then exploit information on individuals' careers before opening a firm to establish that there is a strong degree of sorting in entrepreneurial teams along observed and unobserved heterogeneity. A novel theory of career choices and team formation rationalizes why similarity in entrepreneurs' overall talent and dissimilarity in their specialization lead to better firm outcomes, providing insights into the micro-foundations of firm growth.

*We thank Gadi Barlevy, Francois Gourio, Sasha Indarte, Emi Nakamura, Morten Ravn, Tom Schmitz, Jon Steinsson, Vincent Sterk, Gianluca Violante, and participants in seminars at UCL, Queen Mary University London, CREI, Goethe University Frankfurt, the Inter-American Development Bank, Northwestern University, Bank of England, the Chicago FED, the LSE Junior Macro Conference, and the NYU Abu Dhabi Conference on New Frontiers for Policy Evaluation for helpful feedback. Elena Casanovas provided excellent research assistance. The views expressed reflect only the view of the authors and not those of the Bank of England or its policy committees.

Acabbi: eacabbi@emp.uc3m.es
Alati: andrea.alati@bankofengland.co.uk
Mazzone: luca.mazzone@cerge-ei.cz
Morazzoni: m.morazzoni@ucl.ac.uk

1 Introduction

Firm productivity stands as a fundamental driver of economic growth and forms the cornerstone of macroeconomic models of firm dynamics.Understanding what determines the dispersion in firm performance means understanding how micro-level frictions shape macro-level outcomes, such as the production capacity of a given country (see, among others, Syverson 2011 and Bartelsman, Haltiwanger and Scarpetta 2013). Along these lines, research has established that firms' heterogeneity, especially at birth, is highly indicative of their life-cycle trajectories (Sterk, Sedláček and Pugsley 2021) – including differences in firm selection at entry (Bhandari et al. 2022), their initial pool of workers (Choi et al. 2023) and early investments into physical capital (De Haas, Sterk and Van Horen 2022). This paper takes a different angle, and focuses on how the sorting of entrepreneurs into teams affects the distribution of firms in equilibrium.

While between 30 and 40% of privately held firms in advanced economies is multi-owned,[1] limited attention has been devoted to the sorting patterns of entrepreneurial teams, although the literature suggests that entrepreneurs in teams have similar industry experiences as workers (Feliz, Karmakar and Sedlácek 2021). Our aim is to investigate more broadly how entrepreneurs select co-founders and whether they have similar (or dissimilar) talent and skills, and then offer an explanation as to why these sorting patterns influence firm outcomes. Combining novel theoretical framework and empirical evidence, we show that entrepreneurial sorting in talent and skills is key for capturing the determinants of heterogeneity in productivity, especially at firm inception.

We first build a model that encompasses the core dynamics of team sorting within a framework of entrepreneurship and career choice. In our model economy, individuals are endowed with a combination of skills across multiple dimensions. Equilibrium wages in each occupation provide a price for each single skill. If choosing to become entrepreneurs, individuals use a combination of their entire skills set, and then demand labor depending on relative wages. Crucially, entrepreneurial teams' productivity emerges as a combination of each skill at the individual level. We assume that, before career and entrepreneurial choices are made, each individual has a chance to meet another one, randomly sampled from the same joint skills distribution. Teams are formed when both individuals in the match prefer starting a firm in a team rather than alone, or supplying labor as a worker.

The model yields four main intuitions: (i) conditional on their overall productivity, individuals whose skill endowment is more dispersed are more likely to become members of entrepreneurial teams; (ii) while individuals with a higher productivity draw are more likely to be entrepreneurs, the productivity of members of entrepreneurial teams is on

[1] In our dataset, for example, entrepreneurial teams constitute approximately 30% of all privately-held firms, employ more than 40% of their workforce, and account for more than 40% of their total gross sales.

average higher than that of single entrepreneurs; (iii) firms with heterogeneity in the skill composition of team members are larger and more productive; (iv) firms with positive sorting in the overall productivity of team members are larger and more productive.

Then, we analyze empirically the formation and performance of entrepreneurial teams, leveraging comprehensive administrative data that covers the universe of employer-employee matches in Portugal from 1985 to 2019, which can also be linked to the balance sheets of privately-held firms. This dataset is unique in two important dimensions: first, on top of several demographic variables, it records the entire occupational trajectory of millions of individuals, including their transitions in and out of self-employment. Second, it makes it possible to investigate the performance of firms started by the entrepreneurs in our sample, including those firms founded by a team.

To test the predictions of our theoretical framework, we need measures of individuals' talent and skills, which we recover by tracing the career paths of workers and entrepreneurs alike in our sample. First, we exploit information on the occupations agents hold before starting a firm, together with official EU-wide crosswalks between occupations and essential competences, to provide an individual-level measure of skills. Second, we use yearly wages to estimate agents' unobserved heterogeneity – a measure of their talent – following the methods in Abowd, Kramarz and Margolis (1999) and Bonhomme, Lamadon and Manresa (2019). Crucially, for individuals who become entrepreneurs, we focus on their working careers before the first entrepreneurial spell.

By analyzing entrepreneurs' employment histories, we find evidence in support of the prevalent view in entrepreneurship, dating back to Lucas (1978) and confirmed in Levine and Rubinstein (2017), suggesting entrepreneurs are positively selected with respect to workers on overall productivity. However, we also establish a novel empirical fact, which is consistent with the second prediction of our framework: entrepreneurs in teams are positively selected with respect to single entrepreneurs on overall productivity. In our model, for someone to prefer team entrepreneurship over their outside options, they must have relatively high levels of at least one skill and to have met a potential business partner with relatively high levels of at least one complementary skill. Along this line – and again consistent with the first prediction of our model –, we find that individuals with unbalanced skills in our sample are more likely to be part of entrepreneurial teams.

It is important to stress that the key novelty of our empirical strategy is to disentangle different sorting dimensions within entrepreneurial teams, distinguishing between similarity in talent (e.g., cognitive and managerial ability) and complementarity in skills (e.g., technical vs. business expertise). In turn, the analysis of entrepreneurs' employment histories before teaming up – linked to firm-level outcomes – lends further empirical support for the third and fourth predictions of the model. Specifically, our findings indicate that positive sorting along latent types correlates with higher firm sales, total factor productivity (TFP), and survival rates. However, we also document that teams exhibiting greater skill diversity tend to achieve better firm-level outcomes, suggesting that complementarity in competences enhances firm success.

Overall, by integrating empirical evidence with a model of career choice and team formation, we make three contributions to existing works on the aggregate consequences of entrepreneurial dynamics. First, we add to the research on entrepreneurial selection by highlighting the importance of co-founder choice in determining firm-level trajectories. Second, we provide novel insights into knowledge spillovers within teams, analyzing entrepreneurial sorting patterns along observed and unobserved traits. Finally, we build on studies examining the drivers of firm success by linking entrepreneurial team composition to long-run firm outcomes, both in the data and in a model. The tension between positive sorting and skill heterogeneity provides novel insights into the micro-foundations of firm growth. We also find limited evidence that financial or cyclical factors primarily drive team formation, reinforcing the argument that, unlike workers sorting early into startups (Bias and Ljungqvist 2023), entrepreneurial sorting is linked to intrinsic attributes rather than external constraints.

Related Literature. Our work relates to four strands of research. First, we contribute to the existing body of evidence on the labor market determinants of entrepreneurship. Our empirical analysis is related to Gendron-Carrier (2024), Humphries (2022), and Queiró (2022), who explore the human capital accumulation and career patterns leading to entrepreneurship. Since we focus on the heterogeneous components of individual skills, we also relate to Argan, Indraccolo and Piosk (2024), who use Danish data to show that workers with very specialized skills are less likely to become entrepreneurs.

Second, we contribute to the literature that studies skill complementarities and how they lead to the sorting of individuals in labor markets. In this sense, our focus on two-people teams that are core to their organization is close to Freund (2022), although we analyze teams of entrepreneurs, not workers. Moreover, our theory of mutual learning between entrepreneurs is an extension of Acabbi, Alati and Mazzone (2024), and is also related to Jarosch, Oberfield and Rossi-Hansberg (2021) and Herkenhoff et al. (2024).

Third, we contribute to an emerging literature on entrepreneurial teams. Notably, Choi et al. (2021) shows that the death of a member of a firm's founding team negatively affects its performance, while D'Acunto, Tate and Yang (2024) discuss the role of skill complementarities between entrepreneurs with similar previous industry experiences. We build and expand on these results by explicitly investigating the determinants of team formation along both the vertical and horizontal differentiation of individual skills, and studying – empirically and theoretically – their role for the life-cycle performance of firms.

Finally, our model of career choice builds on the classic *span of control* framework proposed by Lucas (1978) by adding multiple individual skills and entrepreneurial team formation.[2] Importantly, our focus on the direction of sorting across dimensions of individuals' characteristics follows intuitions discussed in Eeckhout and Kircher (2011). A different approach to complementarities in production is discussed in Boerma, Tsyvinski and Zimin (2025) focusing on teams of size three – two workers and a project. In related work, Boerma et al. (2023) explores the coexistence of positive and negative assortative matching when matches are characterized by concave mismatch costs, while Mukoyama and Sahin (2005) discuss the emergence of negative sorting between skills for pairs of workers. Our paper illustrates how, when looking at teams of entrepreneurs, the notions of positive and negative sorting can coexist, depending on whether one looks at skills separately or if all of them are taken into account into a measure of overall talent.

The remainder of this paper is structured as follows. Section 2 develops a theoretical general equilibrium model of career decisions and team entrepreneurship, whose key predictions will be then tested empirically. Section 3 outlines the data sources and provides descriptive statistics for our sample. Section 4 presents the empirical strategy and discusses our results, including robustness checks and alternative explanations. Finally, Section 5 concludes with policy implications and directions for future research.

2 Model

In what follows, we propose a static general equilibrium model of career decisions, characterized by heterogeneous (workers') occupations, entrepreneurship, and entrepreneurial team formation. The economy we analyze is populated by a continuum of risk-neutral agents, each endowed with one unit of time and a heterogeneous combination of skills among N existing ones, which we think of as their human capital.

Agents in this economy have access to three career choices: (i) working as an employee, (ii) founding a single-owned firm, or (iii) forming a multi-owner entrepreneurial team. Entrepreneurs derive profits from firm operations, while workers earn the wages that clear the labor market. Individuals that choose the first option can supply labor in any of the N different occupations, which uses intensively one of the N skills. The formation of entrepreneurial teams follows instead a matching process, where agents search for and pair with potential co-founders if the expected joint profits exceed individual alternatives.

Wages and profits are determined in equilibrium, and so are occupation choices, which depend on the (endogenous) returns to starting a firm relative to working as an employee. The structure of the economy, including the wage-setting mechanism and entrepreneurial teams matching process, leads to equilibrium outcomes in which agents self-select into

[2] Other notable extensions of the entrepreneurial model by Lucas (1978) that already consider multiple dimension of individual ability or skills include Cagetti and De Nardi (2006) and Poschke (2013).

different careers based on their comparative advantage and expected earnings.

The output of firms depends on the human capital of their founders, labor inputs from the external market, and the technology that governs production. Firms demand labor from each occupation depending on their own specialization and on prevailing wages.

2.1 Agent Heterogeneity and Occupational Choices

Each agent is characterized by a N-dimensional skill vector defined by the following:

$$\vec{\theta} = [\theta_1, \theta_2, \dots, \theta_N]$$

Given the combination of their skill levels, individuals will be both *vertically* and *horizontally* differentiated. To use a terminology close to Freund (2022), vertical differentiation can be interpreted more generally as the individual's talent, while horizontal differentiation speaks to their specialization. Formally, talent can be obtained as $\tau_i(\vec{\theta}) = \sqrt{\sum_{j=1}^{N} \theta_j^2}$. Note that the joint dis tribution $G(\vec{\theta})$ captures the heterogeneity in individual skill endowments as a multivariate probability distribution over $\mathbb{R}^N$. Formally, we can write:

$$\vec{\theta} = [\theta_1, \theta_2, \dots, \theta_N] \sim G(\theta_1, \theta_2, \dots, \theta_N)$$

When individuals supply labor to firms, their productivity as workers is occupation-specific and is given by $h_j(\vec{\theta})$ for each occupation $j \in \{1, \dots, N\}$. Importantly, we assume that skills are fully unbundled across occupations, so that $h_j(\vec{\theta}) = h_j(\theta_j)$.[3] Also, the function $h_j(\cdot)$ is increasing and concave. An individual i with skills given by the vector $\vec{\theta^i}$ and who is employed in occupation j will then have earnings given by $w_j \cdot h_j(\theta_j^i)$.

Alternatively, individuals can become entrepreneurs. If they do, their productivity as entrepreneurs is given by $z(\vec{\theta})$, with the function $z(\cdot)$ increasing in all elements of $\vec{\theta}$ and non-negative cross derivatives.[4] Since we abstract from the use of capital in production, the profit maximization problem of a single-owned firm run by individual i is given by:

$$\pi_i\left(\vec{\theta_i}, \vec{L}, \vec{w}\right) = z\left(\vec{\theta_i}\right) \cdot f(\vec{L}) - \sum_{j=1}^{N} w_j L_j \tag{1}$$

where $f(\cdot)$ is increasing and concave. Productivity $z(\vec{\theta_i})$ of individual i thus enters the firm problem as in the general class of models of entrepreneurial *span of control* (Lucas

[3]See Edmond and Mongey (2021) on the assumption that skills can be priced separately.

[4]An important assumption we make is that worker skills can be transferred from individuals as workers to entrepreneurship – Gyetvai and Tan (2023) provide evidence consistent with our model choice.

1978). [5] In addition to starting firms alone, we allow for agents to team up in pairs and run firms jointly. When they do, they face a maximization problem analogous to **Equation 1**. In entrepreneurial teams, however, overall firm productivity is a function of the skill vectors across each of the two members of the team, so that team productivity reads as:

$$z(\vec{\theta}_T) = z(\psi(\vec{\theta}_i, \vec{\theta}_{i'})) \tag{2}$$

for every team where individual i is matched with another individual of generic type i'.

2.2 Equilibrium Characterization

Payoffs. Each firm maximizes profits by choosing labor inputs, so $\vec{L}_i^*$ is firm i's labor demand, and the payoff from starting a single-owned firm is $\pi_i\left(\vec{\theta}_i, \vec{L}^*(\vec{\theta}_i), \vec{w}\right)$. For simplicity of notation, we will indicate the payoff for opening a single-owned firm as $\pi^*_{I,i}\left(\vec{\theta}_i\right)$. Similarly, the payoff of becoming part of an entrepreneurial team for individual i is $\pi^*_{T,i}\left(\psi(\vec{\theta}_i, \vec{\theta}_{i'})\right)$ - and it will depend on the other team member, whose type is i'. As discussed above, the payoff of working in occupation j is $w_j \cdot h_j(\theta^i_j)$.

Matching. Before making career choices, each individual randomly meets another, drawn from the same distribution, $G(\vec{\theta})$. This implies that there will be a non-zero measure of entrepreneurial teams for some sections of the skills space $\vec{\theta}$, where suitable matches can be formed, and a measure zero of teams when $\vec{\theta}_i$ is such that no entrepreneurial team can be formed, whatever type i' is i meeting. Defining choice sets below will clarify this.

Choice Sets. The set of any two individuals i and i' that choose to be workers and supply labor in occupation j is a subset of the product space $\Theta_i \times \Theta_{i'}$, and is given by:

$$\mathcal{W}_j = \left\{ \left\{ \vec{\theta}_i, \vec{\theta}_{i'} \right\} \;\middle|\; w_j\, h(\theta_j) \;\geq\; \max\left\{ w_k h(\theta_k), \pi^*_{I,i}\left(\vec{\theta}_i\right), \pi^*_{T,i}\left(\psi(\vec{\theta}_i, \vec{\theta}_{i'})\right)\right\} \quad \forall k \neq j \right\}$$

for any individual i matched with an individual i'. Clearly, this choice set is defined for *pairs* of skill vectors, reflecting the complementarities arising from different potential meetings. We can similarly define the set of individual and team entrepreneurs, given by:

$$\mathcal{E}_I = \left\{ \left\{ \vec{\theta}_i, \vec{\theta}_{i'} \right\} \;\middle|\; \pi^*_{I,i}\left(\vec{\theta}_i\right) \geq \max\left\{ w_j h(\theta_j), \pi^*_{T,i}\left(\psi(\vec{\theta}_i, \vec{\theta}_{i'})\right)\right\}; \forall j \in 1, \ldots, N \right\}$$
$$\mathcal{E}_T = \left\{ \left\{ \vec{\theta}_i, \vec{\theta}_{i'} \right\} \;\middle|\; \pi^*_{T,i}\left(\psi(\vec{\theta}_i, \vec{\theta}_{i'})\right) \geq \max\left\{ w_j h(\theta_{\iota,j}), \pi^*_{I,\iota}\left(\vec{\theta}_\iota\right)\right\}; \forall j \in 1, \ldots, N;\ \iota \in \{i, i'\} \right\}$$

Notice that the set of team entrepreneurs requires a *double coincidence* or, differently said, *bilateral agreement*, meaning that it includes only those pairs where both individuals

[5] Notice that, for an appropriate specification of $h(\cdot)$ that collapses worker heterogeneity into a single type, and for $N = 1$, our framework boils down to the original span of control model by Lucas 1978.

prefer forming an entrepreneurial team to all other outside career options. With the choice sets at hand, we can thus define the aggregate labor demand for each occupation j as:

$$\mathrm{LD}_j \;=\; \int\int_{\mathcal{E}_I} L_j^*\left(\vec{\theta}_i\right) dG\left(\vec{\theta}_i\right) dG\left(\vec{\theta}_{i'}\right) + \int_{\mathcal{E}_T}\int_{\mathcal{E}_T} L_j^*\left(\vec{\theta}_i, \vec{\theta}_{i'}\right) dG\left(\vec{\theta}_i\right) dG\left(\vec{\theta}_{i'}\right)$$

Integrating across agents choosing to work, the labor supply for occupation j becomes:

$$\mathrm{LS}_j \;=\; \int\int_{\mathcal{W}_j} h_j(\theta_j) dG\left(\vec{\theta}_i\right) dG\left(\vec{\theta}_{i'}\right)$$

Finally, note that equilibrium wages are given by the vector $\vec{w} = [w_1, \ldots, w_N]$ and, together with occupation and entrepreneurial choices, they clear the N labor markets.

Timing and Information. While the model is static in nature, agents' key decisions are taken sequentially. Specifically, before any other choice is made, every agent randomly meets another one whose type is drawn from the same joint distribution of skills. As match are formed, all decisions are taken simultaneously: individuals decide whether to open firms together or alone, each firm demands its optimal amount of labor from each occupation, and workers choose occupations based on their skills and prevailing wages.

2.3 Calibration Strategy

Workers. The function $h(\cdot)$ that translates skills into labor productivity is given by $h_j(\theta_j) = \kappa \cdot \theta_j^{\phi}$ and $h_j(\theta_k) = 0$ for all $k \neq j$, with $\phi \in (0, 1)$. Notice that, as $\phi \longrightarrow 0$, workers' heterogeneity disappears, as all individuals supply exactly one unit of labor when working as employees, although their entrepreneurial productivity remains heterogeneous.

Entrepreneurs. Individual skills are aggregated into entrepreneural productivity by the function $z\left(\vec{\theta}\right)$, which is CES with share parameters δ^E and substitution parameter σ^E. A team's skill vector will be:

$$\vec{\theta}_T(i, i') = \left[\psi(\theta_1^i, \theta_1^{i'}), \quad \psi(\theta_2^i, \theta_2^{i'}), \quad \ldots, \quad \psi(\theta_N^i, \theta_N^{i'})\right]^T$$

To calibrate the aggregation of skills within entrepreneurial teams, we adopt a functional form similar in spirit to the "catch-up" technology in Acabbi, Alati and Mazzone (2024). The underlying idea is that each entrepreneur may eventually accumulate the skills in which the other team member is relatively more abundant. However, since our model is static, we might want to collapse the skill accumulation dynamics to a single shift. While assuming that the catch-up immediately happens would be akin to assume that

$\psi(\theta_j^i, \theta_j^{i'}) = \max(\theta_j^i, \theta_j^{i'}) \quad \forall j = 1, \dots, N$, a slightly more flexible form is:

$$\psi(\theta_j^i, \theta_j^{i'}) = \frac{\log\left(\xi \exp^{a \cdot \theta_j^i} + \xi \exp^{a \cdot \theta_j^{i'}}\right)}{a} \quad \forall j = 1, \dots, N$$

This is a logsum expression, with a penalty term ξ that shifts the expression downward. As $a \to \infty$, the penalty term becomes irrelevant, and $\psi(\theta_j^i, \theta_j^{i'}) \to \max(\theta_j^i, \theta_j^{i'})$.

Firms. We assume the production function of firms to be $z\left(\vec{\theta}\right) f\left(\vec{\theta}\right) \equiv z\left(\vec{\theta}\right) \tilde{L}^{\alpha}$. Labor inputs at the firm level are aggregated using a CES form, with share parameters δ^L and substitution parameter σ^L. We further assume decreasing returns to scale in production – $\alpha < 1$ – to allow for the existence of a non-degenerate distribution of firms in equilibrium.

Distributions. We calibrate the joint distribution of skills $G(\vec{\theta})$ by combining N independent marginal Beta distributions with parameters $\left\{\alpha_\beta^j, \beta_\beta^j\right\}$, for $j = 1, \dots, N$ by means of a Gaussian copula. The copula allows us to model skill correlation separately from their marginal distributions, and to capture it with parameter(s) $\rho_{j,j'}$ for $j \neq j'$.

2.4 Numerical Example with 2 Skills

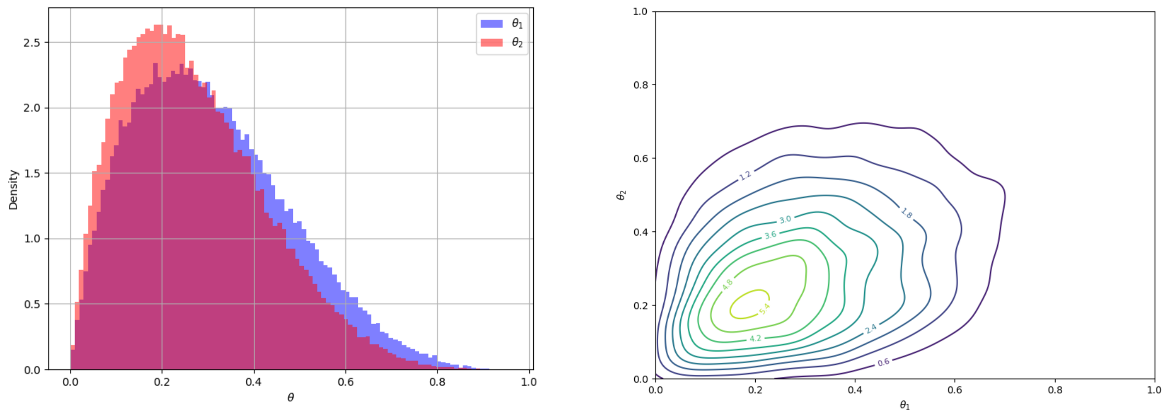
For ease of exposition, we now present a numerical solution of our model in which individuals have only two skills ($N = 2$). With two skills, there are also two occupations, and we can represent individual choices graphically. Note that the current model calibration is to be intended as a qualitative example and not a quantitative exercise.

Table 1: Calibration Parameters

Parameter	Description	Value
α	Production Function Returns to Scale	0.7
$(\alpha_\beta, \beta_\beta)$	Shape of Beta Distribution of Skills	(2.0,4.5)
κ	Scaling Parameter in Labor Productivity	1.1
ρ	Correlation (Gaussian copula) Between Skill Distributions	0.4
ϕ	Worker Type Transformation	0.4
δ^L	CES Share Parameter in Firm Production	0.5
σ^L	CES Substitution Parameter in Firm Production	0.0
δ^E	CES Share Parameter in Entrep. Team Productivity	0.5
σ^E	CES Substitution Parameter in Entrep. Team Productivity	0.3
a	Logsum Parameter	8.0
ξ	Logsum Penalty	0.5

We work under a scenario of perfect symmetry across skills. To make the example more compelling, we make the distribution of θ_1 less skewed, multiplying the shape parameter β_β of θ_1 by 0.925 (see **Figure 1**). **Table 1** summarizes the values of all remaining parameters.

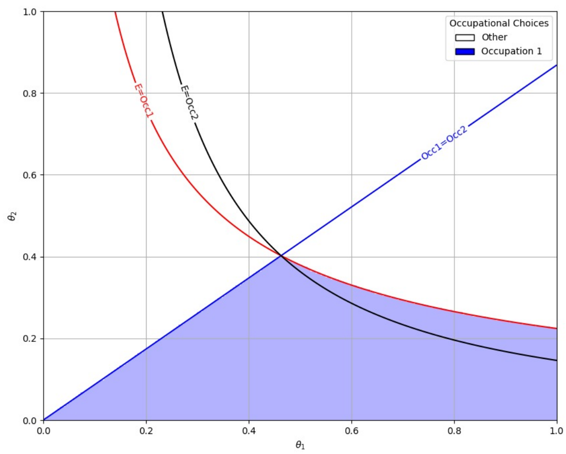
Figure 1: Skills Distribution in 2-Dimensional Model



(a) Marginal Distributions **(b)** Joint Distribution

Partial Matching. Before we discuss the general equilibrium properties of the model, some features of this economy can be discussed by describing the choices of a particular individual. Specifically, we consider the choice set of an agent i, who – at the beginning of the model period – draws a potential match with another agent i' , characterized by the skill bundle $\{\theta_1 = 0.5; \theta_2 = 0.2\}$. Clearly, matched agent i' is a "specialist" in skill 1, but for which values does agent i consider themselves a specialist in skill 1? The area in the $\{\theta_1, \theta_2\}$ space for which individual i chooses occupation 1 is portrayed in **Figure 2**.

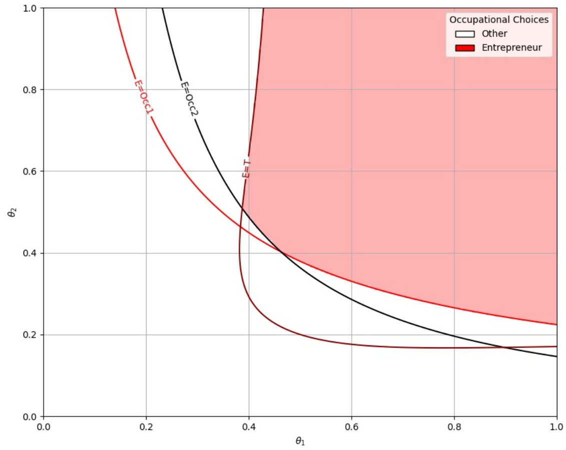
Figure 2: Share of Occupation 1 given match with $[0.5, 0.2]$



The dark blue line represents the locus of points (i.e. combinations of skills in individuals' bundles) for which agent i is indifferent between occupations 1 and 2. Note that, since skill 2 is relatively scarcer, it commands a higher wage. For individuals with high levels of θ_1, however, another relevant boundary is given by the red line, which is the indifference locus between occupation 1 and becoming a single entrepreneur.

The area of single entrepreneurs, in turn, is only partially determined by the

Figure 3: Share of Single Entrepreneurs given match with $[0.5, 0.2]$



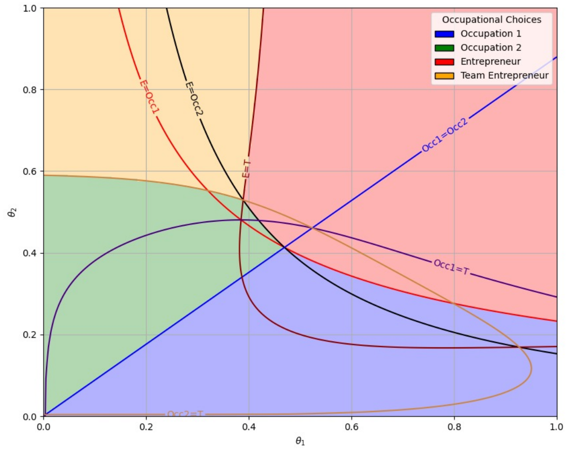
indifference line between entrepreneurship and occupation 1. Indeed, the black line shows the locus of points where individuals have a skill mix that makes them indifferent between becoming entrepreneurs and supplying labor in occupation 2; the maroon line, finally, applies to high-θ_2 types whose θ_1 levels are not high enough to start a firm alone.

The red shadowed area, therefore, covers the values of the skill bundle $\vec{\theta}$ for which single entrepreneurship is preferred to supplying labor in either occupation and to joining an entrepreneurial team, encapsulating the intuition of Lazear (2004) that many entrepreneurs have "*balanced*" skills, or in other words that they are *generalists*.

Putting these elements together, we see for which values agent i will pick occupation 2, or will want to team up with individual i'. **Figure 4** illustrates this. The clear motive for team formation is consistent with D'Acunto, Tate and Yang (2024), who highlight the role of skill complementarities as drivers of entrepreneurial partnerships. However, one additional feature is highlighted by the figure, which is novel in our model: *vertical differentiation*. Individuals in the green area end up going towards occupation 2, because, like those who form teams, their skills are specialized towards θ_2. However, those who become team entrepreneurs have higher overall talent. We therefore see the role of vertical and horizontal differentiation in shaping agents' sorting into different career choices.

The boundaries of the choice regions depend heavily on the skill endowment of the potential team partner drawn by any given individual. Indeed, **Figure A.1** displays the same choice sets but for a different case, in which the potential match has both higher talent and lower specialization. In that scenario, individuals which are specialized both in skill 1 and in skill 2 end up forming a team. However, the set of individuals with a skill 2 specialization that joins a team shrinks, because the matched individual would rather start a firm alone than with an individual that is too close in skills composition, or too distant in talent. These intuitions will shape model predictions also in the general case.

Figure 4: All Choices, given match with $[0.5, 0.2]$



Full Matching. We now consider the case in which each individual gets to meet another one, drawn from the same ex-ante skill distribution $G(\vec{\theta})$, and where career choices, entrepreneurship, and occupational wages are simultaneously and jointly determined in equilibrium. Numerically, the skill distribution is the same as in the previous example, with the marginals and the joint distribution described in **Figure 1**.

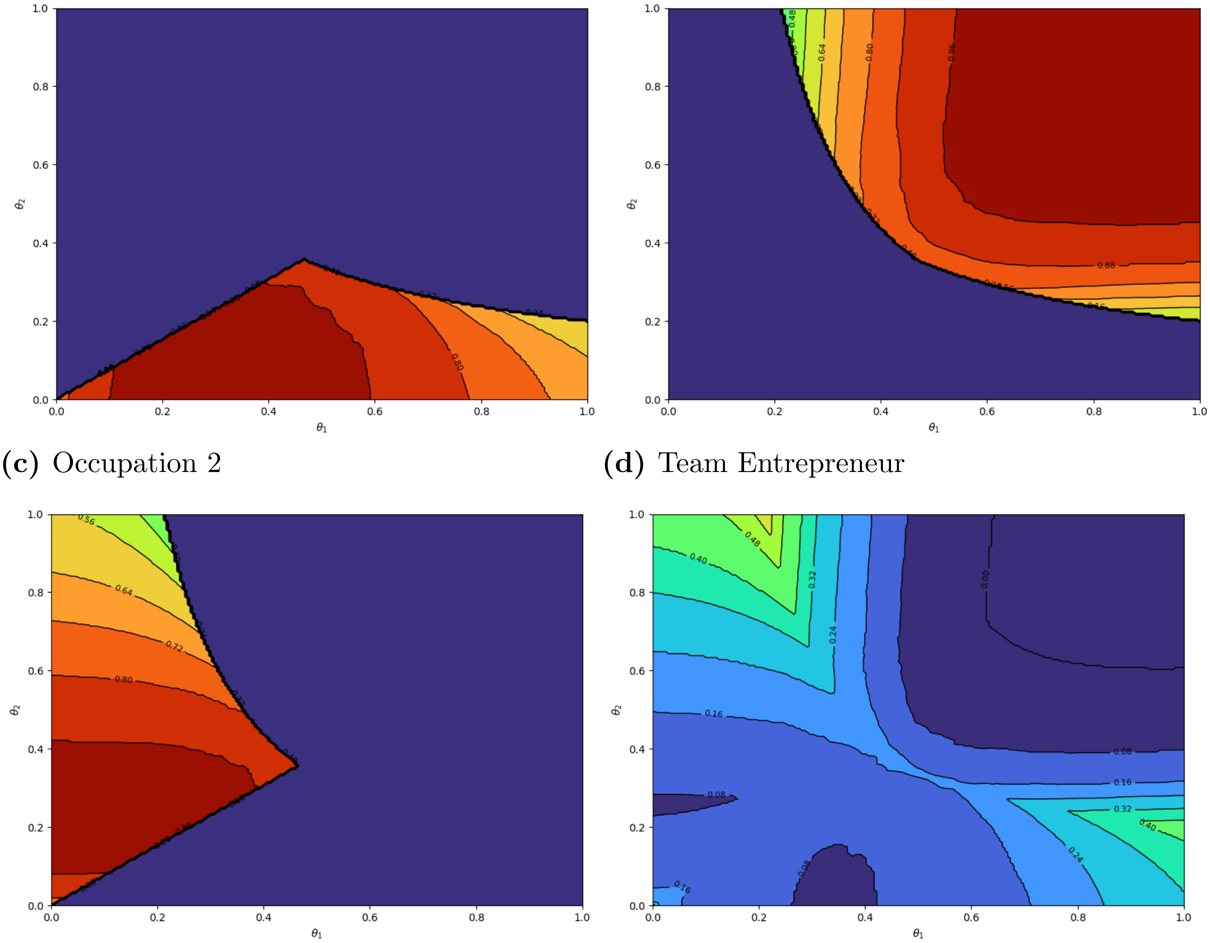
As in the partial matching case, the relative scarcity of skill 2 implies a higher price for that skill - in the general equilibrium exercise, w_2 is 3.25% higher than w_1. About 16% of agents become entrepreneurs, with 41% going into occupation 1, and 43% going into occupation 2. Interestingly, about half of the firms are run by entrepreneurial teams.

Figure 5 displays individual choices over the θ_1, θ_2 space: the color is dark purple when no agent in that point of the skills space takes the given career choice. Warmer colors correspond to higher probability of the choice being taken, with deep red indicating a given choice (e.g., being a single entrepreneur in panel **Figure 5b**) having probability 1.

We thus formalize **Model Prediction I**: *Individuals with unbalanced skills are more likely to become part of entrepreneurial teams, conditional on their oveall talent.* While single entrepreneurs are high-talented generalists, **Figure 5b** shows that both high talent and high specialization increase the likelihood of forming a team. However, because of the high complementarities involved in the choice of opening a firm within a team, no type has certainty of becoming a team member: some high-talented specialists will choose to supply labor in other firms, while others will prefer to start a firm alone. Clearly this depends on the skills of the other potential team member the agent has drawn. As such, it can easily be shown that there is an indifference threshold over the $\{\theta_1, \theta_2\}$ space above which the matched agent becomes desirable as a team partner. Conversely, a team is formed if the individual is in the acceptance region of the matched agent as well.

These dynamics gives no a priori indication regarding the average productivity of

Figure 5: Team Formation and Equilibrium: Shares by Type

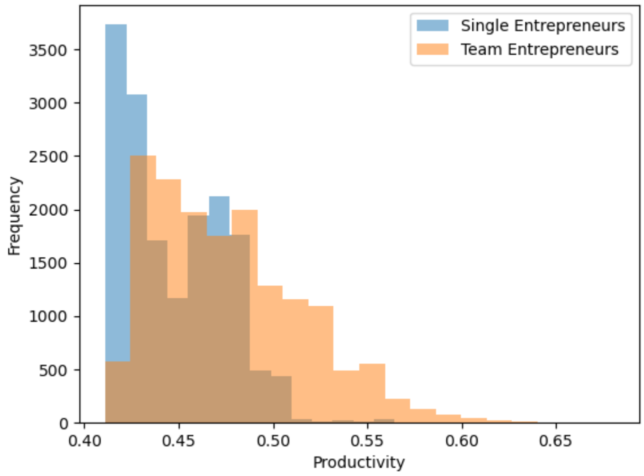


(a) Occupation 1 **(b)** Single Entrepreneur **(c)** Occupation 2 **(d)** Team Entrepreneur

team entrepreneurs vis à vis single entrepreneurs - as individual talent grows, their own acceptance region regarding potential team members shrinks, but the likelihood of entering the acceptance region of their matches increases too. We then state the less obvious **Model Prediction II**: *The productivity of entrepreneurial teams is on average higher than that of single entrepreneurs.* As shown in **Figure 6**, the productivity distributions of teams and single entrepreneurs do, in fact, overlap. However, the one of teams has a fatter right tail, implying that the average team is more productive than the average single entrepreneur. That, crucially, depends on parameter assumptions; as clear from the contour plot in panel **1b** of **Figure 1**, the occurrence of types in the extreme north-east of the skills domain, namely with very high *talent*, is rare. The number of individuals crossing the indifference curve of different career choices will then be informative of one key parameter - the correlation of skills across individuals.

As a final remark, we discuss the role of *horizontal differentiation*; how does the composition of entrepreneurial teams affect firm performance in the model? To this end,

Figure 6: Productivity Distribution of Single Entrepreneurs vs Teams



we compute two metrics of dissimilarity between team members. First, vertical dissimilarity (VD) measures the distance between the talent of the two members, and so is given by $\text{VD} = (z(\vec{\theta}_i) - z(\vec{\theta}_{i'}))^2$. Second, horizontal dissimilarity (HD) is instead computed skill-by-skill across individuals, and is given by: $\text{HD} = \sum_{j=1}^{N} \left(\theta_{i,j} - \theta_{i',j}\right)^2$.

Table 2: Team Composition and Firm Performance

	Productivity	**Sales**
Horizontal Dissimilarity	0.1409***	0.6629***
	(0.0100)	(0.0491)
Vertical Dissimilarity	-0.0379***	-0.2343***
	(0.0100)	(0.0475)

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: Firm productivity for teams is $\theta^T = z\left(\psi(\theta_1^i, \theta_1^{i'}), \ldots, \psi(\theta_N^i, \theta_N^{i'})\right)$. Sales are: $\theta^T \cdot f(\vec{L}^*)$.

The findings in **Table 2** can be summarized by the next two facts. The first one, coming from the first row of coefficients, is stated as **Model Prediction III**: *Firms run by individuals with different skills are larger and more productive.* High skill complementarity leads to *negative* sorting in single skills. The second row of coefficients gives **Model Prediction IV**: *Firms run by individuals with similar levels of talent are larger and more productive.* While individual skills are negatively sorted, positive sorting in overall talent increases firm performance. These stark predictions contrast the patterns of career choice for workers, who simply sort on their strongest skill. In our model, running a firm requires completing an array of tasks that span across specializations. For this reason, generalists are generally better entrepreneurs, as in the Lazear (2004) framework, and heterogeneous skill specialization in teams predict higher performance, as found in D'Acunto, Tate and Yang (2024). In addition, vertical differentiation across individuals leads to a sorting in overall talent that is beneficial to firm performance. A first intuitive implication of the predictions following from **Table 2** is that the distribution of both dissimilarities will be skewed, as most teams still involve relatively unspecialized individuals, but with a long tail. A second one is that, because skill dissimilarity predicts (successful) team formation, while talent dissimilarity does the opposite, the distribution of skill dissimilarities in equilibrium has a fatter tail. Both implications are evident from the plot of the equilibrium distribution of dissimilarities, presented in **Figure A.2**. In the next sections, we will explore how these predictions hold in the data, discuss the role of alternative channels, and finally quantify the aggregate contribution of team sorting to productivity and output.

3 Data

Our theory of career choices, featuring different occupations, entrepreneurship and entrepreneurial team formation, has four key predictions regarding the matching patterns of individuals in teams, and on how the (dis)similarity in entrepreneurs' skills and talent relate to firm performance. Verifying these predictions empirically requires a dataset that encompasses workers and entrepreneurs' careers, records the performance of single-owned and team-owned firms, and in which it is possible to infer or observe individuals' talent and skills. The following section explains how we put together such dataset for Portugal, and outlines few initial descriptives on entrepreneurial teams.

3.1 Sample Construction

Our main source of data is the Portuguese administrative employer-employee sample from *Quadros de Pessoal* (hereafter: QP), which contains information on roughly 4 millions of individual-firm matches per year, from 1985 to 2019. QP uses an administrative mandatory survey for all private firms, filled in during the October of every year, with information on the workforce composition of the firm for the reference month. It is very detailed in terms of individual characteristics, as it reports age, gender, nationality, education level, occupation codes[6], earnings and hours (both contractual and extra), contract characteristics (part-time vs. full-time, permanent vs temporary or seasonal contracts) and hierarchical qualification within the firm. On the firm side, it includes information on their industry code, geographical location, legal status[7], total employment, sales and founding year. Although Portugal is a relatively

[6]Over the years, the occupational classification in Portugal has changed three times. In this paper, we exploit an harmonized classification, based on ISCO-08, at the 3-digit occupational level.

[7]We have very detailed information on whether the firm is incorporated or not, and its specific kind of legal entity. This allows us to to identify precisely which firms, by their own nature, are privately owned.

small country, its firm distribution compares to that of several other OECD countries.[8]

The advantage of using the QP is twofold: first, it contains information on whether individuals are characterized as "employees", "employers" of "self-employed", which makes it possible to identify who the owner(s) (and founder(s))of firms are whenever these are privately held.[9] Second, by recording yearly labor market information for all individuals in the Portuguese labor force (each with a unique id), its longitudinal dimension allows us to observe agents' transitions between working and entrepreneurial spells. This is key for the scope of our analysis, because it enables us to observe agents' careers before starting a firm and/or before teaming up with another entrepreneur(s).

Then, to provide a measure of individual skills, we exploit the European Skills, Competences, Qualifications and Occupations (ESCO) database, which provides EU-wide links between occupations and essential skills or competences required to workers.[10] Specifically, ESCO reports a zero to one index on the intensity that each skill or competence is used in each 3-digit occupation. Different aggregation levels are available, raging from 296 granular groups to 74 or 9 very coarse categories of skills.[11]

Table 3: Descriptive Statistics from *Quadros de Pessoal*, entrepreneurs

Variable	Mean	SD	Median	P25	P75	N
Age	44.52	10.29	44	37	52	4,027,361
Age at Founding	41.75	10.00	41	34	49	3,518,134
College %	16.28	36.92	0	0	0	3,886,812
Firm Age	12.56	12.58	9	4	17	4,027,361
# Employees	2.19	0.97	2	1	3	3,849,674
Firms Owned	1.06	0.39	1	1	1	4,027,361
# Founders	1.49	1.15	1	1	2	4,027,361
# Owners	1.64	1.01	1	1	2	4,027,361
Log Sales	13.89	1.31	12.21	11.30	13.27	3,801,998

The table reports descriptive statistics for entrepreneurs in the sample, covering all years from 1985 to 2018. Sales are deflated by the 2010 CPI.

Finally, for a subsample of firms active or started between the years 2004 and 2018, we can also retrieve balance-sheet variables from the *Sistema de Contas Integradas das Empresas* (hereafter: SCIE) – for instance regarding firms' capital and debt structure, as

[8]The distribution of firms in Portugal resembles closely the one of Italy, for example. Also, note that the employment share of Portuguese firms with 10+ employees is only between 7 and 9 percentage points (p.p.) lower than that of French or German firms in the same size category.

[9]We identify as "self-employed" professionals carrying out their activity with at most one employee assisting them throughout all years observed in the data. All other employers work in firms with multiple employees at some point in time.

[10]We use the latest mapping between skills and occupations in ESCO v1.2 published in May 2024.

[11]Examples of coarse categories are *Management* or *Communication, Cooperation and Creativity*, while very granular skills required by occupations can be *Hammering, nailing and riveting*, *Tending and breeding aquatic animals*, *Analyzing business operations* or *Managing budgets or finances*.

well as their intermediate inputs. However, given the limited time frame of SCIE, we use it for robustness checks and additional analyses, and keep QP as our baseline dataset.[12]

Table 3 reports summary statistics for all QP firms of which we are able to identify the set of owners. These are 65% of firms in QP, covering 66% of sales and 76% of employment.[13]

3.2 Descriptives of Entrepreneurial Teams

To start, we provide few definitions for entrepreneurs anf firm-ownership in our sample. We consider *owners* all those individuals who are employers within a firm at a given point of its life-cycle. We instead identify as *founders* all individuals listed as owners within the first 3 years of the stated foundation year. Clearly, these two definitions tend to overlap, especially in the first 10 years of a firm's life-cycle; however, given the scope of our research question, our baseline is to focus on founding entrepreneurs.[14] In addition, note that we consider a firm to be single-owned or single-founded if it is associated to only one owner or one founder respectively, and multi-owned or multi-founded otherwise.

Figure 7 below illustrates that entrepreneurial teams are a common and relevant macroeconomic phenomenon in Portugal. Specifically, firms with more than one entrepreneur make up for roughly 40% of overall employment and sales in our sample,[15] albeit they represent slightly less than 30% of all privately held Portuguese firms. Moreover, two elements are worth stressing. First, entrepreneurial teams are a cross-industry and cross-years phenomenon, which means that our results do not hold for a specific time-frame and/or a specific sector only. Second, more than 3/4 of entrepreneurial teams are formed by two individuals, which makes the mapping to our theoretical framework relatively straightforward.

In **Table 4** we report some descriptive statistics regarding the characteristics of "single" founders with respect to team founders. Overall, no very stark difference stands out in terms of demographics for the two groups. Team founders tend to be slightly older, but are less likely to have already been the owner of a firm previously. In terms of experience and earnings team founders, who eventually found relatively more successful firms in the long run, are actually *less* likely to have got a degree, have been managers, and have lower cumulative earnings.

[12]In terms of overlap, the merge between SCIE and QP covers 97% of total employment and 88% of sales. Note that SCIE does not contain firms within the finance and insurance sectors and public services.

[13]See **Appendix B.1** and **Table C.1** for details on the characteristics of the QP and descriptive statistics at the worker level.

[14]In terms of demographics and previous careers, founders tend to be on average slightly younger and less educated than owners. Relatedly, founders' average earnings before entrepreneurship are slightly lower, and the average length of their entrepreneurial spells slightly longer compared to owners.

[15]Note that we exclude entrepreneurs from the labor headcount when computing employment shares.

Figure 7: Distributions of Firms, Employment and Sales by Number of Entrepreneurs

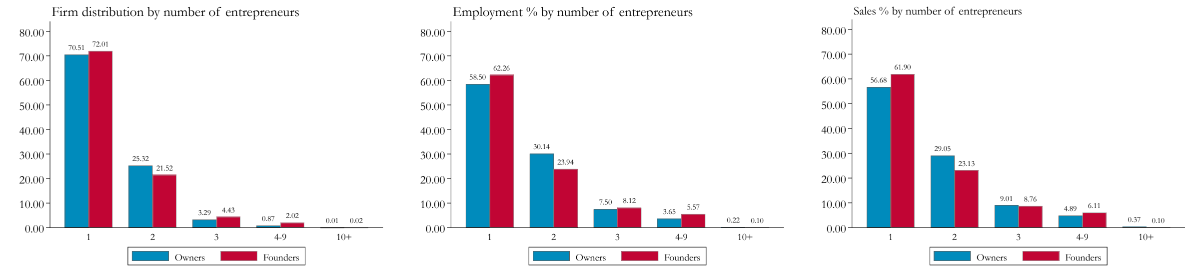


Table 4: Characteristic of single and team founders

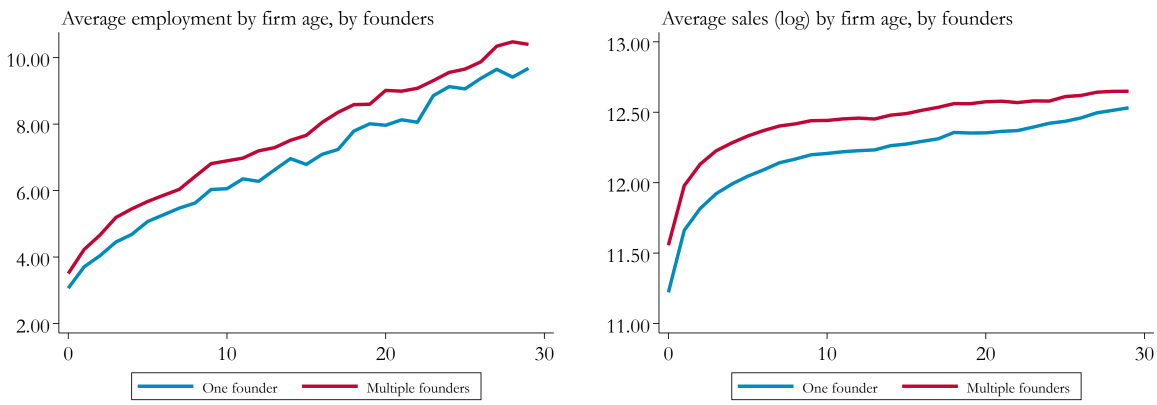
	Mean	SD	Median	P25	P75	N
Single founders						
Age at foundation	40.6	10.4	33	40	48	434,642
Sex	.302	.459	0	0	1	435,005
Sh. High educated	.159	.366	0	0	0	410,978
Sh. previously manager	.145	.352	0	0	0	192,555
Last wage	12,039	11,268	5,710	8,373	14,062	205,368
Previous 5y avg. earnings	11,444	10,234	5,640	8,124	13,412	205,472
Cumulative earnings	62,182	87,856	12,400	31,867	75,072	205,522
Previous employee jobs	5.03	4.2	2	4	7	205,522
Sh. previous entrep. exp.	.174	.379	0	0	0	435,005
Team founder						
Age at foundation	44.8	10.3	37	44	52	325,738
Sex	.289	.453	0	0	1	325,738
Sh. High educated	.128	.334	0	0	0	308,110
Sh. previously manager	.128	.334	0	0	0	131,117
Last wage	11,633	10,558	5,717	8,304	13,487	139,899
Previous 5y avg. earnings	11,058	9,532	5,688	8,075	12,911	139,945
Cumulative earnings	57,287	78,607	12,306	30,713	69,935	139,971
Previous employee jobs	4.91	4.1	2	4	7	139,971
Sh. previous entrep. exp.	.139	.346	0	0	0	325,738

The table reports descriptive statistics regarding entrepreneurs being solo founders of their firm, or members of teams. The characteristics are measured in the foundation year. All nominal values are deflated by the 2010 CPI. The share of individuals with previous entrepreneurial experience identifies entrepreneurs who, during their lives, have opened more than one firm.

3.3 Firm Performance of Entrepreneurial Teams

In what follows, we outline few distinctive characteristics – in terms of performance – of firms founded by either a single or a team of entrepreneurs. The left panel in **Figure 8** shows that firms with multiple founders have higher average employment than single-founded ones, and consistently so over their life-cycle. A similar conclusion holds when comparing average (log) sales of businesses with one or multiple founders (in the right panel of **Figure 8**).[16] We also confirm this finding for a balanced panel of firms, to ensure that results are not mainly driven by selection in and out of the sample of surviving firms.

Figure 8: Average Life-Cycle Employment and Sales by Number of Entrepreneurs



Firms with more than one founder register higher growth in the first 10 years of operations, as reported in **Figure C.2**. Moreover, **Figures C.3** and **C.4** show that multi-founded firms have higher labor productivity – computed as yearly sales over employment (or, alternatively, wage-bill) – and significantly lower exit rates over the life-cycle compared to single-founded firms. It is key to stress again that our results are not driven by time- or industry-specific patterns, and highlight that entrepreneurial teams tend to have better firm-level performances within given industries and years.

We can then exploit the subsample of firms for which we have balance sheet data and estimate yearly firm-level TFP, following common production function estimation strategies in the literature (Gandhi, Navarro and Rivers (2020)). In particular, assuming sector-level input elasticities, gross output Y in sector s for firm j at time t is defined as:

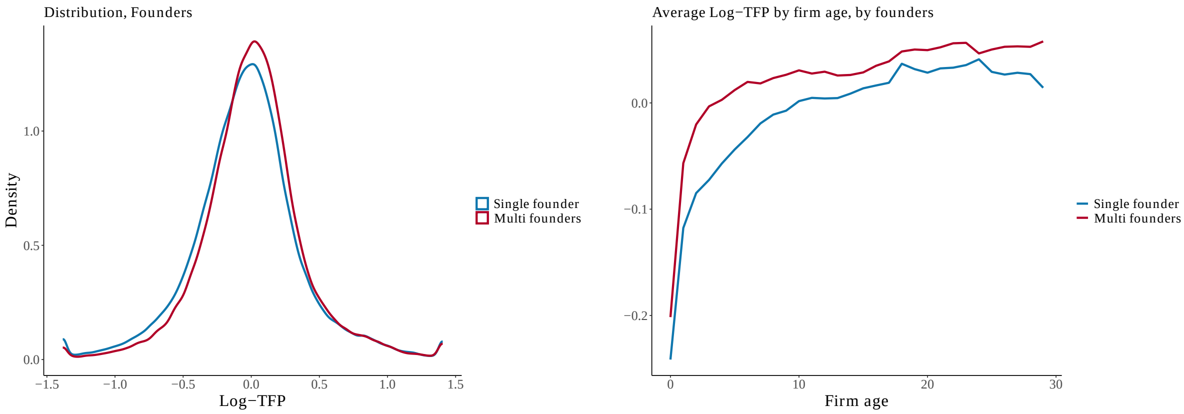
$$Y_{j,t} = e^{TFP_{j,t}} F_{s(j)}(K_{j,t}, L_{j,t}, M_{j,t})$$

where K is computed via perpetual inventory methods (PIM) with sectoral depreciation rates (*OECD-STAN)* and deflators (*EU-KLEMS*), L is total employment (headcount) and M are intermediate inputs (services and materials). Since TFP has a strong sectoral component, we further residualize our estimates using sector × year fixed-effects (FE). **Figure 9** highlights that TFP is higher in firms founded by a team, with the distribution on the left panel showing a higher mean and a fatter right tail.

Summarizing the evidence presented so far, our initial exploratory analysis seems to provide support to **Model Prediction II**: *The productivity of entrepreneurial teams is on average higher than that of single entrepreneurs.* It is important to stress that, through the

[16]When using the definitions of single and multi-owned firms (instead of single and multi-founded), we observe even larger differences in their average employment and sales, as reported in **Figure C.1**. This could be due to the fact that successful single-founded firms may attract more owners as they grow.

Figure 9: Residualized Firm-level log-TFP by Number of Entrepreneurs



The figure reports the distribution of average firm-level log-TFP (left) and the average firm-level TFP by firm-age (right) in the matched SCIE-Quadros de Pessoal sample for firms with single and multiple founders. Firm-level TFP is estimated using Gandhi, Navarro and Rivers (2020) separately for each sector (one-digit) and then residualized by sector and year effects.

lens of the theoretical framework presented in Section 2, the higher average productivity of multi-founded relative to single-founded firms is due to two key mechanisms: (i) a stronger selection process of individuals into team entrepreneurship compared to single entrepreneurship, and the (ii) sorting (in talent and skills) between individuals in teams.

In particular, our theory predicts that individuals with balanced skills are more likely than others to become entrepreneurs, whereas those with unbalanced skills tend to take specialist roles outside of entrepreneurship. Therefore, in order for a given agent to prefer team entrepreneurship over their outside options, they must have relatively high levels of at least one skill and, equally important, to have met a potential business partner with relatively high levels of at least one complementary skill. Indeed, it is not two talented entrepreneurs, but rather two *complementary* in skills and highly *talented* entrepreneurs that predicts better firm performance. The next section explore in detail these two mechanisms, by specifically taking Model Predictions I and III to our data.

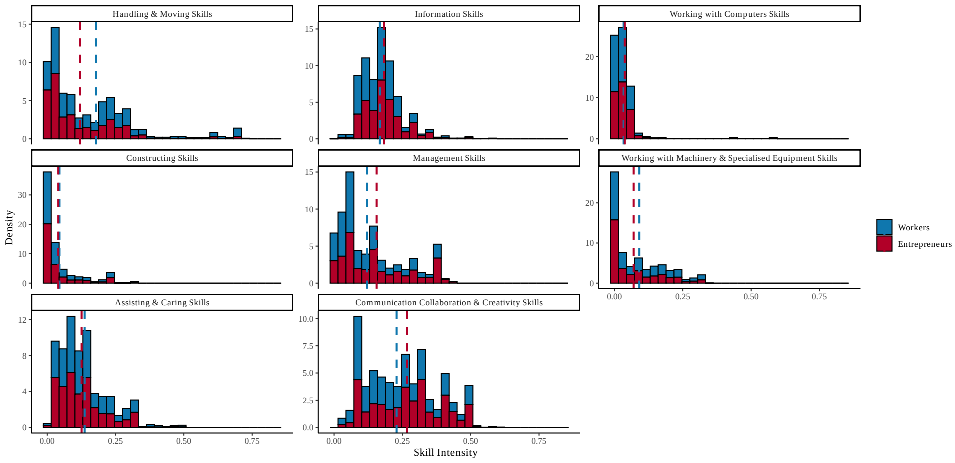
4 Sorting Patterns of Entrepreneurial Teams

4.1 Measuring Skills, Talent and Similarities

In order to explore the sorting patterns of individuals within entrepreneurial teams and link them to the predictions of our theory, we first need to define and measure three key variables: Agents' (i) skills, (ii) talent, (iii) and the similarity of (i) and (ii) among individuals in entrepreneurial teams. We explain how we construct these variables below.

Skills. Exploiting the ESCO mapping between 3-digit occupations and the essential skills required to workers, we construct time-varying measures of (cumulative) individual skills, exploiting information on agents' occupational history until any point in time t, and weighting ESCO's skill indexes by the years agents spent in each occupation. Recall that different aggregation levels are available within the ESCO database, so we construct two measures of individual skills, one using 8 coarse skill categories (mostly for graphical purposes) and one based on 74 finer ones. Importantly, for workers that eventually become entrepreneurs, we consider employment histories before their first entrepreneurial spell.

Figure 10: Skill Distribution for Workers and Entrepreneurs



The figure presents the distribution of the broader ESCO categories for workers in our sample and for the subset that eventually become entrepreneurs. Dashed lines report the averages for each distribution.

Figure 10 plots the first of these two measures of skill intensity (based on 8 skill groups) for workers and entrepreneurs in our sample, using their cumulative occupational history until any given time t. It is interesting to note that, relative to agents that will stay workers their entire careers, those who eventually become entrepreneurs over their life-cycle – whether alone or in a team – have on average higher levels of managerial competences, as well as higher levels of communication, collaboration and creativity skills.

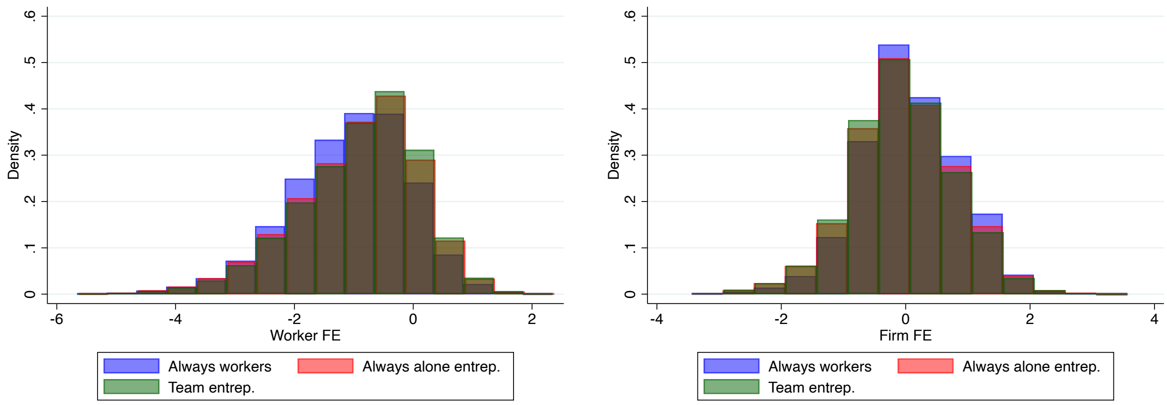
While the richness of the QP data allows us to observe individuals' entire occupational history and use it jointly with the ESCO database for a measure of their competences, this is only one of the two elements defining agents' skill bundles for the scope of this analysis. Specifically, through the lens of the theoretical framework presented in Section 2, individuals' career choices are informed not only by the relative composition of their skill bundles, but also by how talented agents are, namely by their overall level. For workers, it is the level and relative combination of skills that define occupational choices and agents' final wages. Moreover, for single and team entrepreneurs, the level and relative combination of skills map directly into the productivity of their firms. So, how can we measure talent? We exploit again agents' working histories, but now focus on their wages.

Talent. In our QP sample, we observe yearly wages for all individuals employed in any given t and for all the years in which they are working. This is true also for those who eventually become entrepreneurs: in particular, as the average age of an entrepreneur is 45 years old, we have on average 20 years of wage histories pre-entrepreneurship. Borrowing the strategy from Abowd, Kramarz and Margolis (1999) (hereafter: AKM), we hence estimate unobserved worker and workplace heterogeneity via the following regression:

$$\log(w_{i,t}) = X'_{i,t}\beta + \alpha_i + \psi_{j(i)} + \epsilon_{i,t} \tag{3}$$

where $X_{i,t}$ include age^2 and year FE, α_i measures latent worker quality, and $\psi_{j(i)}$ measures latent workplace quality. For individuals that become entrepreneurs at least once in their career, worker types (or FE) reflect their type as workers before the first entrepreneurial spell, while firm types (or FE) are intended as their past workplace types, *before* they opened their own firm.[17] We then plot the distribution of these estimated worker and workplace FEs in **Figure 11**, distinguishing between individuals that remain workers their entire career, those that become entrepreneurs at least once but always alone, and those that become entrepreneurs at least once and at least once in a team.

Figure 11: Distribution of worker and firms' FEs, by entrepreneurial type



The figures present the distributions of workers' and firms' fixed effects, as estimated by a standard AKM specification as in equation 3. The figure pools all years, with fixed effects coming for every year from AKM specifications estimated on a 5 years backward looking rolling window. For entrepreneurs, the relevant fixed effects come from the last year before the first entrepreneurial spell. Individuals are identified as "always workers" if they never undertook any entrepreneurial activity, "always alone entrepreneur" if they were at any point entrepreneurs, but never participated to a team, or "team entrepreneurs".

Two observations can be made looking at **Figure 11**. First, there is positive selection into entrepreneurship based on worker qualities, and a small negative selection based on workplace (in this case: past workplace) qualities. This result supports the prevalent view in entrepreneurship, dating back to Lucas (1978), suggesting entrepreneurs are positively selected on overall productivity compared to workers. Second, and a key contribution of our empirical analysis, entrepreneurs who open a firm with a team at least once in their careers tend to be more positively selected on their worker type (i.e. their FE) compared

[17] We refer the reader to **Appendix B.4** for details regarding the AKM estimation.

to single entrepreneurs, as well as to individuals who will always be workers. This finding connects to **Model Prediction II**, and helps explain why firms by entrepreneurial teams are more productive than those by single entrepreneurs, as we will further clarify later.

Similarity. As both our measures of talent and skills are highly-dimensional and difficult to compare across teams of more than two-individuals, we reduce the dimensionality of these quality metrics by computing two similarity indexes. These indexes summarize how similar founding teams are in relation of their members' cumulative skills as workers and AKM FEs. Specifically, for each founding team, we compute the average of the pairwise Gower index across skills and AKM FEs for each entrepreneur in the team.[18] Formally, the index over C characteristics for the pair (i, j):

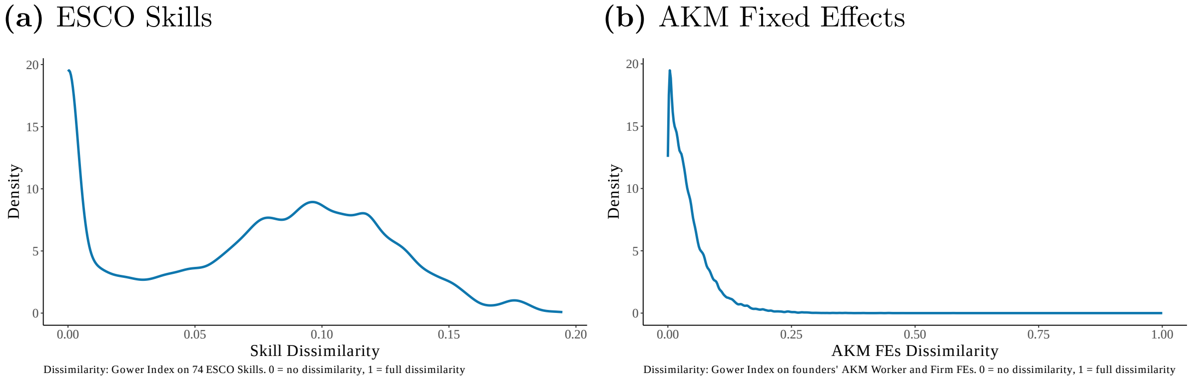
$$G_{i,j} = \frac{\sum_{c=1}^{C} w_{i,j,c} g_{i,j,c}}{\sum_{c=1}^{C} w_{i,j,c}},$$

with

$$g_{i,j,c} = \begin{cases} \frac{|c_i - c_j|}{R_c}, & \text{for continuous variables. } R_c \text{ is the range of } c. \\ \mathbb{I}\{c_j = c_i\}, & \text{for binary or ordinal variables.} \end{cases}$$

By construction, the index is bounded between zero and one, where the lower bound indicates perfect similarity and the upper bound perfect dissimilarity. Hence, founding teams with a high average value of any or both indexes – skills and AKM FEs – are composed by more heterogeneous entrepreneurs compared to teams with lower scores. **Figure 12** plots the distributions of these similarity indexes for the teams of business founders in our sample. What emerges is that entrepreneurial teams are more dissimilar in latent types – i.e. their talent – than they are in skills, consistent with the model-implied distributions of team entrepreneurs' skills and type similarity depicted in **Figure A.2**.

Figure 12: Skill and Type Similarity in Founding Teams



The figure reports the distributions for the Gower indexes across seventy-four ESCO skills (a) and the estimated AKM fixed effects (worker and past employer) in (b) for each founding team in our sample. Gower indexes are computed pairwise within each team and then averaged.

[18]To compute skill similarity within teams, we exploit the finer 74 ESCO skill categories.

4.2 Predictors of Team Formation

With these measures of individuals' talent and skills, we proceed to validate **Model Prediction I** regarding which characteristics are good predictors of joining a team of founders by running the following linear probability model at the founder-team level:

$$\mathbb{I}\{|\text{Co-founders}|_{f(i)} > 0\} = \beta_0 \log(\bar{w})_i + \beta_1 \hat{\alpha}_i + \beta_2 \hat{\psi}_{j(i)} + \beta_3 \sigma(\boldsymbol{\theta}_i) + \beta_4 \mathbb{I}\{\text{Serial}_i\} + \boldsymbol{\beta}' \mathbf{X}_i + \Phi + \epsilon_i, \quad (4)$$

where $\mathbb{I}\{|\text{Co-founders}|_{f(i)} > 0\}$ is a dummy variable that takes a value equal to one if entrepreneur i has at least one co-founder in the founding team of firm f. On the right-hand side, we include the log of (cumulative) previous labor market earnings, $\log(\bar{w})_i$, the estimated AKM fixed effects, $\hat{\alpha}_i$ and $\hat{\psi}_{j(i)}$, the standard deviation of entrepreneurs' ESCO skills $\sigma(\boldsymbol{\theta}_i)$ as a measure of their specialization,[19] and a set of fixed effects and additional controls to account for individual characteristics or sector-time variation at the time of the team formation. We report the results of this estimation in **Table 5**.

Two main findings emerge. First, high previous labor market earnings reduce the likelihood of joining a founding team with a fellow entrepreneur. This result is consistent with *some* entrepreneurial teams emerging as a response to the presence of liquidity constraint, as discussed in Evans and Jovanovic (1989). Second, higher talent (or quality), proxied either by the AKM FE as a worker or by the AKM FE of the last workplace before becoming an entrepreneur, is a good predictor of the likelihood of having a co-founder, as shown in Columns 1 through 3. The degree of specialization, instead, increases the likelihood of forming a founding team only when controlling for the level of individual abilities, either by their skills' levels (Columns 2 and 3), or by including a fixed effect for the previous occupation held as a worker (Column 3).[20] These results, although only suggestive, are in line with **Model Prediction I**: *Individuals with unbalanced skills are more likely to become part of entrepreneurial teams, conditional on their overall talent.*

4.3 Team Composition and Firm Performance

In what follows, we combine all these elements together and ask what the production function of entrepreneurial teams is in our data. Are entrepreneurs in teams positively sorted with respect to their worker (and/or past workplace) types? Are they negatively sorted with respect to their skills? And, finally, in which directions do the sorting patterns of team entrepreneurs – along their talent and skills – affect firm-level performance?

To start, **Figure 13** plots the correlation coefficient between the worker (on the left) and workplace types (on the right) of entrepreneurs within 2-member teams. Clearly,

[19] A high standard deviation in the measures of individual skills implies a more specialized entrepreneur, as it signals they had occupations characterized by high values of the ESCO indexes only on few skills.

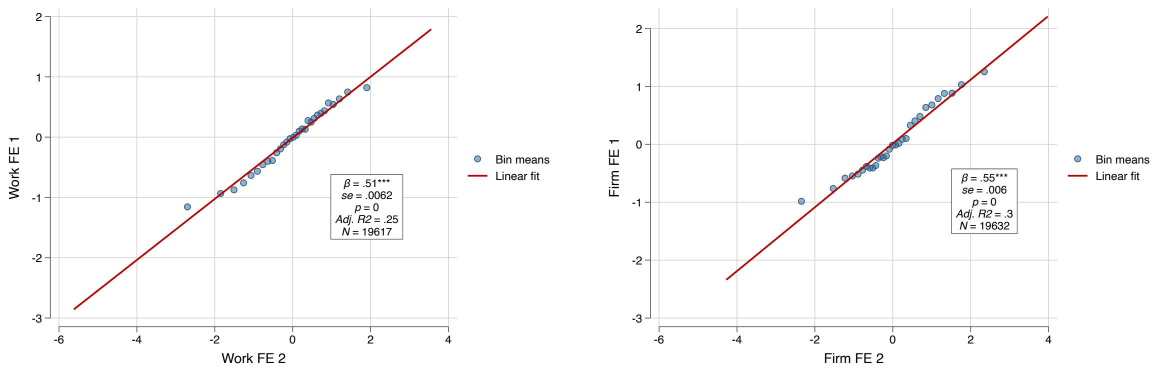
[20] Approximately 90% of entrepreneurs is employed in 2 or fewer occupations (with more than 50% reporting only one), therefore the last occupation as a worker is a good proxy of their overall skill level.

Table 5: Team formation

Dependent Variable:	$\mathbb{I}\{\lvert\text{Co-Founders}\rvert > 0\}$		
Model:	(1)	(2)	(3)
Prev. (Log) Cumulative Earnings	-0.017***	-0.016***	-0.015***
	(0.004)	(0.004)	(0.001)
Work FE	0.027***	0.026***	0.026***
	(0.006)	(0.006)	(0.001)
Firm FE	0.014***	0.014***	0.013***
	(0.002)	(0.002)	(0.002)
S.D. in Skills	-0.128**	0.070	0.114**
	(0.037)	(0.067)	(0.056)
Fixed-effects			
College	Yes	Yes	Yes
Sex	Yes	Yes	Yes
Age at found.	Yes	Yes	Yes
Recession × 2dgts Sector	Yes	Yes	Yes
Prev. Occupation			Yes
Additional Controls			
Skill Levels	No	Yes	No
Fit statistics			
Observations	211,844	211,844	204,990
R^2	0.048	0.050	0.054

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Figure 13: Correlation of Worker and Past Workplace Types for Entrepreneurial Teams

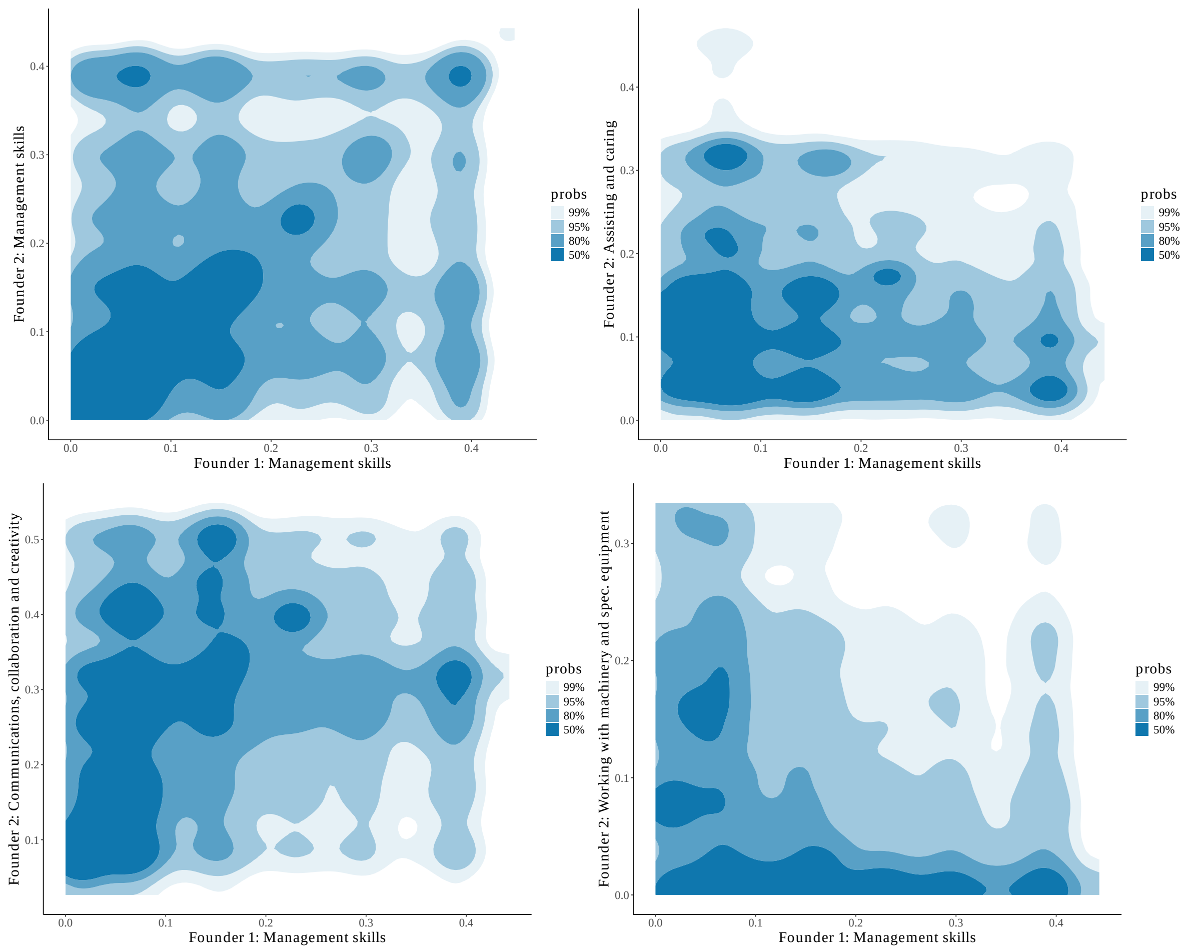


The figures present binned scatterplots of individuals' (left) and workplaces (right) AKM fixed effects for entrepreneurs in two-member teams. Fixed effects are estimated for every year on a 5 years backward looking rolling window. The fixed effects come from the last year before the first entrepreneurial spell.

there is a strong positive correlation in talent between entrepreneurs within teams. This suggests that – empirically – entrepreneurial teams tend to be characterized by a *vertical similarity* in latent types, further reconciling the mechanisms at work in our model with the evidence from the data. Moreover, note that this result holds true when residualizing the correlations in **Figure 13** by a host of individuals' covariates, and when following an alternative strategy by Bonhomme, Lamadon and Manresa (2019) to estimate worker and workplace types (we report these robustness checks in **Figures C.7** and **C.6**).[21]

As a second step, **Figure 14** plots the density of entrepreneurial teams, where on the x-axis is the intensity of managerial skills for one firm founder, and on the y-axis the intensity of other skills for the other founder.[22] For graphical purposes, we exploit the broad 8 ESCO skill groups, and darker colors in the figure represent higher densities. What emerges from this analysis is that there is a high degree of heterogeneity in observable skills for teams with two founders, and that entrepreneurial teams are more likely to be formed by individuals with complementary skills. This suggests that – empirically – entrepreneurial teams tend to be characterized by a *horizontal dissimilarity* in skills, further reconciling the mechanisms at work in our model with the evidence from the data.

Figure 14: Distribution of Skills in Founding Teams



The figure reports the joint distribution of skills in two-founders teams for the managerial skill of one founder against selected skills of the other founder in the team.

[21] We refer the reader to **Appendix B.4** for details on the clustering procedure.
[22] **Figure C.8** reports all sixty-four pairwise skill distributions of two-founders teams.

Having established that teams are characterized by horizontal dissimilarity in entrepreneurs' skills and vertical similarity in their talent, we finally quantify the effect of these sorting patterns on firm performance. To this end, we estimate the following:

$$\overline{\log(y)}_{f,t} = \beta_0 \overline{\hat{\alpha}}_f + \beta_1 \overline{\hat{\psi}}_f + \beta_2 \Delta_f(\hat{\alpha}, \hat{\psi}) + \beta_3 \Delta_f(\boldsymbol{\theta}) + \Phi + \epsilon_f, \tag{5}$$

where $\overline{\log(y)}_{f,t}$ is the 3 years moving average of log sales, $\overline{\hat{\alpha}}_f$ and $\overline{\hat{\psi}}_f$ are the founders' average worker and past workplace FEs, $\Delta_f(\hat{\alpha}, \hat{\psi})$ is the average dissimilarity of the entrepreneurs' types in the founding team and $\Delta_f(\boldsymbol{\theta})$ is the average dissimilarity of founders' skills within the founding team[23], and Φ is a set of fixed effects that control for sector and time variation and the firm incorporation type. Importantly, we use firm-level sales to maximize the number of data points. Unfortunately, other metrics such as TFP are available only for the reduced SCIE sample, and the estimation of the AKM FEs and the skill dissimilarities are already demanding on the data, as they require information on past careers (both wages and occupations) for all team members.

Table 6: Founding team characteristics and firm performance

Dependent Variables:	$MA_3[Log(Sales)]_{t=5}$					$MA_3[Log(Sales)]_t$
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Work FE	0.204***				0.379***	0.359***
	(0.036)				(0.013)	(0.007)
Avg. Firm FE	0.147***				0.215***	0.252***
	(0.016)				(0.014)	(0.009)
FE dissimilarity		-0.765***		-0.924***	-0.428*	-0.253***
		(0.130)		(0.143)	(0.187)	(0.018)
Skill dissimilarity			0.897***	1.03***	0.978***	0.851***
			(0.108)	(0.073)	(0.067)	(0.054)
Fixed-effects						
Incorporation type	Yes	Yes	Yes	Yes	Yes	Yes
Sector × Founding Year	Yes	Yes	Yes	Yes	Yes	Yes
Firm Age						Yes
Sector × Year						Yes
Fit statistics						
Observations	26,816	6,174	6,734	5,708	5,708	129,150
R^2	0.265	0.261	0.251	0.258	0.287	0.303

*Clustered (Incorporation type) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 6 reports the main coefficient of interest from the specification in **Equation 5**. In Columns (1) to (4) we document the effects of each regressor and we report our preferred specification in Column (5) for the cross-section of firms five years after foundation, while in Column (6) we run a version of specification (5) for all firms in our sample, irrespective of their foundation year. First, the coefficients on the average AKM fixed effects are positive and significant, indicating that founding teams that – on

[23]Dissimilarity measures are the pairwise Gower indexes for each founder in the team, then averaged.

average – are formed by better workers, tend to perform better in the in the medium-run. However, the coefficient on the measures of similarity across FE – that we take as a proxy for the team *vertical dissimilarity* – are both economically and statistically significant, indicating that founding teams that are 10% more vertically dissimilar are associated with approximately 4% less sales five years after foundations. Founding teams that instead are *horizontally dissimilar* – proxied by our measure of skill dissimilarity – tend to over-perform relative to less heterogeneous teams. Also in this case, the effects are both statistically and economically significant, indicating that firms founded by a 10% more diverse team enjoy almost 10% higher sales five years after incorporation. When pooling all the available years, the results remain very stable, as shown by Column (6). Both these results are consistent with the model's predictions in **Table 2**.

4.4 Robustness Analysis

As a concluding remark, we briefly discuss few robustness exercises, which lend support to the importance of the sorting of entrepreneurs into teams based on their talent and skills, in addition to other relevant motives for which team entrepreneurship may occur.

4.4.1 Financial Frictions and the Business Cycle

A valid concern is that our results on the patterns of entrepreneurial sorting and their relevance for firm performance could be driven just by the existence of financial frictions, or that teams were simply a response to negative (or positive) movements of the business cycle. On the former point, we are able to exploit information on firms' financial variables from the merge of QP with SCIE balance sheets. **Table C.2** shows that entrepreneurs' talent, skills, and the similarity of these within teams are relevant for life-cycle sales beyond the effect of firms' initial capital. On the latter concern, **Table C.5** clarifies that the stock and the flow of entrepreneurial teams have no main cyclical component.

4.4.2 Changes to Team Composition

To further highlight the relevance of founders within entrepreneurial teams in terms of the performance of their firms, we perform a similar exercise to Choi et al. (2023) and check what happens to firm sales when a founding member leaves the team. Since leaving a team could be endogenous to firm's performance, we condition on those individuals that leave the firm (not close to retirement age) and disappear forever from the QP, namely from the Portuguese labor force. This can happen, for instance, in case of death or migration. **Table C.3** shows that there is a negative and statistically significant relation between a founder leaving the firm and the change in firm sales, controlling for a host of variables, including that same firm fixed effect.

5 Conclusions

We study the sorting of individuals into entrepreneurial teams and establish sorting as a critical determinant of firm productivity and long-run performance. The paper proposes a simple theoretical model of career and entrepreneurial choices. In the model, individuals with complementary yet unbalanced skills are more inclined to join entrepreneurial teams, resulting in higher productivity for team-based ventures compared to solo entrepreneurship. Our empirical analysis leverages comprehensive administrative data from Portugal. By linking entrepreneurs' pre-firm occupational trajectories to subsequent firm performance, we demonstrate that positive sorting along talent, and negative sorting along skills specialization, is associated with larger firms, increased sales, and improved survival rates. Our findings suggest that the micro-level dynamics of team formation are driven primarily by intrinsic attributes rather than by external financial or cyclical constraints. Our paper contributes to the joint understanding of the equilibrium interactions between firm entry and exit and labor market dynamics. The framework naturally lends itself to the analysis of policies aimed at improving quantity and quality of firm entry.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333. ISBN: 00129682.

Acabbi, Edoardo Maria, Andrea Alati, and Luca Mazzone. 2024. "Human Capital Ladders, Cyclical Sorting, and Hysteresis." *Cyclical Sorting, and Hysteresis (March 28, 2022).*

Argan, Damiano, Leonardo Indraccolo, and Jacek Piosk. 2024. "Teach the Nerds to Make a Pitch: Multidimensional Skills and Selection into Entrepreneurship."

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. "Cross-country differences in productivity: The role of allocation and selection." *American economic review*, 103(1): 305–334.

Bhandari, Anmol, Tobey Kass, T May, E McGrattan, and Evan Schulz. 2022. "On the nature of entrepreneurship." *SOI Working Papers, Internal Revenue Service.*

Bias, Daniel, and Alexander Ljungqvist. 2023. "Great Recession Babies: How Are Startups Shaped by Macro Conditions at Birth?" *Swedish House of Finance Research Paper*, , (23-01).

Boerma, Job, Aleh Tsyvinski, and Alexander P Zimin. 2025. "Sorting with Teams." *Journal of Political Economy*, 133(2): 421–454.

Boerma, Job, Aleh Tsyvinski, Ruodu Wang, and Zhenyuan Zhang. 2023. "Composite sorting." National Bureau of Economic Research.

Bonhomme, Stephane, Thibaut Lamadon, and Elena Manresa. 2019. "A Distributional Framework for Matched Employer Employee Data." *Econometrica*, , (87): 1–71.

Cagetti, Marco, and Mariacristina De Nardi. 2006. "Entrepreneurship, frictions, and wealth." *Journal of political Economy*, 114(5): 835–870.

Choi, Joonkyu, Nathan Goldschlag, John C Haltiwanger, and J Daniel Kim. 2021. "Founding teams and startup performance."

Choi, Joonkyu, Nathan Goldschlag, John Haltiwanger, and J Daniel Kim. 2023. "Early joiners and startup performance." *Review of Economics and Statistics*, 1–46.

D'Acunto, Francesco, Geoffrey A. Tate, and Liu Yang. 2024. "Entrepreneurial Teams: Diversity of Skills and Early-Stage Growth."

De Haas, Ralph, Vincent Sterk, and Neeltje Van Horen. 2022. "Start-up types and macroeconomic performance in Europe."

Edmond, Chris, and Simon Mongey. 2021. "Unbundling labor." Working Paper.

Eeckhout, Jan, and Philipp Kircher. 2011. "Identifying sorting—in theory." *The Review of Economic Studies*, 78(3): 872–906.

Evans, David S, and Boyan Jovanovic. 1989. "An estimated model of entrepreneurial choice under liquidity constraints." *Journal of political economy*, 97(4): 808–827.

Feliz, Sonia, Sudipto Karmakar, and Petr Sedlácek. 2021. "Serial entrepreneurs and the macroeconomy."

Freund, Lukas. 2022. "Superstar Teams: The Micro Origins and Macro Implications of Coworker Complementarities." *Available at SSRN 4312245.*

Gandhi, Amit, Salvador Navarro, and David A Rivers. 2020. "On the identification of gross output production functions." *Journal of Political Economy*, 128(8): 2973–3016.

Gendron-Carrier, Nicolas. 2024. "Prior Work Experience and Entrepreneurship: The Careers of Young Entrepreneurs."

Gyetvai, Attila, and Eugene Tan. 2023. "The Role of Human Capital Specificity in Entrepreneurship."

Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon M Phillips. 2024. "Production and learning in teams." *Econometrica*, 92(2): 467–504.

Humphries, John Eric. 2022. *The causes and consequences of self-employment over the life* *cycle.* The University of Chicago.

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg. 2021. "Learning from coworkers." *Econometrica*, 89(2): 647–676.

Lazear, Edward P. 2004. "Balanced skills and entrepreneurship." *American Economic Review*, 94(2): 208–211.

Levine, Ross, and Yona Rubinstein. 2017. "Smart and illicit: who becomes an entrepreneur and do they earn more?" *The Quarterly Journal of Economics*, 132(2): 963–1018.

Lucas, Robert E. 1978. "On the size distribution of business firms." *The Bell Journal of Economics*, 508–523.

Mukoyama, Toshihiko, and Aysegül Sahin. 2005. "Patterns of Specialization." mimeo. Concordia University and Federal Reserve Bank of New York.

Poschke, Markus. 2013. "Who becomes an entrepreneur? Labor market prospects and occupational choice." *Journal of Economic Dynamics and Control*, 37(3): 693–710.

Queiró, Francisco. 2022. "Entrepreneurial human capital and firm dynamics." *The Review of Economic Studies*, 89(4): 2061–2100.

Sterk, Vincent, Petr Sedláček, and Benjamin Pugsley. 2021. "The nature of firm growth." *American Economic Review*, 111(2): 547–579.

Syverson, Chad. 2011. "What determines productivity?" *Journal of Economic literature*, 49(2): 326–365.

A Model Appendix

Figure A.1: Bilateral Meeting with $\theta_1, \theta_2 = [0.45, 0.45]$ and Occupation Choices

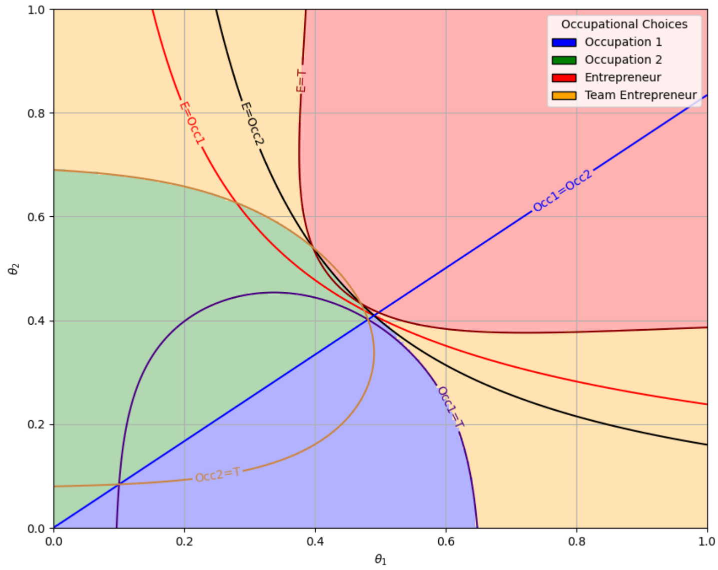
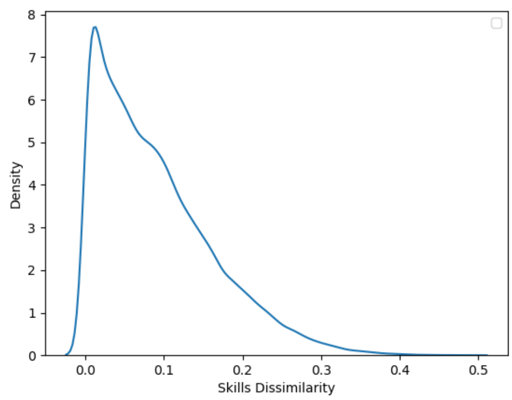
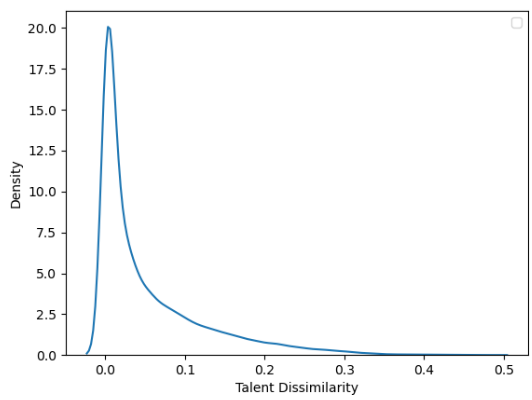


Figure A.2: Specialization and Talent Similarity in Founding Teams: Model

(a) Horizontal Differentiation



(b) Vertical Differentiation



B Data Appendix

B.1 Quadros de Pessoal

The main data source is the *Quadros de Pessoal* (hereafter QP) for the 1985-2019 period. The data are gathered annually by the Portuguese Ministry of Employment through an questionnaire that every establishment is obliged by law to fill in. The dataset does not cover the public administration and non-market services, whereas it covers partially or fully state-owned firms, provided that they offer a market service. The dataset covers virtually the entire population of firms with at least one employee. The dataset contains a snapshot of firms' employment in October each year, and when relevant firms also report the identity of the individuals self-identifying as employers. It contains information on industry, hiring date, the kind of job contract (fixed-term or open-ended), the effective number of hours worked, and different types of compensation. This implies that jobs (hence earnings, days worked and daily wages) are not recorded for a worker who is not employed in October. The dataset is hierarchically composed by a firm-level dataset, an establishment-level dataset and a worker-level dataset.

The firm level dataset made available to us contains information on the firm location at NUTS 2 level,industry of operation (CAE rev. 1 until 1994, rev. 2 until 2002, rev. 2.1 until 2006 and rev. 3, based on NACE-Rev. 2 Statistical classification of economic activities in the European Community), total employment, total sales, ownership structure and legal incorporation. Analogous information is available on the establishment-level dataset. The worker level dataset provides detailed information on worker characteristics and contracts. Information included comprehends workers' gender, age, nationality, detailed occupational code (the *Classificação Nacional de Profissões* (CNP94) up to 2009 and the *Classificação Portuguesa das Profissões* (CPP2010) from 2010 onward, which is based on ISCO08 International Occupational Classification Codes), detailed educational level, qualification within the firm[24]. At the contract level it is possible to know the precise hiring date, the kind of contract (various typologies that generally define the contract as fixed-term or open-ended, from 2000), the hours arrangement (full-time versus part-time), the effective number of hours worked, and information on the compensation. More specifically, for each worker it is possible to obtain information on the base pay, any extra paid in overtimes or other extra-ordinary payments and other irregularly paid components. There is no information on social security contributions. As regards employers, the dataset reports detail on their hierarchy and occupation within the firm, but information on compensation is almost entirely missing.

We perform several minimal checks on the data to eliminate inconsistencies in individuals identification and demographic characteristics over time. We follow Caliendo et al. (2020) and Mion, Opromolla and Sforza (2022) in harmonizing the sectoral codes across years, and use firms own changes in occupational definitions for continuing contracts to create a

[24]As regards the qualification categories, the Portuguese Decree-Law 380/80 established that firms should indicate the qualification level as in the Collective Agreement. If this is not available, firms should select the qualification level of the worker. These categories are based on the degree of complexity of tasks that the worker performs within the firm (from more basic, routine tasks to more discretionary managerial ones). The categories are defined within a 9 levels hierarchy, that we simplify into three broad categories.

frequency-based transition table between occupational codes. For each worker, we select the main job as the highest paid job during the year. We report in **Table C.1** descriptive statistics for workers in the sample, covering all years from 1985 to 2018.

B.2 Sistema de Contas Integradas das Empresas

The *Sistema de Contas Integradas das Empresas* (henceforth SCIE) is a firms level balance-sheet and income statements database, created by the Instituto Nacional de Estatisticas (hereby INE), combining several administrative and survey sources from various other Portuguese institutions. Our dataset consists of a repository of yearly economic and financial information on the universe of non-financial corporations operating in Portugal from 2004 to 2019. It includes information on sales, balance-sheet items, profit and loss statements, and cash flow statements (after 2009) for private firms in Portugal (with the sxclusion of the public sector, finance and insurance businesses).[25]

The dataset contains a great amount of information on enterprises' balance sheets and income statements, but has limited information on sole proprietorships. We use the dataset to obtain information on total assets, fixed assets, interest expenditures, cash-flow and capital expenditures (after 2009), cash balances, exports and export status, value added and profits.

The coverage of SCIE in the QP is not complete, but is extremely high. Firms present in both datasets account for 98% of the total number, 96% of employment and 96% of sales, for the years in which the data exists.

B.3 Variables definition for the entrepreneurs dataset

We identify as owners all individuals who are identifies as "employers" in the QP worker level records. Of these, we identify as founders all owners present in the firm within three years of its foundation date.

Entrepreneurs can be further characterized as *serial* when they own multiple enterprises at the same time, and/or *sequential*, if they ever own more enterprises but not necessarily at the same time.

For all entrepreneurs with a work history, we obtain characteristics regarding their past work career *before* their first spell as entrepreneurs.[26] We calculate quantiles of several characteristics for their work careers upon becoming entrepreneurs: size of the

[25]After 2009, in order for the data to comply with international accounting standards, there has been a major overhaul of the variables definitions in the dataset, from the *Plano Oficial de Contabilidade* (POC) to the *Sistema de Normalização Contabilística* (SNC). In all our computations, unless otherwise noted, we have personally gone through a variables' harmonization process.

[26]We can identify work histories for 44% of owners in the data.

firm, sales, last five years of earnings, cumulative career earnings, tenure, age of the firm for the last employer. We also calculate, when possible conditional of belonging to the relevant connected set, worker and firm fixed effects as in Abowd, Kramarz and Margolis (1999).

Eventually, we are able to identify owners for 65% of firms, covering 66% of sales and 76% of employment in the QP.

B.4 AKM specifications

In order to extract worker and firms fixed effects (hereafter: AKM) as in Abowd, Kramarz and Margolis (1999), we run the following regression:

$$\log(w_{i,t}) = X_{i,t}'\beta + \alpha_i + \psi_{j(i)} + \epsilon_{i,t}$$

where $X_{i,t}$ include age^2 and year FE, α_i measures latent worker quality, and $\psi_{j(i)}$ measures latent workplace quality. The estimation of the fixed effects relies on the concept of connected set, that is the set of all firms connected by worker mobility. In order to properly disentangle the individual and workplace effects one needs to have workers moving across different firms. This in turn implies that if firms do not experience worker flows with firms in the connected set, no estimation is feasible for them. Given the presence of some very small (and isolated) firms in our dataset, the connected set does not cover the entirety of the labor market.

One way to overcome this limitation is to give up the estimation of workplace effects, and aim at estimating effects corresponding to more broadly defined categories that can expand the connected set. That is the approach in Bonhomme, Lamadon and Manresa (2019*b*), who employ a K-means clustering algorithm (MacQueen et al., 1967, Lloyd, 1982) to the empirical cumulative distribution function of wages at the firm level to characterize broadly defined "firm-types". We use for robustness analysis the same technique to identify 10 clusters of firm types, by pooling all years in the datasets for the clustering procedure.[27]

As we want individual and workplace effects for entrepreneurs to be proxies of their talent and career characteristics *before* their entrepreneurial career starts, we estimate them only for the years before the first entrepreneurial spell. This amounts to estimating our AKM model or backward-looking rolling windows of years. Specifically, for every year in the data we run the AKM specification on the connected set estimated on the current year of analysis and the five years prior. Then, for entrepreneurs, we assign to

[27]The procedure is typically proposed to attenuate the so-called "limited mobility bias" problem (Andrews et al., 2008), which is however relevant for variance-decompositions and calculation of sorting with the estimates. We mainly use it to expand our connected set of estimation.

them the most recently estimated individual fixed effect as a proxy of skill or talent on the workplace, and the most recent firm effect as a proxy of the unobserved quality of the last workplace before the decision of undertaking an entrepreneurial activity.

C Appendix Tables and Figures

Table C.1: Descriptive Statistics from *Quadros de Pessoal*, workers

	Mean	SD	Median	P25	P75	N
Age	37.2	11.2	36	28	45	55,436,196
Sh. Female	.413	.492	0	0	1	55,436,196
Sh. High educated	.105	.307	0	0	0	54,197,088
Sh. Managers	.0562	.23	0	0	0	49,044,808
Sh. Temp. contracts	.283	.451	0	0	1	36,586,988
Sh. Part-time	.129	.335	0	0	0	55,427,944
Tenure	7.83	8.66	5	1	12	55,436,196
Yearly wage	11,985	10,046	9,100	6,356	14,440	55,436,196
Firm size	1,164	3,533	59	12	417	55,436,196
Num. jobs	1.02	.404	1	1	1	55,436,104

The table reports descriptive statistics for workers in the sample, covering all years from 1985 to 2018. Wages are deflated by the 2010 CPI. The detail on temporary vs. permanent contract is only available from 2000 onwards.

Figure C.1: Average Life-Cycle Employment and Sales by Number of Entrepreneurs

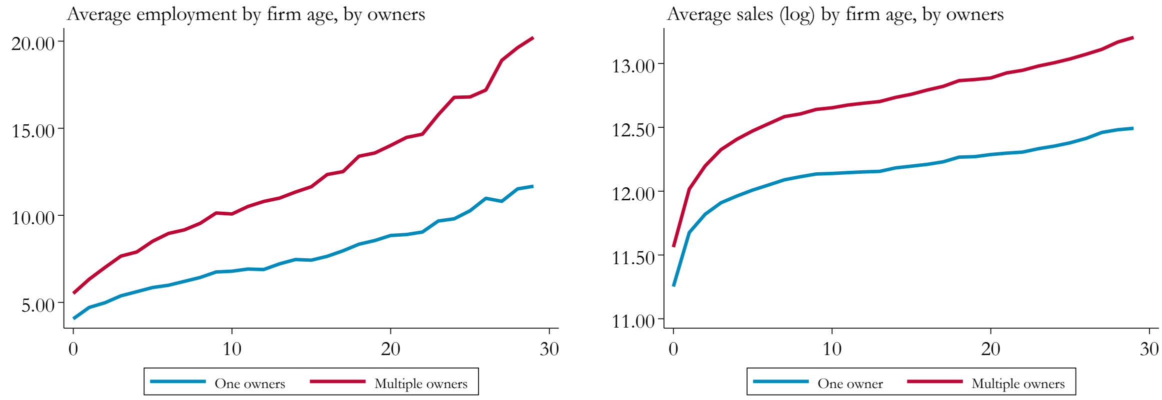


Figure C.2: Life-Cycle Employment and Sales Growth by Number of Entrepreneurs

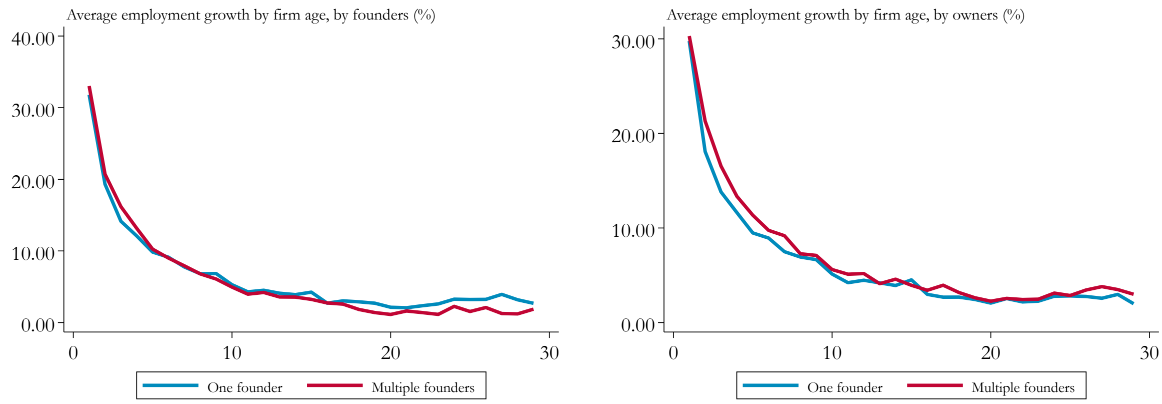


Figure C.3: Average Life-Cycle Labor Productivity by Number of Entrepreneurs

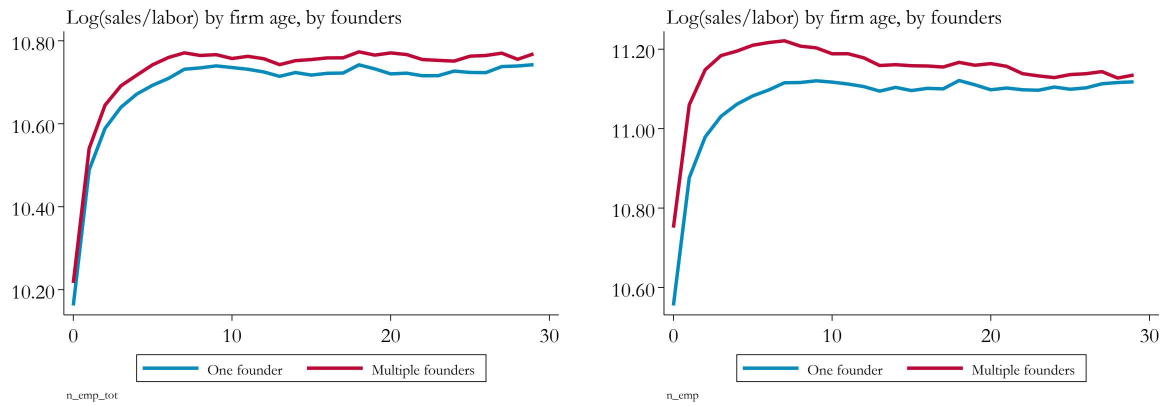


Figure C.4: Exit Rates by Number of Entrepreneurs

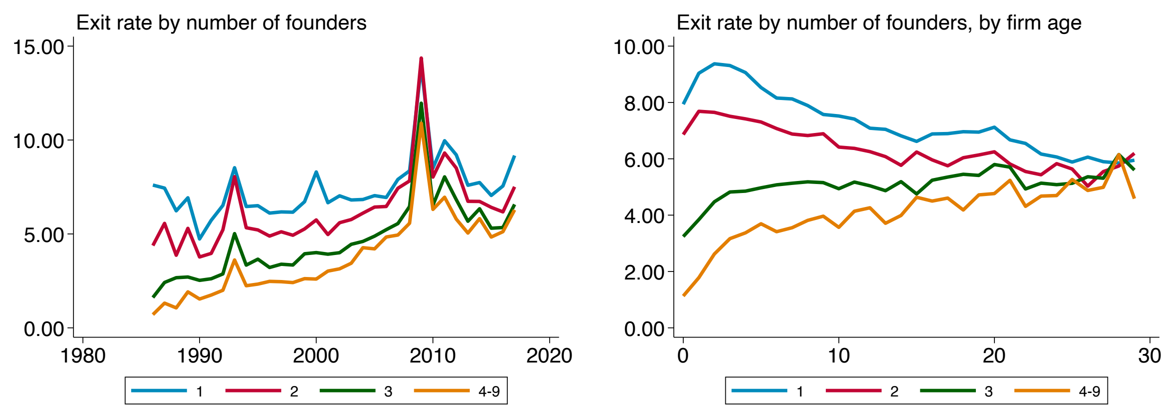


Figure C.5: Exit Rates by Number of Entrepreneurs

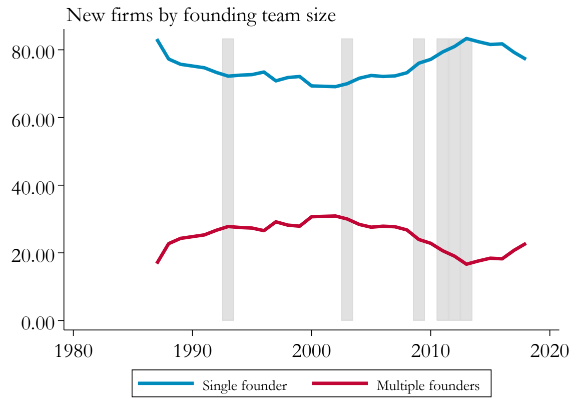
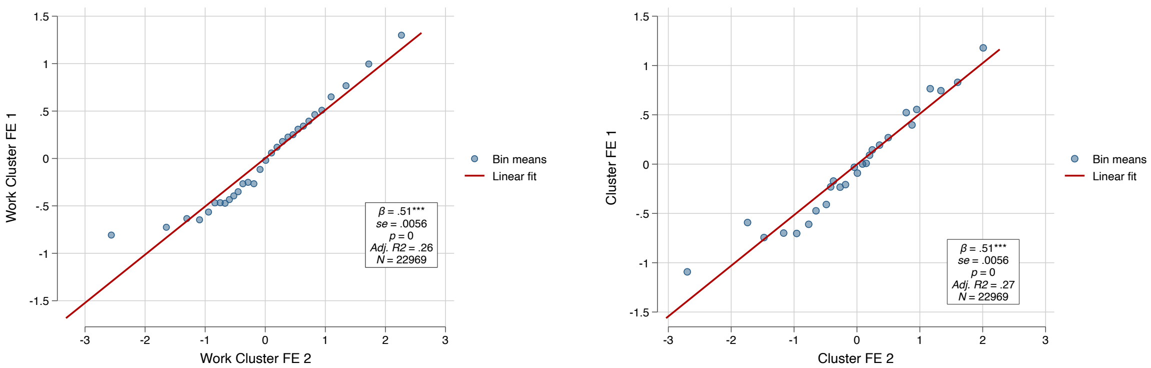
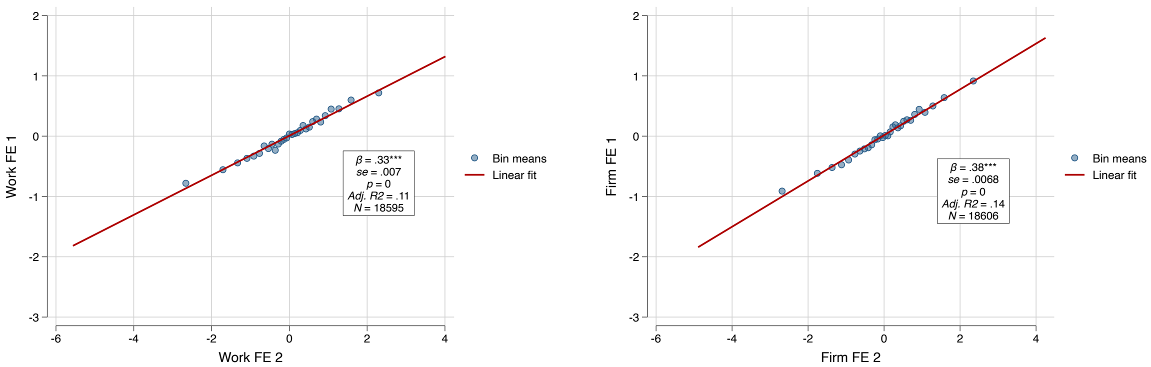


Figure C.6: Correlation of Worker and Past Workplace Types for Entrepreneurial Teams, clustered workplaces



The figures present binned scatterplots of standardized individuals' (left) and workplaces (right) AKM fixed effects for entrepreneurs in two-member teams. For this estimation a K-means clustering is used to identify 10 clusters of firms types, based on the empirical cumulative distribution functions of earnings within firms, pooled across all years, as in Bonhomme, Lamadon and Manresa (2019*a*). Fixed effects are estimated for every year on a 5 years backward looking rolling window. The fixed effects come from the last year before the first entrepreneurial spell.

Figure C.7: Correlation of Worker and Past Workplace Types for Entrepreneurial Teams, residualized effects



The figures present binned scatterplots of standardized individuals' (left) and workplaces (right) AKM fixed effects for entrepreneurs in two-member teams. Fixed effects are estimated for every year on a 5 years backward looking rolling window. We plot residuals obtained by regressing fixed effects on age, year, gender, college education, dummies for same sector, profession, qualification, earnings quintiles, firm size, being a "sequential" entrepreneur and being colleagues, for both members of the teams. The fixed effects come from the last year before the first entrepreneurial spell.

Figure C.8: Full Distributions of Skills in Two-members Teams

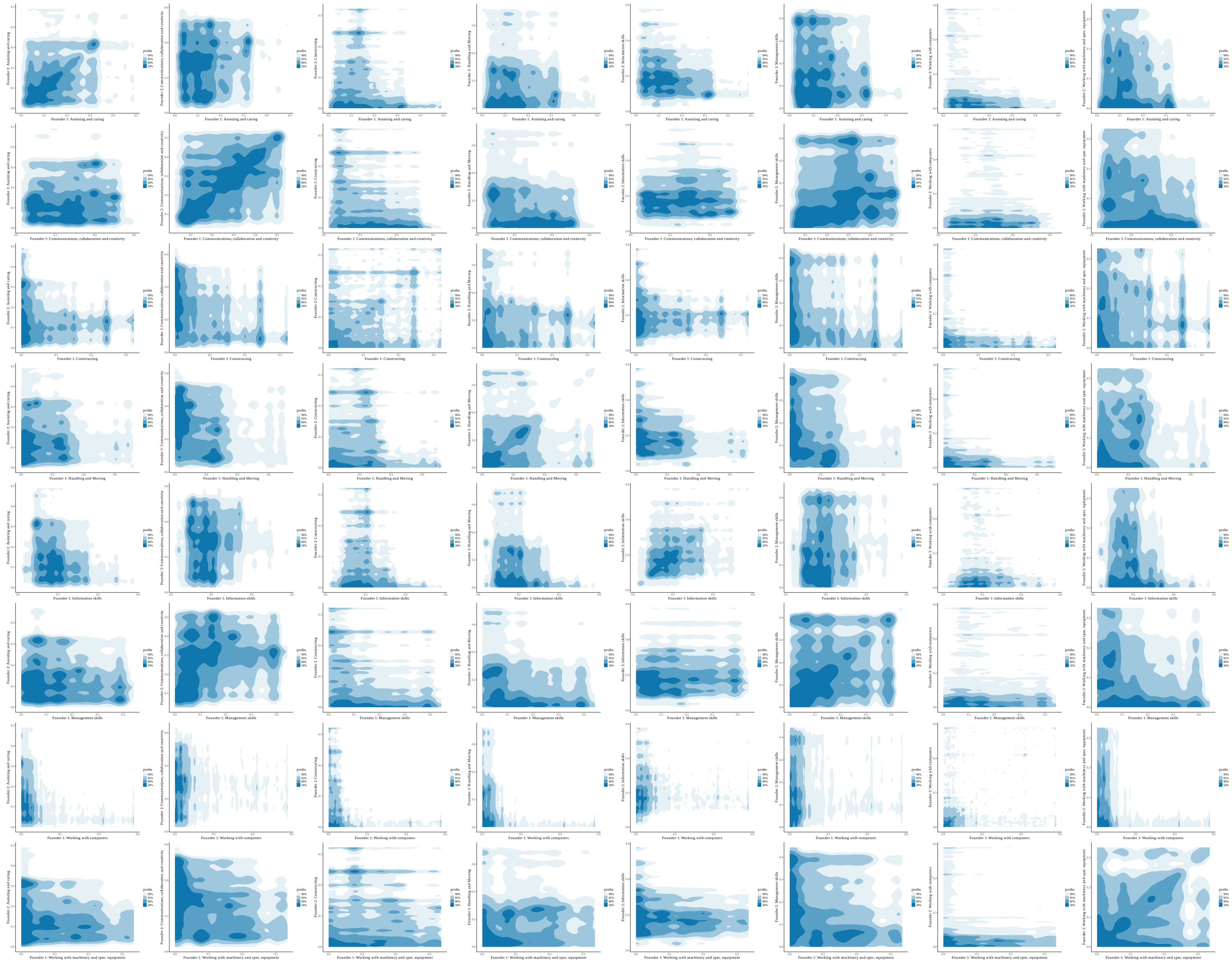


Table C.2: Founding team characteristics and firm performance

	Log-Sales		
	(1)	(2)	(3)
	log_sales_SCIE_w	log_sales_SCIE_w	log_sales_SCIE_w
	b/se	b/se	b/se
Worker FE	0.180***	0.195***	0.127***
	(0.010)	(0.007)	(0.005)
Firm FE	0.125***	0.151***	0.161***
	(0.012)	(0.008)	(0.005)
SD Worker FE	-0.031**	-0.007	-0.003
	(0.014)	(0.011)	(0.008)
SD Firm FE	-0.021	-0.034***	-0.024***
	(0.016)	(0.012)	(0.008)
Initial total assets	0.450***	0.436***	0.480***
	(0.006)	(0.004)	(0.003)
Share of college grad.	0.180***	0.263***	0.338***
	(0.021)	(0.016)	(0.012)
Share of Women	-0.303***	-0.362***	-0.296***
	(0.023)	(0.017)	(0.011)
Observations	22,833	41,982	94,321
Number of firms	8,101	11,715	17,391
R-squared	0.326	0.324	0.390

Table C.3: Changes in Team Composition and Firm Performance

	(1)	(2)	(3)	(4)	(5)	(6)
	Log sales changes	Log sales changes	Log sales changes	Log sales changes	Log sales changes	Log sales changes
Team Member Leaves	-0.0702***	-0.0512***	-0.0428***	-0.0434***	-0.0430***	-0.0576***
	(-10.35)	(-8.07)	(-6.90)	(-6.97)	(-6.94)	(-7.44)
Number of employees	0.0000701*	-0.00111***	-0.00108***	-0.00107***	-0.00104***	-0.00101***
	(2.45)	(-8.54)	(-8.49)	(-8.51)	(-8.38)	(-5.53)
Implied age of the firm	-0.00387***	-0.0112***	0	0	0	0
	(-29.52)	(-12.46)	(.)	(.)	(.)	(.)
Firm FE		Y	Y	Y	Y	Y
Year FE			Y	Y	Y	Y
Sector FE			Y	Y	Y	Y
Nat-Ju FE				Y	Y	Y
Geo FE				Y	Y	Y
Under 55						Y
R-squared	0.00758	0.131	0.139	0.139	0.142	0.156
Observations	2253730	2194457	2194457	2194457	2194455	740007

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333. ISBN: 00129682.

Andrews, Martyn J, Len Gill, Thorsten Schank, and Richard Upward. 2008. "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society Series A: Statistics in Society*, 171(3): 673–697.

Bonhomme, Stephane, Thibaut Lamadon, and Elena Manresa. 2019*a*. "A Distributional Framework for Matched Employer Employee Data." *Econometrica*, , (87): 1–71.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019*b*. "A distributional framework for matched employer employee data." *Econometrica*, 87(3): 699–739.

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.

Lloyd, Stuart. 1982. "Least squares quantization in PCM." *IEEE transactions on information theory*, 28(2): 129–137.

MacQueen, James, et al. 1967. "Some methods for classification and analysis of multivariate observations." Vol. 1, 281–297, Oakland, CA, USA.

Mion, Giordano, Luca David Opromolla, and Alessandro Sforza. 2022. "The Value of Managers' Export Experience: Lessons from the Angolan Civil War." *The Review of Economics and Statistics*, 1–26.