Gender differences in entrepreneurship: the role of family*

Giovanna Vallanti† Belén Rodríguez Moro‡

April 15, 2026

Preliminary work – Please, do not circulate

Abstract

This paper examines gender differences in the intergenerational transmission of entrepreneurship among Ph.D. graduates in Italy. Using administrative survey data, we show that entrepreneurial parents substantially increase the likelihood of entrepreneurial entry, but this effect is significantly weaker for women—especially for daughters of entrepreneurial mothers. We document that a nontrivial share of the gender gap in entrepreneurship reflects differential returns to similar family backgrounds. Exploiting province-level variation in gender norms, proxied by historical voting behavior in the 1981 abortion referendum, we find that transmission to daughters is further attenuated in more traditional environments. The results highlight how family background and local norms jointly shape gender disparities in high-skilled entrepreneurship.

Keywords: Economics of Gender, Entrepreneurship, Occupational and Intergenerational Mobility

JEL Codes: J16, L26, J62

*We are grateful to the participants of GRAPE 2025 and of the workshop on 2026 Gender Gaps in Innovation and Entrepreneurship at the University of Foggia for helpful comments and suggestions.

†University Luiss Guido Carli,gvallanti@luiss.it

‡European University Institute, belen.rodriguez@eui.eu

1 Introduction

Women remain significantly underrepresented in entrepreneurship across advanced economies, with gaps that are especially pronounced in capital-intensive and high-growth sectors (Halabisky and Shymanski, 2023). These disparities carry both equity and efficiency implications: when talent is systematically misallocated across occupations along gender lines, aggregate productivity suffers (Hsieh et al., 2019). In response, understanding the sources of gender gaps in entrepreneurial entry has become a central question for both researchers and policymakers. A large literature points to differences in access to finance (Alesina et al., 2013; Guzman and Kacperczyk, 2019), networks (Howell and Nanda, 2024), and local role models (Bell et al., 2019) as proximate barriers. This paper shifts attention to a closer institution: the family.

Intergenerational persistence in self-employment and entrepreneurship is well documented (Lindquist et al., 2015). Parental entrepreneurship may transmit financial capital (Dunn and Holtz-Eakin, 2000), business networks (Bosma et al., 2012), risk preferences (Dohmen et al., 2012), and direct exposure to entrepreneurial environments (Bell et al., 2019). Yet this transmission need not operate symmetrically across sons and daughters. If family background also conveys gendered expectations about entry costs into male-dominated occupations, or if the occupational signals embedded in parental entrepreneurship are received differently by gender, then standard role-model predictions—which emphasize stronger same-sex transmission—may fail. We examine this question in a setting that holds human capital constant.

We examine whether the intergenerational transmission of entrepreneurship differs between sons and daughters in a high-skilled population. We study Ph.D. graduates in Italy using administrative survey data from the *Inserimento Professionale dei Dottori di Ricerca*, which tracks labor market outcomes five years after completion. This setting is well suited to isolating family-background effects: individuals share high levels of human capital and career orientation, yet substantial gender gaps remain in entrepreneurial occupational choice (Muscio and Vallanti, 2024). Entrepreneurship in this context reflects differential career selection rather than educational sorting.

Our baseline results reveal strong intergenerational persistence in entrepreneurship overall. Sons of entrepreneurial parents are substantially more likely to become entrepreneurs than sons of non-entrepreneurs. For daughters, this association is markedly weaker, and the asymmetry in intergenerational transmission accounts for a meaningful share of the observed gender gap in entrepreneurial entry. When we separate transmission by the parent's sex, a striking pattern emerges: having an entrepreneurial mother raises sons' entrepreneurial entry but yields little or no corresponding effect for daughters. This finding directly contradicts the prediction that same-sex transmission should be strongest, a pattern documented in other intergenerational contexts (Lindquist et al., 2015; Olivetti et al., 2013).

A natural interpretation of the asymmetric transmission pattern is that daughters fail to acquire entrepreneurial skills from their parents. investigate this mechanism directly. Building on the view of entrepreneurship as a bundle of transferable competencies—such as opportunity recognition, resource mobilization, and adaptive problem-solving—rather than a discrete occupational status, we construct an occupation-level measure of entrepreneurial skill intensity. To do so, we match the full universe of skill descriptors in the European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy to the OECD entrepreneurial competency framework. For each ESCO skill descriptor, we compute semantic similarity scores against the set of competencies defined by the OECD framework using sentence-level language model embeddings and classify each skill as entrepreneurial or non-entrepreneurial. We then aggregate these skill-level indicators to the occupation level, weighting by the number of skills associated with each occupation, to obtain continuous and binary measures of entrepreneurial skill content. This procedure allows us to examine the intergenerational transmission of entrepreneurial competencies independently of occupational status.

Because implementation at the three-digit ISCO-08 level requires linkage to restricted administrative records, estimates based on the most granular occupational classification must be produced within ISTAT's secure research environment. These results will be incorporated in a subsequent version of the paper.

We further examine whether the estimated gender differential varies with local gender norms. Using province-level voting outcomes from the 1981 Italian referendum on abortion restrictions as a proxy for local gender attitudes, we assess whether the intergenerational transmission of occupational entrepreneurship differs across normative environments. As with the detailed occupational codes, province identifiers for all waves are only available within ISTAT's secure research environment. Consequently, we currently present results based on the final wave only, which substantially reduces statistical power. Incorporating the full sample is left to future work.

Literature review. To be completed.

2 Data and Sample

2.1 Data Sources

Ph.D. survey Our analysis draws on individual-level data from the *Inserimento Professionale dei Dottori di Ricerca* (IPDR) survey, administered by CINECA in collaboration with Italian universities and the Ministry of University and Research (MUR). We pool the 2014 and 2018 survey waves, which follow Ph.D. graduates from the 2008, 2010, 2012, and 2014 cohorts five years after degree completion. This five-year horizon is long enough to capture entry into entrepreneurship as a deliberate career choice rather than an immediate post-graduation response, and short enough to attribute occupational outcomes to individual decisions and family background rather than late-career dynamics.

The IPDR combines rich information on early-career labor market outcomes with detailed demographic, educational, and family background characteristics. In addition to gender, age, nationality, and region of current residence, the survey records field of study, institution of graduation, Ph.D. duration, and teaching activities during doctoral training. Crucially for our purposes, it collects retrospective information on parents' occupations, which allows us to measure parental entrepreneurial status and study intergenerational transmission within a highly selected and observably similar population.

Labour Force Survey To provide population-level benchmarks and broaden the external validity of our findings, we complement the IPDR data with microdata from the Italian Labour Force Survey (*Rilevazione sulle Forze di Lavoro*, RCFL), conducted quarterly by ISTAT. The RCFL covers a representative sample of the Italian resident population aged 15 and over and collects harmonised information on employment status, occupation (coded at the three-digit ISCO level), industry, and contractual arrangements, as well as individual characteristics including gender, age, educational attainment, and household composition. We use the annual cross-sectional files corresponding to the survey years 2014 and 2018 to construct occupational reference distributions and to assess whether the entrepreneurial patterns we document among Ph.D. graduates differ meaningfully from those observed in the broader highly educated workforce.

Entrepreneurial skills. To characterise the entrepreneurial content of occupations, we construct an occupation-level entrepreneurial skills index by combining data from the ESCO (European Skills, Competences, Qualifications and Occupations) taxonomy with the EntreComp Entrepreneurial Competency Framework. ESCO, maintained by the European Commission, provides a structured multilingual classification of occupational skills and competencies; in the version we use (v1.1.0), it covers 13,492 individual skill descriptors—short natural-language phrases describing a specific ability, knowledge domain, or behavioural disposition—each linked to occupations at the four-digit ISCO-08 level. To identify which of these skills carry entrepreneurial content, we rely on the EntreComp framework (Margherita et al., 2016), jointly developed by the European Commission and aligned with the OECD entrepreneurial competency literature. EntreComp organises entrepreneurial competencies into three areas—*Ideas & Opportunities*, *Resources*, and *Into Action*—each comprising five competencies, for a total of fifteen core competencies (see Appendix G for the full list). These three areas correspond to the three sub-dimensions of our index: *technical*, *business*, and *personal* entrepreneurial skills.

2.2 Samples

Life Survey Sample

Ph.D. Sample. We select individuals containing information on employment status, and parental employment status, as well as province, bachelor grade

Entrepreneurial activity. Our main outcome is a binary indicator equal to one if an individual is self-employed and engaged in entrepreneurial activity five years after Ph.D. completion. This includes owning a business, working in a family business, or participating in a cooperative. We exclude regulated liberal professions—such as lawyers, private physicians, and architects—from the definition, since entry into these occupations is governed by licensing requirements that are distinct from entrepreneurial entry into market-facing businesses. This definitional choice is conservative: it focuses the analysis on entrepreneurship as a productive and competitive activity rather than as a proxy for self-employment in protected sectors.

Parental entrepreneurship. Parental entrepreneurial status is constructed analogously using retrospective information on parents' occupations. We distinguish four mutually exclusive categories indexed by $p \in \{\text{none, father, mother, both}\}$: no entrepreneurial parent (baseline), entrepreneurial father only, entrepreneurial mother only, and both parents entrepreneurial. This classification mirrors the empirical specifications in Section **??**. Table **??** reports the distribution of own and parental entrepreneurship in the full sample and by gender. Entrepreneurial mothers are relatively rare—2.7% of respondents have an entrepreneurial mother, compared to 17.1% with an entrepreneurial father—reflecting the substantially lower rate of female entrepreneurship in the Italian parent generation. We account for the resulting small cell sizes in estimation and inference throughout.

Other controls. All specifications control for Italian citizenship, undergraduate

Table 1: Entrepreneurship by Parental Entrepreneurial Status

	Parental entrepreneurial status				
	None	Father	Mother	Both parents	Total
Full sample					
Not entrepreneur	18,038	3,086	590	673	22,387
	(78.4%)	(13.4%)	(2.6%)	(2.9%)	(97.2%)
Entrepreneur	447	125	28	32	632
	(1.9%)	(0.5%)	(0.1%)	(0.1%)	(2.7%)
Total	18,485	3,211	618	705	23,019
	(80.3%)	(14.0%)	(2.7%)	(3.1%)	(100%)
Women					
Not entrepreneur	8,931	1,704	314	383	11,332
	(77.2%)	(14.7%)	(2.7%)	(3.3%)	(98.0%)
Entrepreneur	166	46	8	13	233
	(1.4%)	(0.4%)	(0.1%)	(0.1%)	(2.0%)
Total	9,097	1,750	322	396	11,565
	(78.7%)	(15.1%)	(2.8%)	(3.4%)	(100%)
Men					
Not entrepreneur	9,107	1,382	276	290	11,055
	(79.5%)	(12.1%)	(2.4%)	(2.5%)	(96.5%)
Entrepreneur	281	79	20	19	399
	(2.5%)	(0.7%)	(0.2%)	(0.2%)	(3.5%)
Total	9,388	1,461	296	309	11,454
	(82.0%)	(12.8%)	(2.6%)	(2.7%)	(100%)

Notes: The table reports the number of individuals by their own entrepreneurial status and parental entrepreneurial background. Percentages (in parentheses%) refer to the share of the relevant subsample (full sample, women, or men%).

final grade as a proxy for pre-Ph.D. ability, and Ph.D. cohort fixed effects. Richer specifications add scientific field fixed effects and university-province fixed effects. We classify Ph.D. fields into three broad categories: STEM, Social Sciences, and Humanities.[1]

Table 2 reports summary statistics by gender. Employment rates five years after graduation are high and broadly similar for women and men, while women work fewer paid hours and earn substantially less on average—a gap of roughly 390 per month, or approximately 19%. Women are more likely to remain employed in Italy, whereas men display modestly higher international mobility. Family characteristics differ modestly by gender, with women somewhat more likely to be married or to have children at the time of the survey.

2.3 Entrepreneurial occupational measure

We classify occupations according to the share of their associated ESCO skills that are classified as entrepreneurial, aggregated to the three-digit ISCO-08 level. The construction of this index proceeds in two steps: first, we classify individual ESCO skill descriptors as entrepreneurial or non-entrepreneurial; second, we aggregate these classifications to the occupation level.

Classifying skills. Entrepreneurial content is identified by applying a *semantic ensemble classifier* to the natural-language ESCO skill descriptors. Each descriptor and each EntreComp competency are embedded into a common vector space using a sentence-transformer model, and cosine similarity is computed between each skill and each of the three EntreComp dimensions (technical, business, and personal).[2] This approach identifies skills with entrepreneurial *conceptual content* even when they do not use explicit business terminology—for instance, a skill such as "coordinate resource allocation under uncertainty" may align strongly with resource mobilisation and coping with risk despite not mentioning entrepreneurship directly. As a robustness check, we also construct an alternative measure based on keyword matching (**?**), which flags skills containing explicitly entrepreneurial language derived from the EntreComp framework; results are qualitatively unchanged across

[1] STEM includes Mathematics and Computer Sciences, Physical Sciences, Chemical Sciences, Earth Sciences, Biological Sciences, Agricultural and Veterinary Sciences, Civil Engineering and Architecture, and Industrial and Information Engineering. Social Sciences include Economics and Statistics, Political and Social Sciences, and Legal Sciences. Humanities include Ancient, Philological, and Literary Sciences, as well as Historical, Philosophical, and Educational Sciences. Medical Sciences are excluded, as Ph.D. candidates in these fields are typically employed in clinical or hospital settings during training, leading to career trajectories that are structurally distinct from those of research-track graduates.

[2] We use the `all-mpnet-base-v2` sentence-transformer model (Reimers and Gurevych, 2019).

Table 2: Sociodemographic Characteristics of Ph.D. Graduates by Gender

	Ph.D. Graduates by Gender	
	Women	Men
Entrepreneurs	2.01	3.46
Employment		
Employed (5 years after Ph.D.)	79.8	81.1
	(0.401)	(0.392)
Weekly Paid Hours	35.2	39.5
	(13.0)	(12.8)
Work in Italy	88.7	81.8
	(31.6)	(38.6)
Monthly Earnings (€)	1,694	2,082
	(817)	(1,082)
Family Background and Others		
Married	46.9	41.6
	(49.9)	(49.3)
Has Children	38.8	32.9
	(48.7)	(47.0)
Father with College Degree	32.7	33.4
	(46.9)	(47.2)
Mother with College Degree	29.7	29.1
	(45.7)	(45.4)
Italian Citizenship	97.0	96.5
	(14.0)	(18.3)
Obs.	11,565	11,454

Notes: This table presents sociodemographic characteristics of Ph.D. graduates by gender. Columns report statistics for women and men separately. Standard deviations are provided in parentheses. Data is pooled from Ph.D. cohorts of 2008, 2010, 2012, and 2014.

the two classifiers (see Appendix G.3 for a full description of both methods and their cross-tabulation).

To give a sense of what the classifier captures, Table G.3 in Appendix G.4 reports examples of skills at different points of the entrepreneurial score distribution using the ensemble method. Among skills classified as entrepreneurial, one finds descriptors such as “identify business opportunities”, “manage financial resources”, and “develop a business plan”—skills that map directly onto the EntreComp competencies of opportunity recognition, financial literacy, and planning and management. At the other end of the distribution, skills such as “operate cutting machinery”, “apply hygiene standards”, and “monitor production equipment” score near zero, reflecting the absence of any conceptual or terminological overlap with the entrepreneurial competency framework.

Aggregating to occupations. For each three-digit ISCO-08 occupation, we compute the entrepreneurial skill share as the proportion of its associated ESCO skills—both essential and optional—that are classified as entrepreneurial at a given threshold. This normalisation by the total number of skills per occupation ensures that the index reflects the *relative* entrepreneurial content of an occupation rather than the sheer number of skills listed for it in ESCO. The working dataset covers 129 Italian three-digit occupational categories, spanning 89 distinct ISCO-08 three-digit groups.

The face validity of the resulting measure is confirmed by Table 3, which reports the top and bottom occupations by entrepreneurial skill intensity. Occupations at the top of the distribution—Finance Professionals (ISCO 241), Sales, marketing and development managers (ISCO 122), and managing directors and chief executives (ISCO 112)—are precisely those one would expect to demand entrepreneurial competencies most intensively. Conversely, manufacturing labourers (ISCO 932), heavy vehicle drivers (ISCO 833), and mining and construction labourers (ISCO 931) appear at the bottom, consistent with the broader literature on task content and occupational complexity (Autor and Dorn, 2013). Occupations around the median of the distribution—mid-level technicians, laboratory analysts, and social workers—plausibly require some degree of self-direction and resource management without being primarily entrepreneurial in character.

3 Institutional Context: Labour Outcomes and Parental Background of Ph.D. Graduates in Italy

3.1 Post-graduate entrepreneurial activity in Italy

Table 4 documents aggregate trends in Italian entrepreneurship by gender and education using LFS data for 2014–2018. The gender gap in entrepreneurship is large and persistent: the male rate is consistently more than double the female rate throughout the period. Strikingly, the share of entrepreneurs holding a Master's or PhD degree is markedly higher among women than among men, and widens over the period to reach 11.4 percentage points by 2018—implying that highly educated workers account for a disproportionate and growing share of female entrepreneurship in Italy.

Table 5 compares entrepreneurial rates between PhD graduates and the broader high-skill workforce. PhD graduates enter entrepreneurship at rates three to four times higher than comparably educated LFS workers, confirming that doctoral training—with its combination of research autonomy and knowledge-transfer exposure—

Table 3: Top and Bottom 10 Occupations by Entrepreneurial Skill Intensity: Ensemble method

Rank	**ISCO-3**	**Occupation**	**Score**
Top 10 — highest ensemble score			
1	241	Finance Professionals	1.000
2	122	Sales, Marketing and Development Managers	0.825
3	112	Managing Directors and Chief Executives	0.817
4	133	Information and Communications Technology Service managers	0.813
5	331	Financial and Mathematical Associate Professionals	0.811
6	243	Sales, Marketing and Public Relations Professionals	0.705
7	121	Business Services and Administration Managers	0.704
8	332	Sales and Purchasing Agents and Brokers	0.640
9	252	Database and Network Professionals	0.637
10	431	Numerical Clerks	0.630
Bottom 10 — lowest ensemble score			
1	932	Manufacturing Labourers	0.000
2	931	Mining and Construction Labourers	0.040
3	835	Ships' Deck Crews and Related Workers	0.044
4	833	Heavy Truck and Bus Drivers	0.045
5	911	Domestic, Hotel and Office Cleaners and Helpers	0.049
6	814	Rubber, Plastic and Paper Products Machine Operators	0.052
7	811	Mining and Mineral Processing Plant Operators	0.055
8	832	Car, Van and Motorcycle Drivers	0.056
9	834	Mobile Plant Operators	0.064
10	813	Chemical and Photographic Products Plant and Machine Operators	0.065

Notes: Share of occupation skills classified by ensemble model. Occupations at the three-digit ISCO-08 level. Scores aggregated from ESCO v1.1.0 skill-occupation pairs.

selects a distinctly entrepreneurially active population. Together, these patterns establish that PhD graduates represent the most policy-relevant margin of female entrepreneurship: a group with the human capital to sustain innovative ventures, yet one where gender gaps remain large.

Table 4: Entrepreneurship and Postgraduate Education by Gender, Italy (LFS 2014–2018)

	Share of Entrepreneurs (%)			Entrepreneurs with Master/PhD (%)		
Year	Men	Women	Gap (W–M)	Men	Women	Gap (W–M)
2014	8.59	3.71	-4.87	16.58	25.52	8.94
2015	8.48	4.27	-4.21	17.10	27.12	10.03
2016	8.09	3.91	-4.18	16.91	27.30	10.38
2017	7.88	3.56	-4.32	17.43	29.28	11.84
2018	7.83	3.39	-4.44	18.32	29.67	11.35

Notes: Shares computed using the Italian Labour Force Survey (LFS). The first panel reports the share of entrepreneurs in the working-age population by gender. The second panel reports the share of entrepreneurs holding a Master's or PhD degree. The gender gap is defined as women minus men.

Table 5: Entrepreneurial rates: PhD graduates vs high-skill labour force, Italy

	Women		Men	
Year	PhD survey	LFS (high-skill)	PhD survey	LFS (high-skill)
2014	2.88	0.84	4.91	1.21
2018	1.48	0.96	2.60	1.58

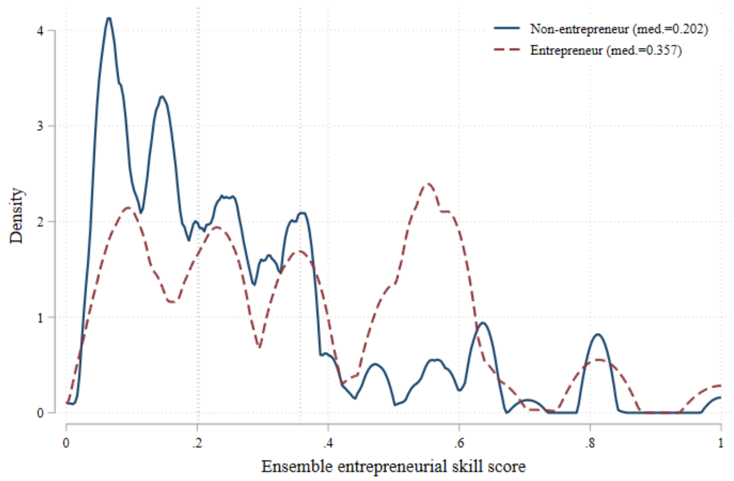
Notes: PhD survey rates are the share of respondents who are entrepreneurs (broad definition: includes autonomous workers, freelancers, and formal entrepreneurs) in each survey wave. LFS rates are the share of self-employed workers among PhD-level employed individuals, from the Italian Labour Force Survey. All rates expressed as percentages.

3.2 Entrepreneurial skill distribution

The binary entrepreneurship indicator captures occupational status but abstracts from the productive content of the activities entrepreneurs perform. We complement it with our occupation-level entrepreneurial skill intensity measure to ask whether entrepreneurs differ systematically from non-entrepreneurs along the skill dimension, and whether high-skill entrepreneurs are qualitatively distinct from the broader self-employed population.

Figure 1 plots the distribution of the ensemble entrepreneurial skill score separately for individuals who are entrepreneurs and those who are not. The two distributions are clearly distinct. Non-entrepreneurs cluster at the left of the skill distribution (median = 0.202), while entrepreneurs exhibit a substantially right-shifted distribution (median = 0.357), with considerable mass at high skill-intensity values. The rightward shift is not driven by a single spike: the entrepreneurial distribution is multi-modal, consistent with entrepreneurs spreading across a variety of high-skill market-facing occupations rather than concentrating in a single sector. This pattern validates the skill measure as a meaningful dimension along which entrepreneurial career choices differ from non-entrepreneurial ones, and motivates its use as an outcome in Section 4.

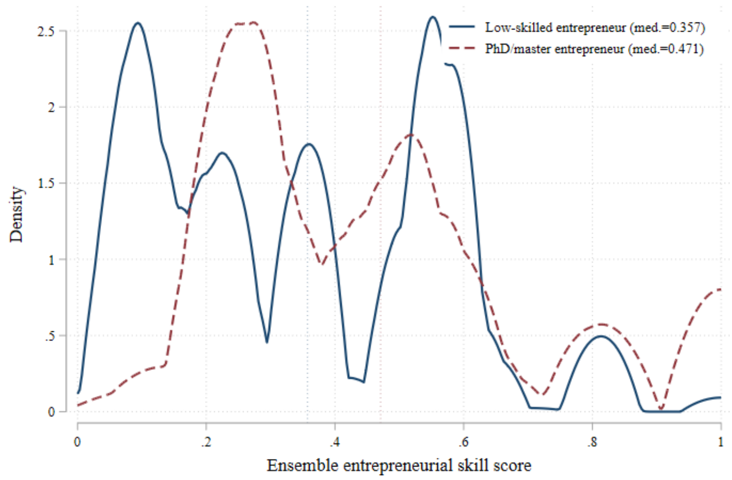
Figure 1: Entrepreneurial occupational distribution: entrepreneurs vs non-entrepreneurs.



Notes: Kernel density estimates of the ensemble entrepreneurial skill score for PhD graduates who are entrepreneurs (dashed, median = 0.357) and non-entrepreneurs (solid, median = 0.202). The score is an occupation-level measure of entrepreneurial skill intensity constructed by matching ESCO skill descriptors to the EntreComp framework via sentence embeddings (Appendix G). Each respondent is assigned the score of their three-digit ISCO-08 occupation five years after PhD completion. IPDR waves 2014 and 2018, pooled.

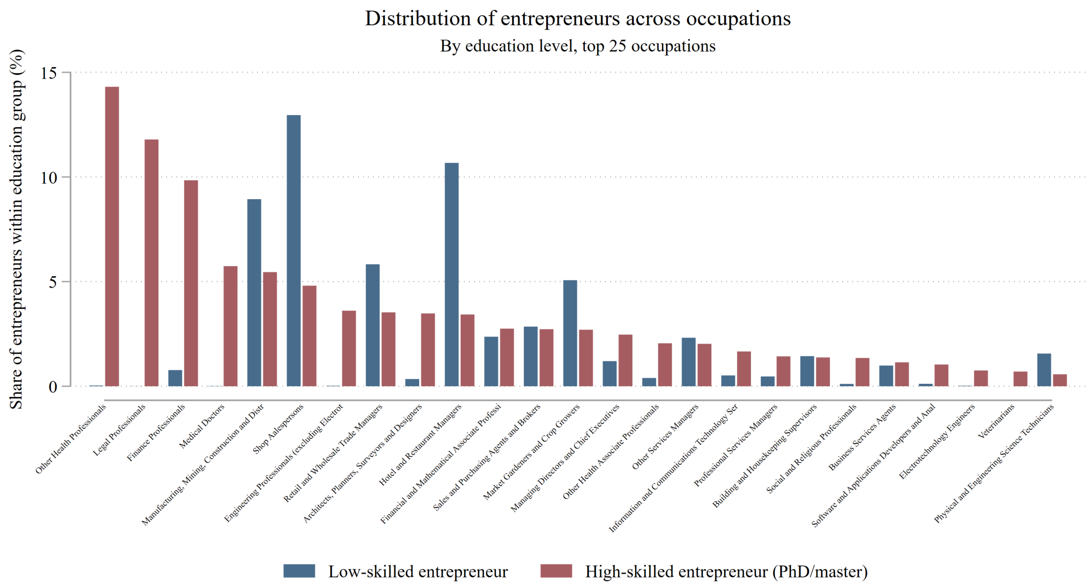
Figure 2 broadens the comparison to the LFS, contrasting the skill distribution of PhD and Master's-level entrepreneurs with that of low-skill entrepreneurs in the broader workforce. The two distributions are clearly separated: high-skill entrepreneurs concentrate at the upper end of the score distribution (median = 0.471), while low-skill entrepreneurs cluster substantially lower (median = 0.357). The gap between these distributions is large relative to the overall spread of the measure, confirming that education reshapes not only the probability of entrepreneurial entry but the qualitative nature of the entrepreneurial activity itself. PhD entrepreneurs occupy occupations with dense entrepreneurial skill content—opportunity recognition, resource mobilisation, strategic planning—while the typical self-employed worker in the broader economy performs activities with far lower entrepreneurial intensity. This distinction justifies treating high-skill entrepreneurship as a substantively separate phenomenon and motivates our focus on the PhD graduate population throughout the analysis.

Figure 2: Entrepreneurial occupational distribution by education level: high-skill vs low-skill entrepreneurs.



Notes: Kernel density estimates of the ensemble entrepreneurial skill score for entrepreneurs holding a post-graduate degree (dashed, median = 0.471) and with lower educational degrees (solid, median = 0.357). The score is an occupation-level measure of entrepreneurial skill intensity constructed by matching ESCO skill descriptors to the EntreComp framework via sentence embeddings (Appendix G). Each respondent is assigned the score of their three-digit ISCO-08 occupation five years after PhD completion. IPDR waves 2014 and 2018, pooled.

Figure 3: Occupational distribution of PhD entrepreneurs.



Notes:

4 Intergenerational Transmission of Entrepreneurial Activity

4.1 Methodology

We estimate the following baseline specification:

$$Y_{i,t,j,s} = \beta_0 + \beta_1 \mathbf{1}\{p_i = a\} + \Gamma X_{i,t} + \gamma_t + \gamma_j + \varepsilon_{i,t,j,s}, \tag{1}$$

where $Y_{i,t,j,s}$ is an indicator equal to one if individual i, who completed a Ph.D. in year t in field s at a university located in province j, is engaged in entrepreneurial activity five years after graduation. The indicator $\mathbf{1}\{p_i = a\}$ captures parental entrepreneurial status, with no entrepreneurial parent as the omitted category. When we examine heterogeneity by the parent's sex, $a \in \{\text{father only, mother only, both parents}\}$.

The vector $X_{i,t}$ includes gender, Italian citizenship, and undergraduate final grade as a proxy for pre-Ph.D. ability, as well as controls for parental education (father and mother college attainment). All specifications include Ph.D. cohort fixed effects γ_t and, in richer specifications, university-province fixed effects γ_j. Standard errors are clustered at the university-province level. We estimate equation (1) by OLS.

To examine whether intergenerational transmission differs by the child's gender, we augment equation (1) with interactions between parental entrepreneurial status and a female indicator. This interaction design allows us to test whether the strength of transmission differs systematically between sons and daughters, and whether any asymmetry varies by the sex of the entrepreneurial parent.

4.2 Empirical results on the transmission of entrepreneurial status

Table 6 reports estimates from the pooled specification. Having at least one entrepreneurial parent increases the probability of entrepreneurial entry by 1.7 to 2.0 percentage points, depending on the specification—an effect that is large relative to the sample mean of 2.7%, implying a near-doubling of the entrepreneurship rate for individuals with entrepreneurial family backgrounds. This estimate is stable across specifications that progressively add parental education controls, scientific field fixed effects, and university-province fixed effects, indicating that intergenerational transmission is not driven by parental human capital or by sorting into particular disciplines and institutions. Female graduates are 1.8 to 1.9 percentage points less likely than male graduates to become entrepreneurs—a gap almost as large as the transmission effect itself.

The interaction of parental entrepreneurship with a female indicator is negative and statistically significant (column 4): daughters experience a transmission effect roughly two-thirds smaller than sons, and this gender asymmetry accounts for a meaningful share of the unconditional gender gap in entrepreneurial entry. Allowing for gender-specific transmission also increases the estimated transmission coefficient for men, implying that pooled estimates attenuate the true strength of intergenerational persistence by averaging across groups with very different returns to the same family background.

Table 6: Intergenerational Transmission of Entrepreneurship

Whether Individual Is an Entrepreneur	(1)	(2)	(3)	(4)
Entrepreneurial Parent	0.017***	0.020***	0.019***	0.030***
	(0.004)	(0.004)	(0.003)	(0.006)
Female	-0.019***	-0.019***	-0.018***	-0.014***
	(0.003)	(0.003)	(0.002)	(0.003)
Parental Education				
Father with College Degree		0.010***	0.011***	0.011***
		(0.003)	(0.002)	(0.002)
Mother with College Degree		0.001	0.002	0.002
		(0.003)	(0.003)	(0.003)
Female × Entrepreneurial Parent				
× Female				-0.020***
				(0.007)
Observations	19,637	19,637	19,616	19,616
Ph.D. Cohort FEs	✓	✓	✓	✓
Scientific Field FEs			✓	✓
University Province FEs				✓

Notes: This table reports OLS estimates of the probability of becoming an entrepreneur five years after Ph.D. completion. *Entrepreneurial Parent* is an indicator equal to one if at least one parent was an entrepreneur (omitted category: no entrepreneurial parent). Columns progressively add controls for parental education, scientific field fixed effects, and university-province fixed effects. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Heterogeneity by parental sex. Table C.1 disaggregates parental entrepreneurship by the sex of the entrepreneurial parent. The results are striking. For male Ph.D. graduates, having an entrepreneurial father increases the probability of entrepreneurial entry by 2.6 percentage points, and having an entrepreneurial mother by 4.3 percentage points—effects that correspond to roughly 75% and 125% of the baseline entrepreneurship rate, respectively. These estimates confirm that parental entrepreneurship is a powerful predictor of entrepreneurial entry among sons, and that maternal entrepreneurship, despite being rarer, carries at least as strong a signal as paternal entrepreneurship for men.

The picture is markedly different for daughters. The female interactions point in a consistent direction: the intergenerational transmission from fathers to daughters is 1.7 percentage points weaker than from fathers to sons (column 4), and the interaction from mothers to daughters is 4.5 percentage points weaker—a point estimate that largely offsets the large positive effect of maternal entrepreneurship estimated for sons, leaving the net effect for daughters close to zero. These patterns are difficult to reconcile with the prediction that same-sex transmission should be the dominant mode of intergenerational persistence, and the point estimates suggest that entrepreneurial mothers do not generate the same role-model effect for daughters as they do for sons. We note, however, that entrepreneurial mothers are rare in our sample (2.7% of respondents), so the mother-specific estimates carry wider confidence intervals than the father-specific results; the near-zero effect for daughters should be read as an imprecisely estimated null rather than as precise evidence of no effect.

These findings differ from Lindquist et al. (2015), who document strong same-sex transmission in the general Swedish population. As we discuss in Section **??**, this divergence likely reflects the particular population and margin we study: among high-skill graduates making a rare career choice, entrepreneurial mothers may transmit information about the social and economic costs of entering gender-atypical environments rather than acting as straightforward occupational role models.

4.3 Transmission of entrepreneurial competencies and innovation ability.

A natural concern is that the gender asymmetry in entrepreneurial transmission reflects a more general difference in how productive human capital—technical skills, creativity, innovative potential—is transmitted from parents to children. We examine this through two complementary tests. Our primary test exploits a novel occupation-level measure of entrepreneurial skill intensity that we construct by matching the full universe of skill descriptors in the European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy to the competency framework defined by the OECD. For each ESCO descriptor, we compute semantic similarity scores against the set of OECD entrepreneurial competencies using sentence-level language model embeddings, and aggregate these skill-level indicators to the occupation level, weighting by the number of skills associated with each occupation, to obtain continuous and binary measures of entrepreneurial skill content. Assigning respondents their occupation's score based on three-digit ISCO-08 codes allows us to examine whether daughters of entrepreneurial parents acquire entrepreneurial competencies at rates comparable to sons, and whether any observed asymmetry arises instead at the stage of occupational translation. Because this analysis requires linkage to restricted administrative records at the three-digit ISCO-08 level, estimates will be produced within ISTAT's secure research environment and incorporated in a subsequent version of the paper.

As a complementary test available with the current data, we use patenting as an indicator of innovative capacity that is distinct from occupational choice. Table **??** reports results using an indicator for whether the respondent has at least one registered patent application within five years of Ph.D. completion. Having an entrepreneurial father is associated with a higher probability of patenting—an effect of 1.1 to 1.4 percentage points that is marginally significant and consistent with the transmission of technical exposure and networks. Maternal entrepreneurship has no statistically significant effect on patenting for either sons or daughters. Crucially, the interaction between female gender and parental entrepreneurship is small and statistically insignificant across all specifications: the gender asymmetry documented for entrepreneurial entry simply does not appear for patenting.

This contrast already sharpens the interpretation of our main results. If daughters failed to inherit entrepreneurial competencies from their parents, we would expect the gender asymmetry to extend to patenting outcomes. Its absence is consistent with a mechanism in which daughters acquire the relevant abilities from entrepreneurial family backgrounds but are less likely to convert those competencies into entrepreneurial careers—a hypothesis the forthcoming skill-content analysis is designed to test directly.

Table 7: Intergenerational Transmission of Innovation: Patenting after the Ph.D.

	(1)	(2)	(3)	(4)
	Whether Individual Has at Least One Patent			
Entrepreneurial Parent	0.008**	0.008*	0.008	0.010
	(0.004)	(0.004)	(0.006)	(0.008)
Female	-0.035***	-0.035***	-0.028***	-0.027***
	(0.003)	(0.003)	(0.004)	(0.004)
Parental Education				
Father with College Degree		-0.004	0.001	0.001
		(0.004)	(0.004)	(0.004)
Mother with College Degree		0.003	0.002	0.002
		(0.004)	(0.003)	(0.003)
Female × Entrepreneurial Parent				
× Female				-0.002
				(0.008)
Observations	19,637	19,637	19,616	19,616
Ph.D. Cohort FEs	✓	✓	✓	✓
Scientific Field FEs			✓	✓
University Province FEs				✓

Notes: This table reports OLS estimates of the probability of having at least one patent application within five years after Ph.D. completion. *Entrepreneurial Parent* is an indicator equal to one if at least one parent was an entrepreneur (omitted category: no entrepreneurial parent). Columns progressively add controls for parental education, scientific field fixed effects, and university-province fixed effects. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Finally, Table E.3 in the Appendix reports OLS estimates of the time elapsed between PhD graduation and the first post-doctoral job, regressing it on an entrepreneur indicator, a female indicator, and their interaction. A positive and significant coefficient on $Entrepreneur \times Female$ would indicate that female entrepreneurs take systematically longer to secure their first position than male entrepreneurs with otherwise similar backgrounds — consistent with entrepreneurship operating as a fallback option for women rather than a primary career choice. In the pooled sample, this is precisely what we find: female entrepreneurs take roughly half a year longer than their male counterparts to find a first job (column 1), and this gap nearly doubles among individuals with at least one entrepreneurial parent (column 2). The pattern is concentrated in Social Sciences and Humanities and largely absent in STEM.

4.4 Robustness.

The results of the previous estimations show the existence of gender differences in the intergenerational transmission of entrepreneurship from parents to their offspring among highly-skilled individuals. However, several potential factors could confound these results. Doing a Ph.D. and becoming an entrepreneur are two endogeneous decisions that generate substantial selection within our sub-sample of entrepreneur Ph.D. graduates. In Section 3 we document gender differences in Ph.D. graduates and entrepreneurial rates across fields of study. To the extent that these gender differences in self-selection into the Ph.D. and into entrepreneurship are correlated with the gender differential effect in the transmission of entrepreneurship, we cannot interpret the intergenerational estimates as causal. In this section we tackle different potential confounding issues related to the decision of doing a Ph.D. and to the later decision of becoming an entrepreneur.

Gender differences in Ph.D. self-selection Women and men with entrepreneurial parents who choose to pursue a PhD may already be a non-random sub-sample. But for this to bias $\hat{\beta}_3$, the *direction and magnitude* of this selection must differ by gender, i.e. $\delta_3 \neq 0$ in (5) for pre-determined observables.

Table E.1 reports $\hat{\delta}_1$, $\hat{\delta}_2$, and $\hat{\delta}_3$ for each of eight pre-determined characteristics: nationality, PhD funding status (grant and public grant), undergraduate performance (honours degree), geographic origin (born in South), and location of PhD institution (South/North). In no case is $\hat{\delta}_3$ statistically significant at conventional levels, and the estimates are economically small throughout. Parental entrepreneurship does create some degree of selection ($\hat{\delta}_2$ is non-zero for some variables), but this selection is equally present for women and men. The joint test of $H_0 : \delta_3 = 0$ across all balance variables cannot be rejected ($p =$ *BR: add p-value*). We therefore find no evidence that whatever positive or negative selection into the PhD that entrepreneurial family background induces is operating differentially across genders.

Gender differences in entrepreneurial self-selection. Although the relevance of accumulating high-skilled human capital is increasing for entrepreneurs, there could be still problems if entrepreneur is seen as a second employment choice for Ph.D. graduates. In E.4 we check whether the job finding rates of women and men after graduation differ by computing the number of years from graduation until individuals start their first job after the Ph.D. We observe that entrepreneur women take longer time to find their first job relative to men, and that this greater among the subsample of individuals with entrepreneurial parents. This suggest that women may see entrepreneur as a second option or that they set up busisness in different induestries that those of their parents relatve to male entrepreneurs.

Transmission of self-employment. We also plan to check the transmission of self-employment to compare our results with Lindquist et al. (2015).

5 Mechanism: Interaction Between Family and Gender Norms

The results in Section 4 establish two facts. First, intergenerational transmission of entrepreneurship is substantially weaker for daughters than for sons, and the attenuation is largest—and statistically complete—for mother-to-daughter transmission. Second, this asymmetry does not extend to the transmission of patenting activity, suggesting that daughters of entrepreneurial parents acquire entrepreneurial competencies but are less likely to act on them occupationally. A natural question is whether this gap in occupational conversion varies across social environments. If gender norms shape the perceived feasibility, social costs, or expected returns of entrepreneurship for women, then intergenerational transmission to daughters should be more attenuated in environments where such norms are more restrictive.

Motivated by this reasoning, we examine whether the gender asymmetry in occupational transmission varies systematically with local attitudes toward women's economic roles. We do not interpret this analysis as identifying a specific intrahousehold mechanism. Rather, local gender norms serve as a conditioning variable that allows us to test whether the mapping from entrepreneurial family background to entrepreneurial entry is sensitive to the normative environment in which graduates were early socialized (Bisin and Verdier, 2001; Fernández et al., 2004; **?**).

5.1 Methodology: province-level heterogeneity in gender norms

We exploit historical variation in gender attitudes using province-level voting outcomes from the 1981 Italian referendum on abortion rights. The referendum proposed restrictions on the law that had legalized abortion in 1978; we use the share of votes cast *against* the restriction proposal as a measure of gender-progressive attitudes. Provinces in the bottom quartile of this distribution—those where support for restrictions was highest—are classified as traditional; provinces in the top quartile are classified as progressive. This historical measure captures deep-rooted differences in gender norms that predate the career formation of the graduates in our sample by at least a generation, alleviating concerns about reverse causality.

We estimate the following reduced-form specification:

$$\begin{aligned} Y_{i,t,j,s} = {} & \beta_0 + \beta_1 \mathbf{1}\{p_i = a\} + \theta\,\mathrm{Female}_i + \alpha\,\mathbf{1}\{p_i = a\} \times \mathrm{Female}_i \\ & + \lambda\,\mathrm{Traditional}_j + \delta_a\,\mathbf{1}\{p_i = a\} \times \mathrm{Female}_i \times \mathrm{Traditional}_j + \Gamma X_{i,t} + \gamma_t + \gamma_j + u_{i,t,j,s}, \end{aligned} \tag{2}$$

where all variables are defined as in equation (1). The province-level indicator

Traditional_j equals one for provinces in the bottom quartile of the referendum distribution. The coefficients δ_a test whether the gender asymmetry in intergenerational transmission is more pronounced in traditional environments, conditional on university-province fixed effects and the full set of controls. The sample is restricted to respondents for whom province of birth is observed, which reduces the estimation sample but preserves the key identifying variation.

5.2 Results

Parental gender heterogeneity Table 8 reports the estimates. Column (1) replicates the baseline gender-interaction specification from Table C.1 in the restricted sample, confirming that the core patterns from Section **??** are present: parental entrepreneurship strongly predicts entrepreneurial entry for men, while transmission to daughters is substantially attenuated, particularly from entrepreneurial mothers. The coefficient on the traditional province indicator is negative and significant: graduates from traditional provinces are about 1.2 percentage points less likely to become entrepreneurs, a meaningful effect relative to the 2.7% sample mean.

Column (2) adds the triple interaction between gender, parental entrepreneurship, and traditional norms. The estimates indicate that the gender asymmetry in intergenerational transmission is further attenuated in more traditional environments. For daughters with two entrepreneurial parents—the group with the strongest exposure to entrepreneurial family environments—the triple interaction coefficient is -0.093 (standard error 0.068). While imprecisely estimated, reflecting the limited number of observations in traditional provinces with both parents entrepreneurial, the point estimate implies a near-complete elimination of any intergenerational transmission advantage for daughters from highly entrepreneurial families in traditional environments. The interaction with the female indicator alone is small and statistically insignificant, suggesting that traditional norms do not operate simply by depressing women's entrepreneurial entry uniformly—rather, they specifically weaken the conversion of entrepreneurial family capital into entrepreneurial careers.

These patterns are consistent with two complementary interpretations. First, in environments with stronger normative constraints on women's economic roles, daughters may form lower expectations about the feasibility of entrepreneurship as a career, even when they grow up in households with direct exposure to entrepreneurial activity (Fernández, 2013). Second, entrepreneurial parents in traditional environments may provide less encouragement or targeted support to daughters, either because they update their own expectations about returns to female entrepreneurship or because, as in Bisin and Verdier (2001), altruistic transmission leads them to steer daughters away from career paths with high social costs. Both channels point toward the same conclusion: family background does not operate in a normative vacuum, and its effect on daughters' career choices depends on the social context in which intergenerational transmission occurs.

The estimates are robust to alternative definitions of traditional provinces based on different quartile thresholds and to specifications that replace the binary traditional indicator with the continuous referendum vote share. The qualitative pattern—greater attenuation of female transmission in more conservative environments—is stable across these robustness checks.

Table 8: Family Background, Gender Norms, and Entrepreneurship

	(1)	(2)
Outcome: Whether Ph.D. Graduate Is an Entrepreneur		
Entrepreneurial Parent	0.037***	0.036**
	(0.011)	(0.016)
Female	-0.004	-0.007
	(0.004)	(0.007)
Female × Entrepreneurial Parent		
× Female	-0.031**	-0.019
	(0.012)	(0.019)
Gender Norms (Province)		
Traditional province	-0.011**	-0.012*
	(0.005)	(0.007)
Interactions with Gender Norms		
Female × Traditional		0.005
		(0.008)
Entrepreneurial Parent × Female × Traditional		-0.022
		(0.024)
Observations	5,975	5,975
Ph.D. Cohort FEs	✓	✓
Scientific Field FEs	✓	✓
University Province FEs	✓	✓

Notes: This table reports OLS estimates of the probability of becoming an entrepreneur five years after Ph.D. completion. *Entrepreneurial Parent* is an indicator equal to one if at least one parent was an entrepreneur (omitted category: no entrepreneurial parent). *Traditional province* is defined based on referendum voting outcomes on abortion rights. Column (2) additionally allows the effect of parental entrepreneurship to vary jointly by gender and local gender norms. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Robustness. The results are robust to alternative measures of local traditionalism based on categorical comparisons between the most and least progressive provinces. Across specifications, gender-specific transmission patterns remain qualitatively similar.

6 Conclusions

This paper studies gender differences in the intergenerational transmission of entrepreneurship among Ph.D. graduates in Italy. By focusing on a population with homogeneous educational attainment and strong career orientation, we limit confounding from human capital and sorting, yet document a substantial gender gap in entrepreneurial entry: women are about 43% less likely than men to become entrepreneurs within five years of graduation, across virtually all fields.

Our main finding is a pronounced asymmetry in intergenerational transmission. Parental entrepreneurship is associated with a near doubling of the baseline probability of entrepreneurial entry, but this effect is significantly stronger for sons than for daughters. The contrast is particularly stark for maternal entrepreneurship: entrepreneurial mothers substantially increase sons' likelihood of entry, while daughters experience little corresponding effect, despite consistently estimated coefficients near zero across specifications.

This asymmetry does not appear to reflect differential transmission of productive skills, as we find no comparable gender gap in the intergenerational transmission of patenting. Instead, our results point to the role of gender norms. Using historical province-level variation in attitudes toward women, we show that the attenuation for daughters is strongest in more conservative environments, suggesting that norms constrain the translation of skills into entrepreneurial choices.

Ongoing work will extend these findings by incorporating more granular occupational measures and additional waves of data, which require access to ISTAT's secure research environment. These analyses will allow for a more precise assessment of mechanisms and heterogeneity in future versions of the paper.

References

Alesina, A. F., Lotti, F., and Mistrulli, P. E. (2013). Do women pay more for credit? evidence from italy. *Journal of the European Economic Association*, 11(S1):45–66.

Autor, D. H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review*, 103(5):1553–1597.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., and Van Reenen, J. (2019). Who becomes an inventor in america? the importance of exposure to innovation. *The Quarterly Journal of Economics*, 134(2):647–713.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic theory*, 97(2):298–319.

Bosma, N., Hessels, J., Schutjens, V., Van Praag, M., and Verheul, I. (2012). Entrepreneurship and role models. *Journal of economic psychology*, 33(2):410–424.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies*, 79(2):645–677.

Dunn, T. and Holtz-Eakin, D. (2000). Financial capital, human capital, and the transition to self-employment: Evidence from intergenerational links. *Journal of labor economics*, 18(2):282–305.

Fernández, R. (2013). Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review*, 103(1):472–500.

Fernández, R., Fogli, A., and Olivetti, C. (2004). Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics*, 119(4):1249–1299.

Guzman, J. and Kacperczyk, A. O. (2019). Gender gap in entrepreneurship. *Research Policy*, 48(7):1666–1680.

Halabisky, D. and Shymanski, H. (2023). Gender gaps in entrepreneurship remain.

Howell, S. T. and Nanda, R. (2024). Networking frictions in venture capital, and the gender gap in entrepreneurship. *Journal of Financial and Quantitative Analysis*, 59(6):2733–2761.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, 87(5):1439–1474.

Lindquist, M. J., Sol, J., and Van Praag, M. (2015). Why do entrepreneurial parents have entrepreneurial children? *Journal of Labor Economics*, 33(2):269–296.

Margherita, B., Panagiotis, K., Yves, P., and Lieve, V. D. B. (2016). Entrecomp: The entrepreneurship competence framework.

Muscio, A. and Vallanti, G. (2024). The gender gap in phd entrepreneurship: Why balancing employment in academia really matters. *Research Policy*, 53(1):104907.

Olivetti, C., Patacchini, E., and Zenou, Y. (2013). Mothers, friends and gender identity. Technical report, National Bureau of Economic Research.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 3982–3992.

Zhang, M., Jensen, K. N., and Plank, B. (2024). European skills, competences, qualifications and occupations (esco).

7 Appendix

A Data and Sample

A.1 Entrepreneurship by Parental Entrepreneurial Status and Gender

Table A.1: Entrepreneurship by Parental Entrepreneurial Status

	Parental entrepreneurial status				
	None	Father	Mother	Both parents	Total
Full sample					
Not entrepreneur	18,038	3,086	590	673	22,387
Entrepreneur	447	125	28	32	632
Total	18,485	3,211	618	705	23,019
Women					
Not entrepreneur	8,931	1,704	314	383	11,332
Entrepreneur	166	46	8	13	233
Total	9,097	1,750	322	396	11,565
Men					
Not entrepreneur	9,107	1,382	276	290	11,055
Entrepreneur	281	79	20	19	399
Total	9,388	1,461	296	309	11,454

Notes: The table reports the number of individuals by their own entrepreneurial status and parental entrepreneurial background. Parental status distinguishes individuals with no entrepreneurial parents, an entrepreneurial father only, an entrepreneurial mother only, or both parents entrepreneurial.

B Institutional Context

B.1 Italian employment and entrepreneurial rates

Table B.1: Employment statistics of the Italian population

	Men	Women
Entrepreneurship	14.1	6.1
	(34.9)	(23.9)
Has a job	61.5	43.9
	(48.7)	(49.6)
Employed	46.4	39.2
	(49.9)	(48.8)
Age	44.9	45.6
	(13.8)	(13.7)
College educated	13.2	17.5
	(33.9)	(38)

Note: This table reports average rates of entrepreneurship, labor force participation, and employment in Italy, separately for men and women. Data are drawn from the Italian Labour Force Survey (waves 2008, 2014, and 2018). The sample is restricted to individuals aged 18–67.

B.2 Gender gaps in employment by scientific field

Table B.2: Employment Rates by Scientific Field and Gender (Five Years After Ph.D. Graduation)

	Employment rate five years after Ph.D. graduation			
	Total (%)	Women (%)	Men (%)	Women − Men (p.p.)
All sample	80.7	80.3	81.2	-0.9
STEM				
Maths and Computer Sciences	91.89	93.06	91.33	1.74
Physics	92.22	90.54	92.96	-2.42
Chemistry	89.32	88.32	90.63	-2.31
Geology	82.68	81.72	83.58	-1.86
Biology	85.02	85.16	84.77	0.38
Agrarian and Veterinary Sciences	77.07	78.10	75.88	2.22
Civil Engineering and Architecture	68.06	69.72	66.51	3.21
Industrial Engineering	89.54	89.44	89.58	-0.15
Mean STEM	**84.73**	**84.75**	**84.78**	**-0.03**
Humanities				
History, Philology, History of Arts	81.05	81.22	80.75	0.48
Philosophy, Psychology, Pedagogy	78.05	76.84	79.84	-3.01
Mean Humanities	**79.55**	**79.03**	**80.30**	**-1.27**
Social Sciences				
Law	54.86	57.30	52.20	5.09
Economics and Statistics	82.81	84.33	81.16	3.17
Political and Social Sciences	79.98	79.26	80.96	-1.71
Mean Social Sciences	**72.55**	**73.63**	**71.44**	**2.19**

Note: This table reports the employment rates of Ph.D. graduates by scientific field, measured five years after graduation. Column 1 presents the overall employment rate. Columns 2 and 3 show rates separately for women and men, and Column 4 reports the gender difference. Scientific fields are grouped into three broad categories: **STEM**, **Humanities**, and **Social Sciences**. Medical Sciences are excluded.

C Transmission of entrepreneurship by parental sex

Table C.1: Regression Results: Intergenerational Transmission of Entrepreneurship

	(1)	(2)	(3)	(4)
Outcome: Whether Ph.D. Graduate is an Entrepreneur				
Female	-0.018***	-0.018***	-0.017***	-0.013***
	(0.002)	(0.002)	(0.002)	(0.003)
Entrepreneurial Parents				
Father	0.015***	0.018***	0.017***	0.026***
	(0.004)	(0.004)	(0.004)	(0.007)
Mother	0.019**	0.020**	0.018**	0.043**
	(0.009)	(0.009)	(0.009)	(0.017)
Both Parents	0.025***	0.028***	0.026***	0.032**
	(0.009)	(0.009)	(0.009)	(0.016)
Female × Entrepreneurial Parents				
× Father				-0.017**
				(0.008)
× Mother				-0.045**
				(0.019)
× Both Parents				-0.011
				(0.019)
Observations	19,637	19,637	19,616	19,616
Ph.D. Cohort FEs	✓	✓	✓	✓
Parental Education		✓	✓	✓
Scientific Field and Univ. Province FEs			✓	✓

Notes: This table reports OLS estimates of the probability that a Ph.D. graduate becomes an entrepreneur. The omitted category for parental entrepreneurship is having no entrepreneurial parents. Column (4) includes interactions between the Female indicator and parental entrepreneurial status; the interaction with the omitted category is not reported. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

D Transmission of patenting activity by parental sex

Table D.1: Intergenerational Transmission and Patenting

	(1)	(2)	(3)	(4)
Outcome: Whether Individual Has at Least One Patent				
Female	-0.035***	-0.035***	-0.028***	-0.027***
	(0.003)	(0.003)	(0.004)	(0.004)
Entrepreneurial Parents				
Father	0.011**	0.010**	0.011*	0.014*
	(0.005)	(0.005)	(0.006)	(0.008)
Mother	0.003	0.003	0.002	0.008
	(0.010)	(0.010)	(0.010)	(0.019)
Both Parents	0.003	0.003	0.004	-0.009
	(0.009)	(0.009)	(0.011)	(0.015)
Female × Entrepreneurial Parents				
× Father				-0.005
				(0.008)
× Mother				-0.011
				(0.020)
× Both Parents				0.023
				(0.018)
Observations	19,637	19,637	19,616	19,616
Ph.D. Cohort FEs	✓	✓	✓	✓
Parental Education		✓	✓	✓
Scientific Field and Univ. Province FEs			✓	✓

Notes: This table reports OLS estimates of the probability that an individual has at least one patent. Entrepreneurial parents are classified as father only, mother only, or both parents (omitted category: no entrepreneurial parent). Column (4) includes interactions between gender and parental entrepreneurship. All specifications include controls for citizenship, region of residence, undergraduate performance, parental education, scientific field fixed effects, Ph.D. cohort fixed effects, and university province fixed effects. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

E Selection considerations

E.1 Methodology: studying selection to confound gender differential effect

Our coefficient of interest is $\hat{\beta}_3$ from the baseline specification:

$$\text{Entrep}_i \;=\; \alpha + \beta_1 \cdot \text{Female}_i + \beta_2 \cdot \text{EP}_i + \beta_3 \cdot (\text{Female}_i \times \text{EP}_i) + \mathbf{X}_i'\boldsymbol{\gamma} + \varepsilon_i, \qquad (3)$$

where EP_i is a dummy for having at least one entrepreneurial parent. The coefficient $\hat{\beta}_3$ is itself a difference-in-differences estimator. Defining $\bar{y}_e^g$ as the mean entrepreneurship rate for gender $g \in \{W, M\}$ and parental background $e \in \{EP, \text{no-EP}\}$, we have:

$$\hat{\beta}_3 \; = \; \underbrace{\left(\bar{y}^W_{\text{EP}} - \bar{y}^W_{\text{no-EP}}\right)}_{\text{EP effect for women}} \; - \; \underbrace{\left(\bar{y}^M_{\text{EP}} - \bar{y}^M_{\text{no-EP}}\right)}_{\text{EP effect for men}}. \tag{4}$$

Equation (4) has a direct implication for what constitutes a valid selection threat. It is *not* sufficient to show that individuals with entrepreneurial parents differ from those without (i.e. that $\delta_2 \neq 0$ in a regression of observables on parental background). What is required for confounding is that the *degree* of selection induced by parental entrepreneurship is *different for women and for men*: that is, $\delta_3 \neq 0$ in the regression

$$X_i \; = \; \delta_0 + \delta_1 \cdot \text{Female}_i + \delta_2 \cdot \text{EP}_i + \delta_3 \cdot (\text{Female}_i \times \text{EP}_i) + \nu_i, \tag{5}$$

where X_i is any pre-determined characteristic or post-graduation outcome variable whose relationship with entrepreneurship could, through differential sorting, generate a spurious $\hat{\beta}_3$. In what follows, we test $H_0 : \delta_3 = 0$ systematically for each potential source of confounding.

E.2 Balanced test in pre-Ph.D. characteristics

Table E.1: Differential selection into the PhD by gender and parental entrepreneurship

	Dependent variable: pre-determined observable							
	(1) Italian	(2) PhD grant	(3) Public grant	(4) Honors deg.	(5) PhD South	(6) PhD North	(7) Born South	(8) Born North
Female (δ_1)	0.004*	0.003	0.002	0.059***	0.027***	-0.022***	0.019***	-0.003
	(0.002)	(0.007)	(0.007)	(0.007)	(0.005)	(0.007)	(0.005)	(0.006)
Entrep. parent (δ_2)	0.001	-0.008	-0.009	-0.038***	-0.018**	0.030***	-0.008	0.032***
	(0.004)	(0.011)	(0.012)	(0.012)	(0.008)	(0.011)	(0.008)	(0.010)
Female × Entrep. parent (δ_3)	-0.009*	-0.006	-0.013	-0.015	-0.004	-0.006	-0.005	-0.021
	(0.006)	(0.015)	(0.016)	(0.017)	(0.011)	(0.015)	(0.011)	(0.013)
Observations	23,019	23,019	23,019	23,019	23,019	23,019	23,019	23,019

Notes: Each column reports an OLS regression of the indicated pre-determined characteristic on a female dummy (δ_1), an entrepreneurial-parent dummy (δ_2), and their interaction (δ_3). Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table E.2: Differential sorting across disciplines: Gender × Parental entrepreneurship interaction

Scientific discipline	Female × Entrep. parent ($\hat{\delta}_3$)
Maths and Computer Sciences	0.001 (0.006)
Physical Sciences	-0.007 (0.007)
Chemical Sciences	-0.001 (0.008)
Earth Sciences	-0.003 (0.006)
Biological Sciences	0.005 (0.011)
Agricultural and Veterinary Sciences	-0.011 (0.009)
Civil Engineering and Architecture	0.011 (0.010)
Industrial and Information Engineering	-0.020 * (0.012)
Ancient, Philological, and Literary Sciences	-0.008 (0.010)
Historical, Philosophical, and Educational Sci.	0.002 (0.010)
Legal Sciences	0.011 (0.009)
Economic and Statistical Sciences	0.012 (0.008)
Political and Social Sciences	0.008 (0.007)
Observations	23,019

Notes: Each row reports the coefficient $\hat{\delta}_3$ on Female × Entrepreneurial Parent from an OLS regression of a discipline indicator. Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table E.3: Years from PhD graduation to first job: entrepreneurial selection by gender

	Pooled		STEM		Social Sciences		Humanities	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	EP only	All	EP only	All	EP only	All	EP only
Entrepreneur	-0.320**	-0.863***	-0.445	-0.128	0.439	-0.370**	0.028	0.368
	(0.131)	(0.222)	(0.322)	(0.686)	(0.567)	(0.173)	(0.280)	(0.571)
Female	0.100***	0.113**	0.207	0.172	-0.036	-0.002	0.077	0.068
	(0.032)	(0.052)	(0.130)	(0.419)	(0.086)	(0.252)	(0.076)	(0.242)
Entrepreneur × Female	0.475***	0.952***	0.543	1.128	-0.154	-0.498*	1.164***	-0.568
	(0.176)	(0.313)	(0.860)	(1.259)	(0.923)	(0.252)	(0.387)	(0.627)
PhD cohort FE	Yes							
Controls	Yes							
Observations	14,967	2,908	699	110	971	150	1,093	214
R^2	0.007	0.015	0.006	0.010	0.001	0.004	0.006	0.001

Notes: OLS regressions. Dependent variable is years from PhD graduation to first post-doctoral job. columns restrict the sample to individuals with at least one entrepreneurial parent. The key coefficient is Entrepreneur × Female: a positive and significant estimate would indicate that female entrepreneurs take longer to find a first job than male entrepreneurs with otherwise similar backgrounds, consistent with a fallback interpretation. Controls include parental education, nationality, geographic origin, and PhD cohort. Errors clustered at PhD institution. * p<0.10, ** p<0.05, *** p<0.01.

E.3 Balanced Ph.D. Field

E.4 Balanced years-to-first job

E.5 Balanced satisfaction with job

Table E.4: Job quality by gender among PhD entrepreneurs: full entrepreneur sample vs. entrepreneurs with entrepreneurial parents

	All entrepreneurs				Entrepreneurs with entrep. parent			
Outcome	Men	Women	Diff. (W–M)	p-value	Men	Women	Diff. (W–M)	p-value
Job satisfaction	7.63	7.07	0.56	0.000	7.70	7.00	0.70	0.006
Autonomy	8.86	8.60	0.26	0.044	8.89	8.71	0.19	0.397
Career opportunities	7.22	6.14	1.08	0.000	7.30	6.50	0.80	0.029
Weekly hours	43.10	33.06	10.04	0.000	44.69	32.89	11.80	0.000
Part-time	1.08	1.23	-0.15	0.000	1.01	1.18	-0.17	0.000

Notes: Columns report gender differences in mean job-quality measures. For each outcome, the table reports mean values for men (female=0) and women (female=1), the difference (Women − Men), and the p-value from a two-sided t-test of equality of means. The left panel uses all entrepreneurs (entrep=1); the right panel restricts to entrepreneurs with at least one entrepreneurial parent (entrep_parent=1).

F Mechanism

F.1 Family interaction with local gender gap in entrepreneurship

G Construction of the Entrepreneurial Skill Intensity Measure

This appendix documents the construction of the occupation-level entrepreneurial skill intensity measure used in Section **??** of the main text. We describe in turn: the ESCO skill universe (§G.1), the OECD entrepreneurial competency framework used as the classification benchmark (§G.2), the two classification methods—semantic ensemble and keyword matching—and how they are combined (§G.3), the aggregation to the three-digit ISCO-08 level (§G.5), and the validation of the resulting measure (§G.6).

G.1 The ESCO Skill Universe

The European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy (Zhang et al., 2024) is maintained by the European Commission and provides a structured multilingual classification of occupational skills and competencies. At its finest level, ESCO lists individual *skill and competence descriptors*: short natural-language phrases or sentences describing a specific ability, knowledge domain, or behavioural disposition (e.g., "manage financial resources", "identify business opportunities", "coordinate project activities").

The version of ESCO we use (v1.1.0) covers 13492 individual skill descriptors.[3] Each descriptor is linked to occupations at the ISCO-08 four-digit level; we aggregate these links to the three-digit level to match the occupational coding available in our survey data (see §G.5 below). Each skill is also typed as either *essential* (a core requirement for the occupation) or *optional* (relevant but not universally required); in our baseline we use the union of essential and optional skills.

The average three-digit ISCO-08 occupation in our working sample is associated with approximately 39 ESCO skill descriptors, with considerable dispersion: the number of associated skills ranges from 1 to 133. To ensure that occupational scores are not driven by this variation in listing intensity, we normalise by the total number of skills per occupation—that is, we compute *shares* rather than counts (see §G.5).

G.2 The EntreComp Entrepreneurial Competency Framework

We classify individual ESCO skills as entrepreneurial or non-entrepreneurial using the EntreComp framework (Margherita et al., 2016), jointly developed by the European Commission and aligned with the OECD entrepreneurial competency literature. EntreComp organises entrepreneurial competencies into three areas—*Ideas*

[3] We use the English-language version of ESCO throughout. Version 1.1.0 was the latest stable release at the time of data construction.

& Opportunities, *Resources*, and *Into Action*—each comprising five competencies, for a total of fifteen core competencies. Table G.1 reproduces the full framework.

The fifteen competencies are organised into three thematic areas that correspond to our three scoring dimensions: the *technical* dimension (skills related to productive and managerial competencies), the *business management* dimension (skills related to opportunity recognition, resource mobilisation, and market activity), and the *personal entrepreneurial* dimension (skills related to self-efficacy, initiative, and resilience).

G.3 Classification Methods

We classify each ESCO skill as entrepreneurial or non-entrepreneurial using two complementary approaches: a semantic ensemble classifier and a keyword-matching rule. These two methods differ in the dimension of entrepreneurial content they capture—conceptual alignment versus explicit terminological overlap—and together form the basis for the binary indicators used in our empirical analysis.

Approach 1: Semantic ensemble classifier. We embed each ESCO skill descriptor and each EntreComp competency into a common semantic vector space using the `all-mpnet-base-v2` sentence-transformer model (Reimers and Gurevych, 2019), and compute cosine similarity between them. This approach identifies skills that are conceptually close to entrepreneurial competencies even when they do not use explicit business terminology—for example, a skill such as "coordinate resource allocation under uncertainty" may align strongly with resource mobilisation and risk management despite not mentioning "entrepreneurship" directly.

Formally, for each ESCO skill k and each EntreComp dimension $d \in \{\text{tech}, \text{biz}, \text{pers}\}$, we compute:

$$\text{sim}_{k,d} = \frac{\mathbf{v}_k \cdot \mathbf{c}_d}{\|\mathbf{v}_k\| \, \|\mathbf{c}_d\|}, \tag{6}$$

where $\mathbf{v}_k$ is the sentence embedding of ESCO skill k and $\mathbf{c}_d$ is the composite embedding of the EntreComp competencies in dimension d. This yields three component similarity scores per skill (`technical_score`, `business_score`, `personal_score`). These are then combined into a single continuous entrepreneurial score:

$$\text{ensemble}_k = f\left(\text{sim}_{k,\text{tech}}, \ \text{sim}_{k,\text{biz}}, \ \text{sim}_{k,\text{pers}}\right), \tag{7}$$

where $f(\cdot)$ is a weighted combination calibrated to the EntreComp framework weights. Across the full ESCO vocabulary, the distribution of ensemble_k has a mean of 0.068, standard deviation of 0.070, and median of 0.047, with a long right tail consistent with the concentration of entrepreneurial content in a subset of occupations.

Table G.1: EntreComp Entrepreneurial Competency Framework

Area	**Competency**	**Description**
Ideas & Opportunities	Spotting opportunities	Identifying opportunities to create value by exploring the natural, cultural and social environment.
	Creativity	Developing creative and purposeful ideas; turning ideas into action by combining knowledge, skills and resources.
	Vision	Imagining the future; developing a vision to turn ideas into action and creating a coherent narrative.
	Valuing ideas	Making judgements about the value of ideas for others; identifying ways to apply ideas in a variety of contexts.
	Ethical & sustainable thinking	Assessing the impact of ideas, opportunities and actions on social, cultural, economic and environmental dimensions.
Resources	Self-awareness & efficacy	Reflecting on and identifying one's needs, assets and goals; believing in one's ability to influence the course of events.
	Motivation & perseverance	Maintaining focus and drive; overcoming obstacles and persisting through difficulty.
	Mobilising resources	Gathering and managing the material, immaterial and human resources needed to act on opportunities.
	Financial & economic literacy	Understanding how money, markets and economic forces work, and using that understanding to create and sustain value.
	Mobilising others	Inspiring, enthusing and getting others on board; building teams; collaborating with and leading others.
Into Action	Taking the initiative	Initiating and leading tasks; acting on ideas rather than just formulating them.
	Planning & management	Setting goals, defining priorities, establishing work plans and managing processes to achieve objectives.
	Coping with uncertainty	Making decisions under incomplete information; adapting to changing circumstances and being resilient under pressure.
	Working with others	Collaborating effectively; communicating, negotiating and resolving conflicts in different contexts.
	Learning through experience	Drawing lessons from one's own actions and those of others; reflecting on successes and failures to improve performance.

Approach 2: Keyword matching. We compile a curated list of entrepreneurial phrases derived from the EntreComp framework—including terms such as "business plan", "opportunity recognition", "resource mobilisation", "strategic planning", and "risk management". A skill is classified as entrepreneurial if its descriptor contains at least one of these expressions. This rule-based approach captures *explicitly* entrepreneurial language and provides a conservative, interpretable benchmark.

Relationship between the two classifiers. The keyword classifier is a strict subset of the semantic ensemble: all skills flagged by the keyword rule are also classified as entrepreneurial by the ensemble, but the ensemble additionally captures skills with entrepreneurial *conceptual content* that do not use EntreComp-specific terminology. Table G.2 reports the cross-tabulation of the two classifiers at the skill level.

Table G.2: Agreement Across Entrepreneurial Skill Classification Methods

Method	**N classified**	**% of skills**
Panel A: Classification summary (N = 13492 ESCO skills)		
(1) Ensemble	5,397	40.0%
(2) Keywords	2,981	22.1%
(3) TF-IDF	5,397	40.0%
Classified by all 3 methods	1,382	10.2%
Classified by no method	6,159	45.6%

Panel B: Pairwise agreement matrix. **Diagonal**: N classified as entrepreneurial (% of 13492). **Upper triangle**: N classified by *both* methods; in brackets, % of the row method also captured by the column method. **Lower triangle**: overall agreement rate, $\frac{N(\text{both=1})+N(\text{both=0})}{13{,}492}$.

	(1) Ens.	(2) Kw.	(3) TF-IDF
(1) Ensemble	**5,397** (40.0%)	2,981 [55.2%]	3,461 [64.1%]
(2) Keywords	82.1%	**2,981** (22.1%)	2,160 [72.5%]
(3) TF-IDF	71.3%	69.9%	**5,397** (40.0%)

Notes: Keywords $\subseteq$ Ensemble by construction: every keyword-classified skill exceeds the ensemble score threshold, implying full containment from the keywords side. TF-IDF is the only method independent of the cosine similarity score; 1936 skills (14.3%) are classified by TF-IDF alone and not captured by any other method.

In our empirical analysis we use the ensemble measure as our baseline and the keyword indicator as a robustness check, as the former has stronger discriminatory power while the latter offers a tighter, more conservative definition of entrepreneurial skill content.

G.4 Skill classification

Table G.3: Top-10 and Bottom-10 Skills by Ensemble Entrepreneurial Score

Rank	Skill description	Ens. score
Panel A: Top 10 — highest ensemble score		
1	*entrepreneurship*	1.0000
2	*lead the brand strategic planning process*	0.8624
3	*implement sales strategies*	0.8273
4	*social entrepreneurship*	0.7964
5	*business incubation*	0.7759
6	*develop business plans*	0.7574
7	*build a strategic marketing plan for destination management*	0.7248
8	*treasury management system*	0.7121
9	*handle cash flow*	0.7106
10	*business ICT systems*	0.6879
Panel B: Bottom 10 — lowest non-zero ensemble score		
1	*undertake clinical chiropractic research*	0.000236
2	*operate turf management equipment*	0.003371
3	*match product moulds*	0.005792
4	*verify product specifications*	0.006106
5	*electrical equipment components*	0.011278
6	*types of tile*	0.016530
7	*match frames to pictures*	0.016530
8	*maintain event records*	0.016530
9	*design post tanning operations*	0.016530
10	*lay non-interlocking roof tiles*	0.016530

Notes: Source: `skills_classified.csv` (ESCO v1.1.0, latest classification). *Ens. score*: composite cosine-similarity ensemble score. A skill is classified as entrepreneurial if its ensemble score $\geq P_{60} \approx$ .. Panel A scores shown to 4 d.p.; Panel B to 6 d.p. to distinguish near-zero values. Skills with ensemble score exactly equal to zero are excluded from Panel B.

G.5 Aggregation to the Three-Digit ISCO-08 Level

We aggregate skill-level scores to the three-digit ISCO-08 level, which is the finest level of occupational detail available in our survey data. Concretely, for each three-digit ISCO-08 occupation o, we compute the entrepreneurial skill share as:

$$\text{EntSkill}_o = \frac{\sum_{k \in \mathcal{K}_o} \mathbf{1}[\text{ensemble}_k \geq \tau]}{|\mathcal{K}_o|}, \tag{8}$$

where $\mathcal{K}_o$ denotes the set of ESCO skills (essential and optional) associated with occupation o, and $\mathbf{1}[\text{ensemble}_k \geq \tau]$ is an indicator for skill k being classified as entrepreneurial at threshold τ. This formula weights each skill equally within an occupation, so that EntSkill_o is simply the proportion of an occupation's associated skills that are classified as entrepreneurial.

The working dataset covers 129 Italian three-digit occupational categories (*codice professionale*, CP3D), spanning 89 distinct three-digit ISCO-08 groups. The crosswalk between Italian CP3D codes and ISCO-08 three-digit codes is one-to-many in the CP3D-to-ISCO direction, so that multiple Italian categories map to the same ISCO-08 group; in such cases we assign to each CP3D category the skill score computed from its ISCO-08 parent.

We construct two binary indicators from the continuous score:

- **High entrepreneurial skill** (`entskill_h` $= 1$): the occupation falls in the upper portion of the EntSkill_o distribution (ensemble score above approximately the 70th percentile). This indicator equals one for 38 of the 129 occupational categories.

- **Low entrepreneurial skill** (`entskill_lo` $= 1$): the occupation falls in the lower portion of the distribution (ensemble score below approximately the 40th percentile). This indicator equals one for 52 occupational categories.

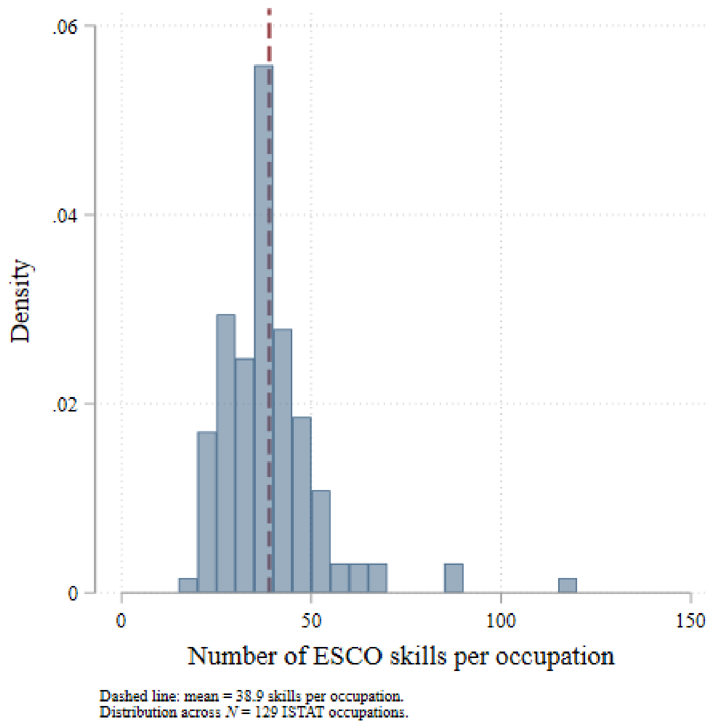
Occupations in the middle of the distribution (39 categories) receive neither indicator and serve as the implicit baseline in specifications that include both binary dummies. Each survey respondent is assigned the score of the three-digit ISCO-08 occupation they hold five years after PhD completion.

Potential limitation: skill overlap across occupations. Some ESCO skill descriptors appear across a large number of occupations (e.g., "communicate effectively"), while others are highly occupation-specific. Cross-cutting skills do not introduce bias into our occupational score *per se*, because our measure is a within-occupation share; however, they reduce the discriminatory power of the index, since a high-frequency entrepreneurial skill contributes equally to the score of all occupations that list it regardless of how central it is to that occupation's core function. As a robustness check, we re-weight each skill's contribution by the inverse of the number of occupations that list it (an inverse-document-frequency scheme), which down-weights generic skills; results are qualitatively unchanged.

G.6 Validation

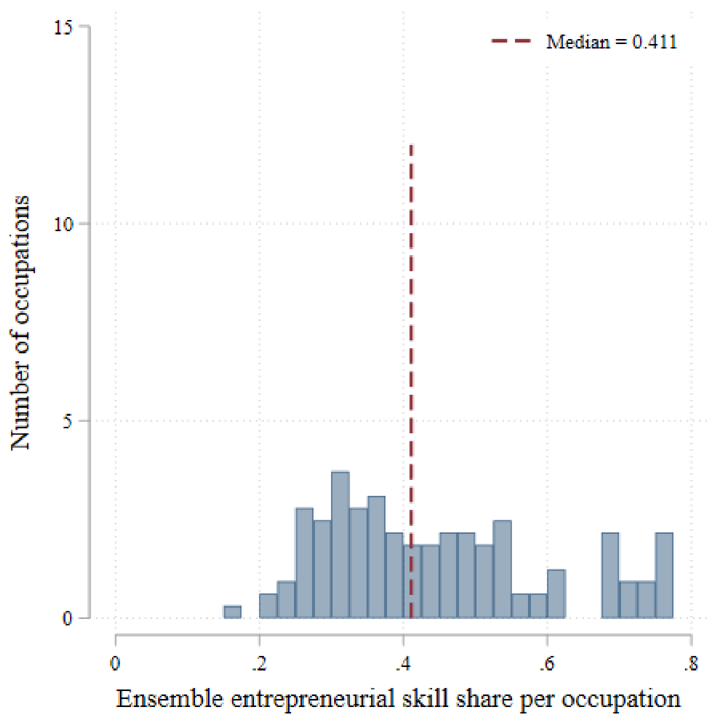
Face validity: top and bottom occupations. Table **??** reports the top-10 and bottom-10 three-digit ISCO-08 occupations by EntSkill_o under the ensemble classifier. The occupations at the top of the ranking—including sales, marketing and development managers (ISCO 122), managing directors and chief executives (ISCO 112), and financial and mathematical associate professionals (ISCO 331)—are those

Figure G.1: Number of skills per occupation



Notes:

Figure G.2: Entrepreneurial score distribution across occupations



Notes:

one would a priori expect to demand entrepreneurial competencies most intensively. At the bottom of the distribution, manufacturing labourers (ISCO 932), mining and construction labourers (ISCO 931), and heavy vehicle drivers (ISCO 833) appear, consistent with a large literature on task content and occupational complexity (**?**). Occupations around the median of the distribution—mid-level technicians, laboratory analysts, and social workers—plausibly require some degree of self-direction and resource management without being primarily entrepreneurial, lending further credibility to the measure.

Comparison between ensemble and keyword classifiers. The ensemble and keyword classifiers agree on the direction of classification for the large majority of occupations, but the keyword measure is more conservative, identifying a smaller share of occupations as high-skill. The gap is largest for managerial and professional occupations whose entrepreneurial content is expressed through general competency language rather than specific business terminology—precisely the cases where the semantic approach adds discriminatory value. We use the keyword indicator as a robustness check in all main specifications and find qualitatively identical results throughout.