De Marzo, Giordano; Mathew, Nanditha; Sbardella, Angelica

Working Paper

Who creates jobs with broad skillsets? The crucial role of firms

ILO Working Paper, No. 94

Provided in Cooperation with:
International Labour Organization (ILO), Geneva

Suggested Citation: De Marzo, Giordano; Mathew, Nanditha; Sbardella, Angelica (2023) : Who creates jobs with broad skillsets? The crucial role of firms, ILO Working Paper, No. 94, ISBN 978-92-2-039266-9, International Labour Organization (ILO), Geneva, https://doi.org/10.54394/KFYG1195

This Version is available at:
https://hdl.handle.net/10419/278563

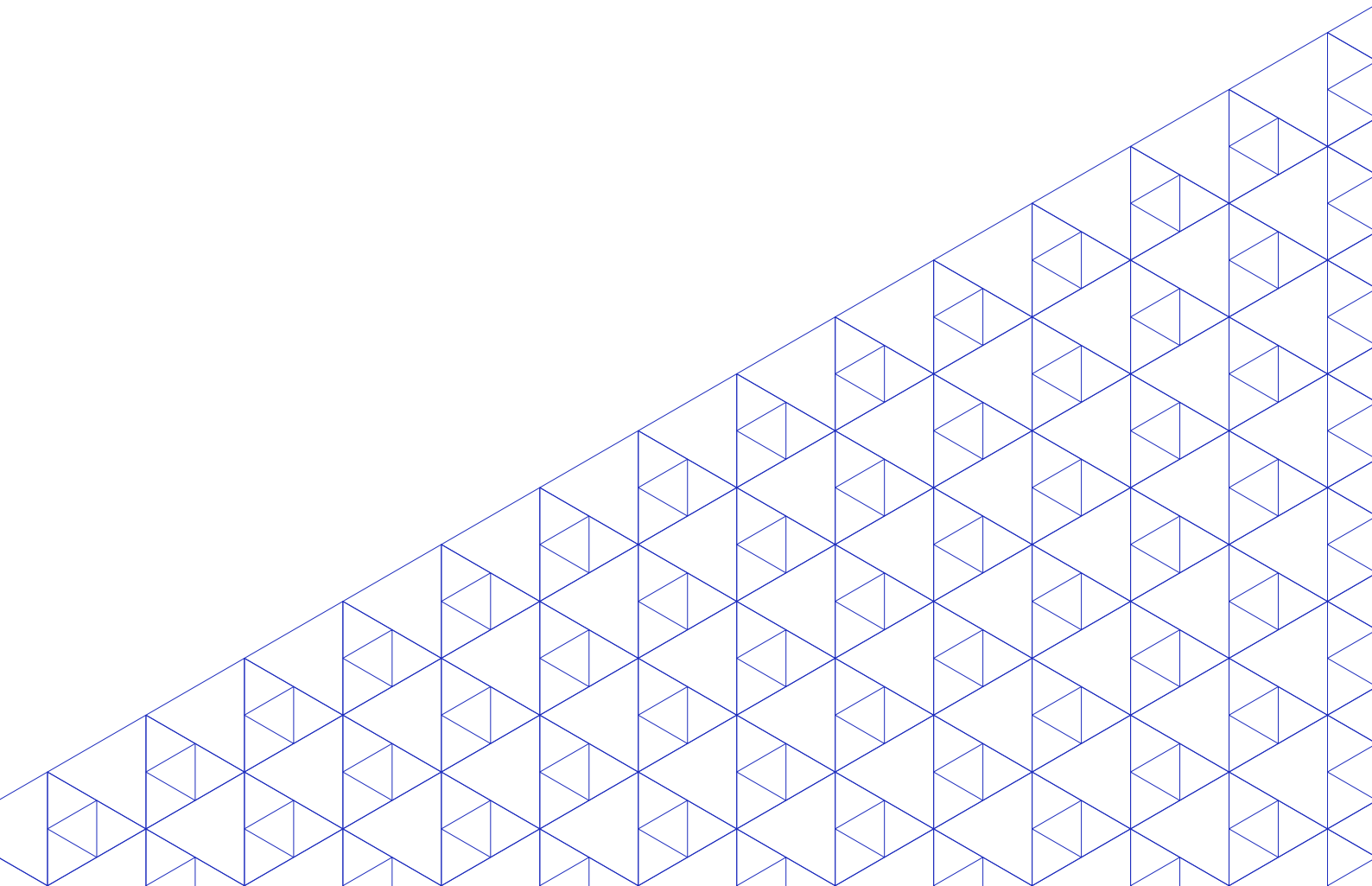
International Labour Organization

ILO Working Paper 94

June / 2023

Who creates jobs with broad skillsets? The crucial role of firms

Authors / Giordano De Marzo, Nanditha Mathew, Angelica Sbardella



ISBN 9789220392706 (print), ISBN 9789220392669 (web PDF), ISBN 9789220392676 (epub), ISBN 9789220392683 (mobi), ISBN 9789220392690 (html). ISSN 2708-3438 (print), ISSN 2708-3446 (digital)

https://doi.org/10.54394/KFYG1195

Information on ILO publications and digital products can be found at: www.ilo.org/publns

ILO Working Papers summarize the results of ILO research in progress, and seek to stimulate discussion of a range of issues related to the world of work. Comments on this ILO Working Paper are welcome and can be sent to RESEARCH@ilo.org, liepmann@ilo.org.

Authorization for publication: Richard Samans, Director RESEARCH

ILO Working Papers can be found at: www.ilo.org/global/publications/working-papers

Suggested citation:
De Marzo, G., Mathew, N., Sbardella, A. 2023. *Who creates jobs with broad skillsets? The crucial role of firms*, ILO Working Paper 94 (Geneva, ILO). https://doi.org/10.54394/KFYG1195

Abstract

Our study investigates the heterogeneity of skill demands within occupations, the firm activities that are associated with demand for broader skill sets, and the firm characteristics that are related to particular skills and different combinations of skills. We use a unique matched database of firm-level data and online job vacancy data for a developing economy, namely, India. Employing a multi-level machine learning technique and an innovative skill taxonomy, we identify and categorize skill requirements of firms. Our empirical analysis provides robust evidence of significant heterogeneity in skill requirements across firms within the same occupations. Additionally, we show that firms demanding diverse skills differ from their counterparts. Firms that are competitive in international markets, as well as those that are more innovative, require digital skills and specific combinations of digital and other skills. Our findings highlight the crucial role played by firms in defining the changing nature of work.

About the authors

Giordano De Marzo is a PhD candidate at Sapienza University Department of Physics, Enrico Fermi Research Centre and Sapienza School for Advanced Studies. He is also a junior research fellow at the Complexity Science Hub Vienna and has been a consultant for ILO, UNU-MERIT and Translated. His research interests cover different topics of Complex Systems, with a focus on the effects of recommendation algorithms on opinion dynamics and social fragmentation.

Nanditha Mathew is a researcher at UNU-MERIT. Nanditha's research interests focus broadly on the microeconomics of innovation and development, in detail, on firm capabilities, firm performance and industrial policy. Nanditha is leading the team on "Conflicting and complementary policies for development" within the new flagship programme of UNU-MERIT on Comprehensive Innovation for Sustainable Development (CI4SD).

Angelica Sbardella is a researcher at the Enrico Fermi Research Centre in Rome and research associate at SOAS University of London. Her research focuses on the application of the economic complexity framework to study industrial competitiveness and development, skills and inequality in the labour market, technological innovation, and how these elements interact in the context of the sustainable transition. Angelica has been a consultant at the European Commission, World Bank Group, ILO, and UNU-MERIT.

Table of contents

List of Figures

List of Tables

▶ Introduction

The demand for new jobs and skills is currently a central topic in the policy and scholarly debates surrounding the changing nature of work. The rapid advances in technological change have the potential to transform the world of work, altering the characteristics of occupations and the skills required to perform them. Several studies investigating different dimensions of the labour market show that the skill requirements and task composition within occupations are changing over time, across countries, industries, firms and technological trajectories (Autor, 2015; Ciarli et al., 2021; Dwivedi et al., 2021; Hershbein and Kahn, 2018).

A significant portion of the literature on skills and tasks connects changes in relative demand across occupations to technological change (Autor et al., 2006, 2008; Autor and Dorn, 2013; Goos et al., 2009; Michaels et al., 2014). However, this literature, often centered around the degree of substitutability between human and automated labor, abstracts from the complex interplay between technology, firm decisions, and their capacity to innovate and reorganize production. By reducing (skill/routine biased) technological change to an exogenous unidirectional process, where firms are passive recipients of new technologies with no ability to shape or influence their development and implementation, it fails to fully recognize the active role of organizations, their decisions on technology adoption or development (Dosi et al., 2000; Ciarli et al., 2021), their internal power relationships and hierarchical layers (Dosi and Marengo, 2015; Cetrulo et al., 2020), and the social and institutional factors that also contribute to shaping the organization of work (Cetrulo et al., 2022; Fernández-Macías and Hurley, 2017; Mishel and Bivens, 2017).

We argue that a missing and crucial element in the discussion on the changing nature of skills is the role of the firm as the locus of the division and organization of labour, also in structuring tasks and skills. It is not only the rate of technological change that determines the scope of human intervention in the production process, but instead the organizational decisions and routines (Dosi and Nelson, 2010; Costa et al., 2021) that considerably regulate whether and which new technologies are to be introduced and the resulting reconfiguration of the amount and characteristics of the labour to be associated to these technologies. Firms may not only differ in the technologies they adopt, but also in how they use the same technology for various business functions, depending on their organizational routines. These differences could lead to variations between firms in the way they remodel their productive and innovation activities, which in turn could have differential effects on employment. Very few works recognize the crucial role of organizations in redesigning their activities and related organizational changes while analyzing the changing demand for skills.

Firms in developing countries often lag behind in technology adoption, which can vary significantly across different business functions (Cirera et al., 2021), and can have important consequences for the labor market. For instance, Martins-Neto et al. (2021) cite the lack of technology adoption as a reason for the scant evidence of job polarization in developing economies. In fact, significant differences in technology adoption among firms, particularly across business functions and sectors, could result in varying demand for skills among firms, even within the same occupation. However, this relationship is not clear-cut: thus far, the literature has not established a definitive association between technology use and changes in the composition of skills (Cirera et al., 2021). This raises the question of which firms demand diverse skills, and how diverse skill requirements are associated with different firm-level activities.

The present paper aims at filling this gap and examines the extent of firm-level differences in the skills demanded within even narrowly-defined occupation categories, and how firm characteristics map into the demand for specific skills. We study this in the case of the labor market of India, using online vacancy data matched with comprehensive firm-level data and relying on an innovative skill taxonomy and methodology to implement it using vacancy data developed

by the ILO Research Department for developing and emerging economies (Bennett et al., 2022) comprising information on digital, other cognitive,[1] manual and socio-emotional skills.

Several recent studies analyse the changing demand for skills using online vacancy data, mainly focusing on the US labor market with Burning Glass Technologies data. These works show within-occupation variation in skill requirements, positive associations between cognitive and socio-emotional skills, wages and firm performance (Deming and Kahn, 2018), with rising educational requirements in high-wage cities and larger states (Blair and Deming, 2020; Modestino et al., 2020). This recent literature also suggests that AI skills are in high demand (Acemoglu and Restrepo, 2020), benefiting highly-qualified professionals, substituting for customer service activities (Alekseeva et al., 2021), and accelerating routine-biased technology adoption (Hershbein and Kahn, 2018). Copestake et al. (2021) focus on Indian service firms and find a rise in the demand and a wage premium for AI adoption in professional services, leading to net labor displacement and reduced wage offers in non-AI jobs. To our knowledge, this is the only study focusing on a developing economy to study the changing nature of skills. Further, Copestake et al. (2021) only analyze AI skills, while we examine how firms differ in their demand for a broader set of skills, and how different typologies of skills and their combinations relate to firm characteristics.

In our work, *firstly*, we analyze the heterogeneity of skills demanded across firms within the same occupation. *Secondly*, we examine which firms demand heterogeneous skill sets, and study the characteristics of the firms that create skill-diversified jobs, by considering a rich set of firm-level characteristics such as size, age, wages, R&D investment, exporting and financial status. Our intuition is that firms that engage in complex activities, with an organizational environment that favours learning, demand for employees equipped with skills instrumental in performing more advanced functions, that will in turn involve more diversified tasks than the standard prescriptions for the job. *Thirdly*, with the aim of pinpointing which firm activities relate to which typology of skills, we investigate how the demand for digital, other cognitive, socio-emotional or manual skills, and their different interactions map into firm characteristics and performance. Our results indicate that firms with a high demand for diverse skills tend to be younger, pay higher wages, and engage in more complex activities. Our findings also show that specific types of skills are associated with different firm activities and outcomes. In particular, digital skills, as well as a combination of digital and manual skills, are related to higher wages. Digital skills are also linked to high engagement by firms in complex activities.

To our knowledge, this is one of the first studies providing evidence of high heterogeneity in the skills demanded by firms, even within the same occupation, and in particular in the context of a developing economy. Furthermore, this paper is also among the first studies distinguishing between digital and other cognitive, which in previous literature were often combined under the umbrella of "cognitive" skills (Deming and Kahn, 2018). As the changing nature of skill demand is often related to technological change, as mentioned above, we argue that digital and other cognitive skills should be studied separately. In fact, by doing so, we are able to disentangle and highlight their distinct relationships with different kinds of firm activities. This distinction is even more relevant in the context of developing economies, for in the initial stages of digitalization adoption firms may require more digital than cognitive skills. Our results emphasize the role of complementary policies that should be implemented at the firm level along with educational policies to tackle changes in the world of work. We further discuss the policy implications of our analysis in the conclusion.

The article is structured as follows. Section 1 details the data sources, their degree of representativeness, and describes the construction of skill and occupational variables, through machine-learning techniques. Section 2 reports evidence on inter-firm heterogeneity in skill demands.

[1] We refer to these skills as "other cognitive skills" since the skill classification by Bennett et al. (2022) includes "digital skills" as a subset of "cognitive skills". In this work, we analyze digital and cognitive skills separately, as you will see in the following sections.

In Section 3, we present the empirical results on firm characteristics related to the demand for skill-diversified jobs and for specific bundles of skills. Section 4 offers our conclusions.

▶ 1 Data

In this study, we use and merge two data-sets, i) the online job vacancy data and ii) firm-level balance sheet data. In the present section we describe the two data-sets with initial statistics, the construction of the occupation and skill variables, the data operations that lead to the final matched data-set, and finally the degree of representativeness vis-à-vis the Indian labour market.

Online job vacancy data from the Naukri portal

General data features. - Our first data source consists of online job vacancies posted on naukri.com, the leading recruitment platform in India.[2] Established in 1997, the platform caters to corporate recruiters, placement agencies, and job seekers. The data is collected through web scrapping and covers the years 2016, 2017, 2019 and 2020. This yields a sample size of around 20 million vacancies, excluding duplicates.[3] Each vacancy consists of a raw text that specifies different characteristics of the job, based on free descriptions. The observations are structured into separated fields, that include the job title, a detailed free-text job description containing the sub-fields "Role" and "Role Category" (which bear similarity with ISCO-08 2-digit and 4-digit occupational codes, respectively), and "Education" (which details the required educational attainments). Moreover, for each vacancy we observe the information on the respective industry, company name, date of the posting, workplace location, the required experience and skills, and the pay-rate – if disclosed by the company. A schematic overview of the vacancy data is provided in Appendix Table A1.

Construction of the occupational variables. - We use machine-learning techniques to map job titles to their corresponding ISCO-08 occupational codes at the 2-digit level. The 2-digit level allows us to differentiate the different roles with enough information, without being at risk of misclassifying when going into too much detail. We start with data from 2016 and 2017 and consider the highly disaggregated "Role" sub-field.[4] We organize all the roles into a single list with 678 unique words, taking into consideration only those roles that appear at least 10 times in the data. We then match the roles with 2 digit ISCO-08 codes by following a set of intuitive rules, that we summarize in Appendix Section B. In such a way we are able to classify 76 per cent of Naukri 2016-2017 job listings. However, the field "Role" is available only for few job posts in the 2019-2020 data. Moreover, while Naukri roles are a good first approximation for labeling firstly ISCO-08 4-digit and secondly 2-digit occupations, they are not always a direct or precise match, and therefore we refine the accuracy of our attributions by using a machine learning (ML) approach. To this end, we use a sub-sample of the 2016-2017 labeled data as a training set for a Natural Language Processing (NLP) algorithm that, instead of focusing solely on the "Role" sub-field, takes into account the whole free-text "Job Title" field – see Appendix Table A1 – and maps it directly to 2-digit ISCO-08 occupational categories. Next, we rely on the same NLP algorithm to classify the remaining 2016-2017 and 2019-2020 vacancies, thereby classifying 71 per cent of the whole sample – 11.1 out of 15.6 million online vacancies. The details of this procedure are provided in the Section B of the Appendix.

Figure 1 shows the time evolution of the shares of job posts in 1-digit ISCO-08 major groups.

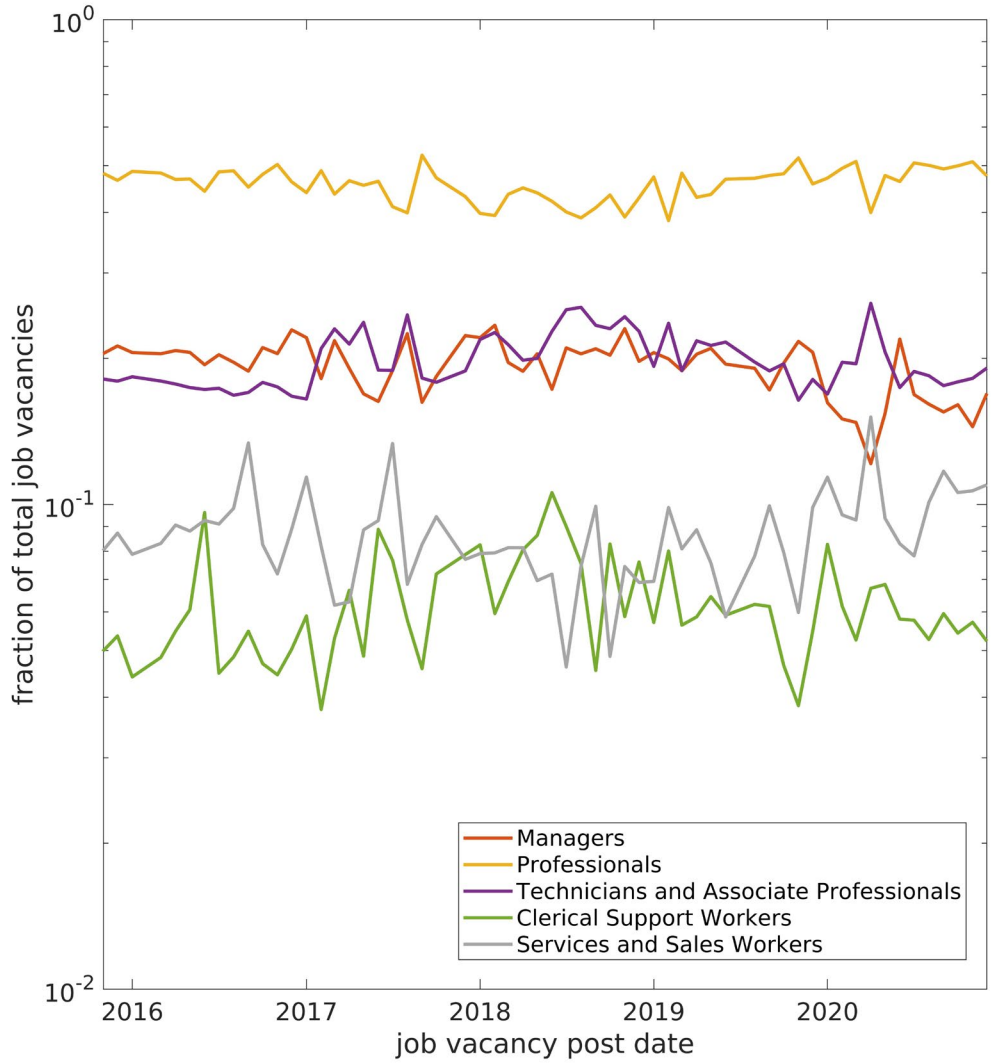
---

[2] See, e.g., https://www.tribuneindia.com/news/brand-connect/top-10-job-search-portals-in-india-372146 and https://economictimes.indiatimes.com/news/how-to/looking-for-a-job-hera-are-top-5-job-portals-to--help-you-find-the-right-job/articleshow/89075753

[3] We use the id associated to each job post to remove the few duplicates present in the naukri.com data-set.

[4] A higher disaggregation allows us to obtain a more accurate matching because a 4-digit misclassification will less likely affect the 2-digit assigned code.

▶ **Figure 1: Occupational shares across time (1-digit level)**



Notes: The figure shows the time evolution of the shares of job posts in 1-digit ISCO-08 occupations present in the Naukri data-set.

We observe a substantial time consistency across shares, that, albeit some within-year monthly variations, are almost stationary over time, indicating that the data is not subject to systematic fluctuations across time. In particular, the bulk of job posts in the Naukri data-set is concentrated in managerial, professional, technical, clerical and sales occupations, with very high percentages in Managers and Professionals and decreasing shares in lower skill-intensive macro occupations, as confirmed by the occupational shares for 2-digit ISCO-08 codes presented in Appendix Section C (see Appendix Figure C1). As can also be observed in the Appendix, as a further step, we benchmark our data-set against the 2017-18 Indian Period Labour Force Survey comparing occupational and industrial shares in the Naukri data against those in the survey – by mapping the industries reported by Naukri into the 2008 Indian National Industry Classification (NIC) at the 2-digit level. Thus, from Appendix Figure C2 it is possible to notice that the online vacancy data is biased towards higher-skilled occupations. However, while very low-skilled jobs are not present in the listings, middle-skilled clerical, service and technical occupations are fairly represented. This is clear also in terms of the sectoral distribution, confirming that, when focusing on online vacancies, the lion share of listings appears to be in the service sector, while low shares

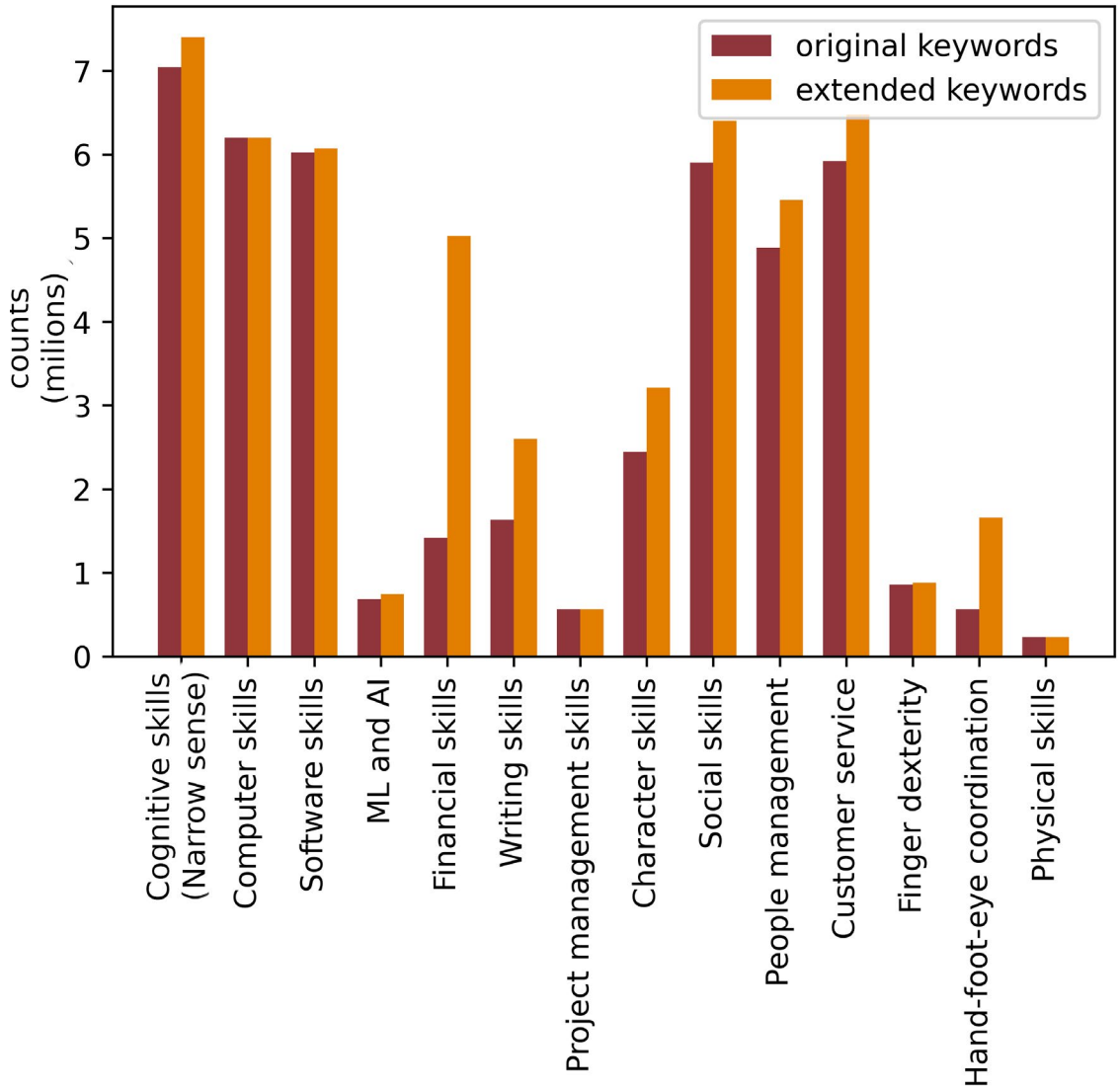
are present in construction and manufacturing, and no job ads are found in agriculture, mining or electricity, as can be observed in Appendix Figure C3. As an external validation, our observations on the occupational and sectoral representativeness of the vacancy data are in agreement to those presented in Copestake et al. (2021), who use similar data for India. Finally, we also geo-localise each vacancy by cleaning the reported city names. We present the distribution of the vacancies by cities in Appendix Figure C4: over 85 per cent of the vacancies are concentrated in large urban areas, with the vast majority located in Bangalore, Delhi and other metropolitan cities (see Appendix C for more details).

Construction of the skill variables. - We create the skill variables using the skill taxonomy developed by Bennett et al. (2022). Drawing from labour economics and psychology literature, this taxonomy distinguishes between the broad categories of cognitive skills, socio-emotional skills, and manual skills, divided into fourteen more detailed sub-categories.[5] These are defined through a comprehensive compilation of keywords drawn from a selection of seminal studies that classify skills using online vacancy data (Deming and Kahn, 2018; Deming, 2017; Hershbein and Kahn, 2018; Kureková et al., 2016), and complemented by studies that similarly exploit different kinds of data sources (Atalay et al., 2020; Autor et al., 2003; Spitz-Oener, 2006). For the various analysis we present in the paper, we divide cognitive skills into two categories: digital skills and other cognitive skills. It is worth noting that while Bennett et al. (2022) include digital skills as a sub-skill within the broader category of cognitive skills, we disaggregated these two types of skills in this study.

Next, we employ a Natural Language Processing algorithm to identify the presence of these keywords in Naukri listings and assign to each vacancy the respective skill sub-categories. To this aim, we clean and tokenize the description of the skills required in each vacancy. We then expand this approach by also considering synonyms of these keywords, identified through web-scrapping of the website www.thesaurus.com homepage (methodological details are again provided in Bennett et al., 2022). By doing so, we are able to assign an average of 5.6 skill sub-categories to each vacancy. The resulting skills distribution is shown in Figure 2, which reveals that cognitive and, to a lesser extent, social, customer service, as well as different kinds of digital skills – especially computer and software competences – are highly represented in the data, whereas, as expected, manual skills are less represented.

[5] Among cognitive skills, the sub-categories are cognitive skills (narrow sense), general computer skills, specific software and technical support skills, machine learning and artificial intelligence skills, financial skills, writing skills, and project management skills. Socio-emotional skills comprise character skills, social skills, people management skills, and customer service skills. Finally, manual skills are divided into finger dexterity skills, hand-foot-eye coordination skills, and physical skills.

▶ **Figure 2: Distribution of skills-subcategories identified in the vacancy data**



Notes: "Original keywords" refer to the keywords presented by Bennett et al. (2022) in their methodology, while "extended keywords" also include synonyms of the original keywords identified through web-scrapping (see below and Bennett et al. (2022) for information on the methodology used). The y-axis reports the skill keyword count in millions.

Firm-level information from the Prowess database

As mentioned above, to study skill requirements of firms in the Indian labour market, we match our vacancy data with detailed firm-level information from the Prowess database, gathered by the *Centre For Monitoring Indian Economy* (CMIE), a private company providing information on Indian firms. The information is collected by CMIE from firms' annual balance sheets and income statements, and covers both publicly listed and non-publicly traded firms from a wide cross-section of manufacturing, services, utilities, and financial industries. These companies account for around 70 per cent of India's industrial output, 75 per cent of corporate taxes, and more than 95 per cent of excise taxes collected by the Indian Government. Prowess is considered the largest firm-level database for India, has allowed researchers to track several dimensions of firm characteristics over time (Goldberg et al., 2010), and has been employed in different studies to investigate firm dynamics and production activities (Coad et al., 2020; Dosi et al., 2017; Goldberg et al.,

2010). Hence, Prowess allows us to observe rich information on the firms posting online vacancies, and to connect the nature of skills they demand to firm characteristics and performance, such as, their growth, profitability, size, age, R&D investments, wages and exporting status. The list of all the variables we exploit in the empirical analysis, along with their definitions and summary statistics, are presented in Table 1.

Data matching. - We match the online vacancy data with the firm-level data for each year, exploiting information on firm names. In such a way, we are able to match 10 per cent of the vacancies. In fact, even though Prowess is the largest firm-level panel available for India, it presents similar caveats to other comparable private firm-level databases. It does not include informal enterprises and has a very low share of micro and small enterprises. This might fundamentally contribute to the large share of unmatched data from online job vacancies and presents a limitation of our study. Even though we are not able to predict how our findings would change when including all firms in the Naukri sample, we believe in the validity of a large part of our results. In fact, higher skill diversification and the lion share of the demand for certain skills (digital and other cognitive) and their combination is likely to be concentrated in medium and large enterprises, which are well represented in the Prowess database. As mentioned above, and in line with both occupational and sectoral distributions, online vacancy data are typically not representative of a country's labour force (e.g., see Deming, 2017; Deming and Kahn, 2018; Hershbein and Kahn, 2018) and our data is no exception in this regard.

▶ Table 1: Employed firm-level variables from Prowess and Naukri data-sets, definition and summary statistics

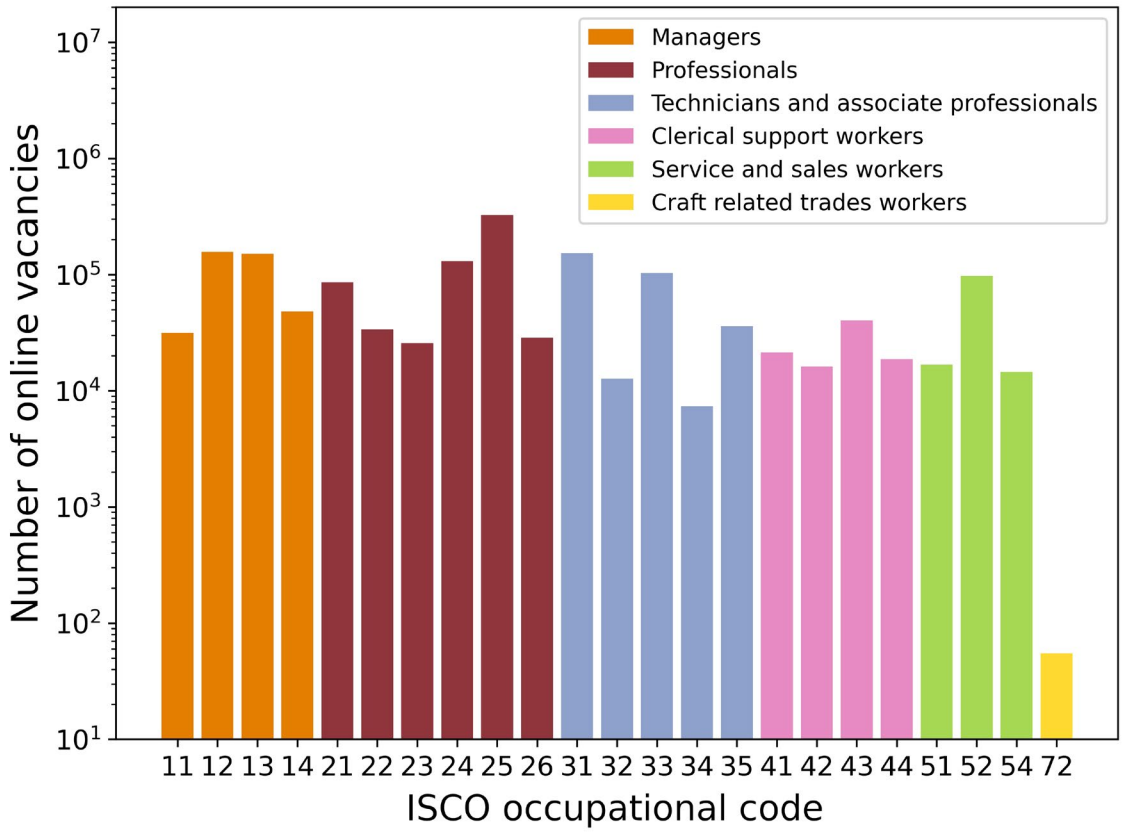
Variable	Definition	Mean	Median	Std. Dev
Firm Growth	Log difference in sales between ***t*** and ***t-1***	0.058	0.071	0.565
Profitability	Profits over sales, relative to the sector average	-1.988	0.063	18.00
Sales	Total sales revenue, goods & services	17113.2	1817.25	141595.3
Firm Size (proxied by Sales)	Log of total sales	5.59	6.00	2.93
Firm Size (proxied by assets)	Log of total assets	7.831	7.835	2.077
Age	Number of years since the year of incorporation of firm	24.868	22	16.49
Wage Intensity	Total wages paid by the firm over total sales	0.213	0.123	0.209
Export intensity	Total value of sales of goods and services from outside India over total sales of the firm	0.303	0	11.68
Export dummy	Takes value 1 if firms has positive sales from selling goods and services outside India	0.226	0	0.418
R&D Intensity	R&D expenditure over total sales of the firm	0.007	0	0.688
R&D Dummy	Takes value 1 if firm spends on R&D	0.146	0	0.353
Skill Div. Index	Entropy of the skill distribution at the firm level	1.77	1.88	0.457

Notes: The construction of the Skill Diversification Index is detailed in section 3. The statistics are calculated for the matched sample of Naukri and PROWESS, comprising a total of 5,237 firms.

The Naukri-Prowess matched data-set features 1,556,394 ads (as mentioned, around the 10 per cent of the original Naukri sample), 5,237 firms, 209 cities (262 in the unmatched sample), with 25 per cent of the vacancies for managers, 39 per cent for professionals, 20 per cent for technicians, 9 per cent for clerical support workers, 8 per cent for service and sales workers, and very

low percentages in craft workers, machine operators, and assemblers.[6] The average number of ads posted by each firm is 297 in a range going from 1 to 98,100, and the "average firm" displays 297 ads, for jobs in 6 occupations located in 4 cities and requiring on average 5 skills.

▶ **Figure 3: Number of jobs posts, at the 2-digit ISCO-08 occupational level (pooled)**



Notes: The figure shows, for the matched data-set, the number of online job posts for each 2-digit ISCO-08 occupational categories, where the y-axis is displayed in logarithmic scale. Color indicates the corresponding 1-digit ISCO-08 groups.

To better understand the occupational distribution of the matched data-set, in Figure 3 we report the number of vacancies for each 2-digit ISCO-08 occupation, a lower level of disaggregation, pooling all job posts over time. We observe that a large number of vacancies cover professional occupations – e.g., in finance, consultancy, manufacturing or ICTs –, and a lower percentage of managerial ones, which are likely filled through less publicly advertised channels. In orange we can observe how the combination of the ISCO-08 codes *12–Administrative and commercial managers and 13–Production and specialized services managers* are the most represented among managerial occupations, while the most atypical roles in code *11–Chief executives, senior officials and legislators* displays a number of job posts lower by one order of magnitude. By contrast, the group of professionals (burgundy), the most represented among the online adverts, is more evenly distributed among 2-digit codes, with the prominent occupations being *25-Information and communications technology professionals, 24-Business and administration professionals*, both with more than 100,000 ads, and *21–Science and engineering professionals*. Among technical occupations (pale blue) *31-Science and engineering associate professionals and 33-Business and administration associate professionals* are the most represented occupations, each with around 100,000

[6] Note that the number of firms included in each regression analysis reported later is limited by the firms that report specific variables used for analysis.

ads; while for services and sales workers (green) we find more than 100,000 ads for *52-Sales workers* and more than 10,000 for *43-Numerical and material recording clerks* (pink). The figure also shows that some occupational groups are not at all covered by the data, including craft and related trades workers, elementary occupations, and skilled agricultural and fishery workers – as was also noted in terms of sectoral distribution of the Naukri vacancy data where jobs in manufacturing and agriculture are almost not present, while a good coverage in services and transport, storage, and communication industries is featured.

▶ 2 Inter-firm heterogeneity in skill requirements

We begin our analysis by studying the within-occupation heterogeneity in skill demand between firms. We thus aim at answering questions such as whether and to what extent the tasks performed by computer programmers differ from one firm to another.

Here, we look at the similarity (or lack thereof) in the skill demand of each occupation across firms by building a skill vector $s^{(i)}$ for each firm i, each occupation and year. $s^{(i)}$ contains the fourteen categories identified in the ILO skill taxonomy, and is thus of dimension N_{skills}= 14. The entries of the skill vector are defined as $s_l^{(i)} = N_{jobs}$ for (l = 1, ·, N_{skills}) where N_{jobs} denotes the number of jobs posted by firm i that require skill l in the chosen occupation (with N_{jobs} = 0 if the skill is not sought after by the firm). Then, for each pair of firms i and j that listed at least one job post in the chosen occupation and year, we compute an index of similarity between the two corresponding skill vectors. In practice, especially because we are interested in the frequency of the different skills demanded for each occupation rather than skill-intensity, we employ a measure of cosine similarity quantifying how similar the demand for skills is between each possible pair of firms that post job listings in a specific occupation in a given year, independently from the number of skills they require.

The cosine similarity CS between the skill vector $s^{(i)}$ of firm i and the skill vector $s^{(j)}$ of firm j can be defined as the dot product between the two vectors divided by the product of their Euclidean norms.

To express this formally:

$$CS(\mathbf{S^{(i)}}, \mathbf{S^{(j)}}) = \frac{\mathbf{S^{(i)}} \cdot \mathbf{S^{(j)}}}{\|\mathbf{S^{(i)}}\| \|\mathbf{S^{(j)}}\|} = \frac{\sum_{l=1}^{N_{skills}} S_l^{(i)} S_l^{(j)}}{\sqrt{\sum_{l=1}^{N_{skills}} S_l^{(i)2}} \sqrt{\sum_{l=1}^{N_{skills}} S_l^{(j)2}}}. \quad (1)$$

In our case, having only non-negative entries in the skill vectors, $CS(s^{(i)}, s^{(j)})$ ranges between 0 and 1, with value 0 indicating orthogonality and thus complete de-correlation, and value 1 indicating that the two vectors are identical. Analysing the distribution of the cosine similarity between the required skills across all pairs of firms for each occupation allows us to observe the within-occupation heterogeneous demand for skills. More specifically, by means of a kernel density analysis, we observe its deviation from a normal distribution – i.e. to understand whether and to what extent the within-occupation skill non-similarity differs from the normal case.

As can be observed in Figure 4, at a first glance, the kernel densities of cosine similarity suggest asymmetry in the tails and significantly fat left-tailed distributions. In the following, we empirically test and quantify in a parametric way the observed empirical distributions and their deviations from a Guassian. With this aim, we rely on the Subbotin family of distributions introduced into economic analysis by Bottazzi et al. (2002), a general and flexible distributional model that allows us to parameterize the two tails of the distribution.

The Subbotin distribution takes the following functional form:

$$f(g;a,b,m)=\frac{1}{2*a*b^{\frac{1}{b}}*\Gamma\left(1+\frac{1}{b}\right)}e^{\left(-\frac{1}{b}*\left|\frac{g-m}{a}\right|^{b}\right)} \quad (2)$$

with Γ (.) standing for the Gamma function, and the three parameters defining the distribution being:

1. *m*, the location parameter which indicates the existence of a general trend in the data;

2. *a*, the scale parameter which determines the spread or dispersion of the distribution;

3. *b*, the shape parameter which indicates the shape/thickness of the tails.

When *b*=2, f (*g;a,b,m*) becomes a Gaussian distribution, and when *b*=1 a Laplace distribution, which are particular cases of the Subbotin family of probability densities. The smaller values of *b* correspond to fatter tails of the distribution. Bottazzi and Secchi (2011) extend the Subbutin distributions to a 5-parameter family of distributions, the Asymmetric Exponential Power (AEP) distribution, able to accommodate for asymmetries in the data, e.g., for asymmetric tails. The AEP parameterizes also the shape (and scale) of the two sides of the distribution. Therefore, in addition to the location parameter m representing the mode, it includes two positive scale parameters a_r and a_l, associated with the distribution width respectively above and below the modal value, and two positive shape parameters b_r and b_l describing the behavior of the right and the left tail, respectively. The AEP density presents the following functional form:

$$f_{AEP}(x;a_l,a_r,b_l,b_r,m)=\frac{1}{C}e^{\left(-\left[\frac{1}{b_l}*\left|\frac{x-m}{a_l}\right|^{b_l}*\theta\left(m-x\right)+\frac{1}{b_r}*\left|\frac{x-m}{a_r}\right|^{b_r}*\theta\left(m-x\right)\right]\right)} \quad (3)$$

where the normalization constant $C=a_l b_l^{1/b_l-1}\Gamma(1/b_l)+a_r b_r^{1/b_r-1}\Gamma(1/b_r)$, and $\theta(x)$ is the Heaviside theta function taking value 1 if $x>0$ and value 0 if $x\leq 0$.

As mentioned above, a value for b_l (b_r) lower than 2 would indicate that the left (right) tail is fat and there are more extreme events in the distribution than what one would expect with respect to a normal. Please notice however that here we are not focusing on the right tail of the distribution since the maximum value is fixed and the cosine similarity cannot be greater than 1.

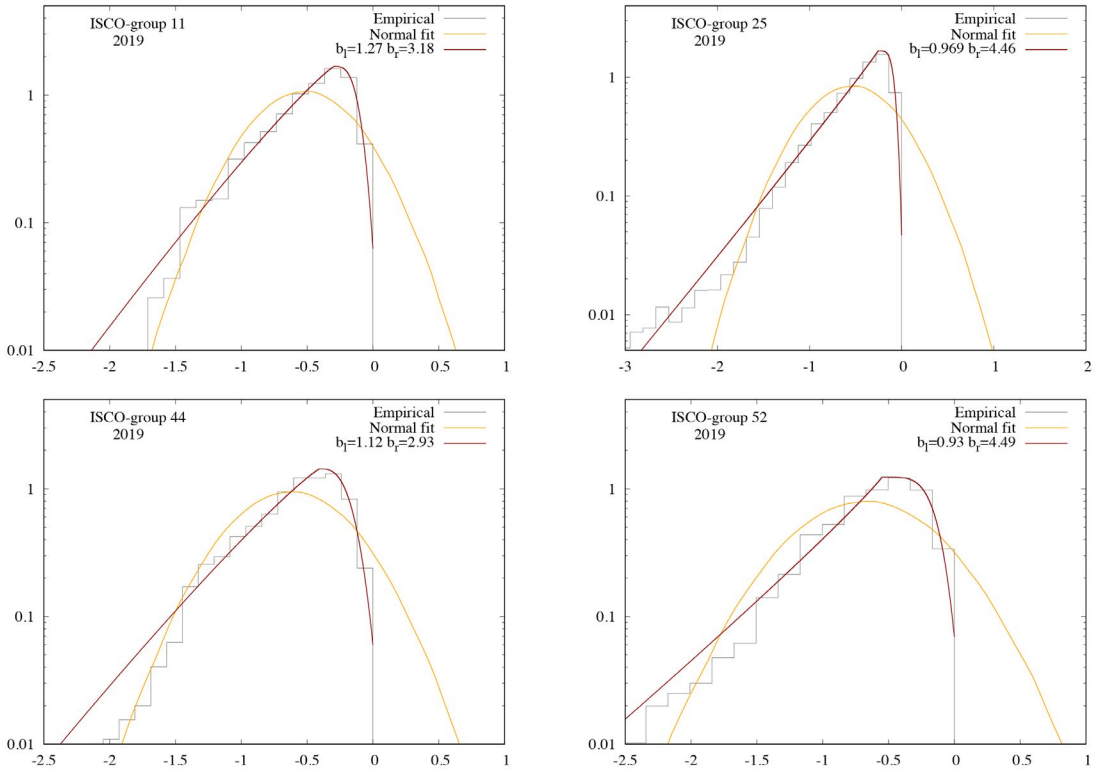
Figure 4 compares the empirical distribution of the cosine similarity with the AEP and normal fit for the selected two-digit occupational categories (*11-Chief Executives, Senior Officials and Legislators, 25-ICT Professionals, 44-Clerical Support Workers and 52-Sales Workers*), with the black curve representing the empirical histogram, the red curve the AEP fit, and the orange curve the normal fit, and the x-axis being shown in log-scale. As noted above, from the figure it is visually detectable that the distributions are fat left-tailed. The maximum likelihood estimates[7] of the AEP parameters are presented in Table 2. The parameter values confirm the visual impression: the b_l parameters are always lower than 2, confirming the presence of a large number of firms demanding

[7] The parameters are estimated by maximum likelihood method (MLE) following Bottazzi and Secchi (2011)

very dissimilar skills sets within the same occupations, and this is widespread across all occupational categories and time.[8]

We therefore find very strong evidence of within-occupation heterogeneity in skills demanded across firms. To our knowledge, this is the first study that empirically documents such widespread differences between firms in skill requirements across occupational categories, emphasising the crucial role of firms and their activities in the changing nature of skills and with potential implications for future studies analysing job and skill dynamics.

▶ **Figure 4: Empirical distribution across firms of cosine similarity (black) together with normal (orange) and Asymmetric Exponential Power (AEP) (burgundy) fits for the occupational categories: *Chief Executives, Senior Officials and Legislators (ISCO 11), ICT Professionals (ISCO 25), Clerical Support Workers (ISCO 44),* and *Sales Workers (ISCO 52).***



[8] Even though we do not report them here for the sake of brevity, the tests were performed for all occupations and years, yielding similar results and documenting fat-tailed distribution in all our data-sample. The analysis is restricted to only occupational categories with sufficient number of observations to run the estimation and hence we have excluded the ISCO-group 72.

▶ Table 2: Estimates of the shape parameters (b_l) and (b_r) of cosine similarity distribution for different ISCO 2-digit categories

Occupations (2-digit)	b_l	b_r	N
Chief Executives, Senior Officials and Legislators	1.127	3.183	248678
Administrative and Commercial Managers	1.064	3.649	1927039
Production and Specialized Services Managers	1.022	4.329	1463295
Hospitality, Retail and Other Services Managers	1.108	3.737	380280
Science and Engineering Professionals	1.101	3.907	725177
Health Professionals	0.991	3.784	125639
Teaching Professionals	1.012	4.581	114341
Business and Administration Professionals	1.042	5.121	2078305
Information & Communications Technology Professionals	0.969	4.459	895029
Legal, Social and Cultural Professionals	1.002	4.411	190512
Science and Engineering Associate Professionals	0.982	5.974	817400
Health Associate Professionals	1.090	3.642	57537
Business and Administration Associate Professionals	1.040	3.798	1169193
Legal, Social, Cultural & Related Associate Professionals	1.264	3.183	17547
Information and Communications Technicians	1.066	3.492	129561
General and Keyboard Clerks	1.020	3.697	163053
Customer Services Clerks	1.000	2.870	98521
Numerical and Material Recording Clerks	9.820	4.142	451221
Other Clerical Support Workers	1.125	2.933	155219
Personal Services Workers	1.043	3.627	74882
Sales Workers	0.930	4.495	356428
Protective Services Workers	1.061	3.946	36065
Drivers and Mobile Plant Operators	1.220	3.653	677

▶ 3 Skill demand and firm characteristics

Having observed the firm-level variation in skill requirements in the previous section, here we focus on the firm characteristics that are associated to different skill demands. Specifically, in Subsection 3.1, we investigate which firms create more "skill-diversified" jobs, while in Subsection 3.2, we explore whether and how different types of skills (or combinations of them) relate to different firm outcomes.

Which firms create "skill-diversified" jobs?

Our goal here is to understand which firms create jobs with diversified skills, that is, to trace the origin of the demand for diversified skills. Also on the basis of what we observed in the previous section, we hypothesize that studying the demand for skills at a mere occupational level may hide how the demand for different skill sets relate to the different activities performed at the firm-level. For instance, while most software engineers require digital skills, a Google India's software engineer might also require socio-emotional skills to perform more complex and diversified tasks, such as strategizing with clients or overseeing co-workers, which might not be necessary in a firm performing less complex activities.

Our analysis will thus help answering the following questions: Which firms ask for a more heterogeneous skill set? Old or young? Big or small? Innovators or non-innovators? Exporters or domestic producers? Do higher-paying firms demand for employees with more diverse skill sets? In the following we address these research questions employing as dependent variable $SDI_{i,t}$ a measure of skill diversity for firm i at time t, which we quantify by means of Shannon entropy. More in detail, for each firm i we compute a normalized skill vector, whose component $p_{i,s}$ represents the number of job ads posted by firm i requiring skill s, $n_{i,s}$, divided by the total number of skills required by the firm, N_i:

$$p_{i,s} = \frac{n_{i,s}}{N_i} = \frac{n_{i,s}}{\sum_s n_{i,s}} \tag{4}$$

Therefore, the skill diversity of firm *i* at time t is defined as follows:

$$SDI_{i,t} = -\sum_s p_{i,s} log\, p_{i,s} \tag{5}$$

$SDI_{i,t}$ is equal to zero for firms that demand only a single specific skill, while is maximal for firms demanding all skills with the same frequency. We therefore estimate firm demand for a diverse skill set with the following model:

$$SDI_{it} = \beta_1 SIZE_{it-1} + \beta_2 AGE_{it-1} + \beta_3 WAGES_{it-1} + \beta_4 PFT_{it-1} + \beta_5 FGR_{it-1} + \beta_6 R\,\&\,D_{it-1} + \beta_7 EXP_{it-1} + b^f + l_{it} + o_{it} + d_t + e_{it} \tag{6}$$

where the independent variables, considered in the previous year $t-1$, include several characteristics of firm i: size ($SIZE_{it-1}$), age (AGE_{it-1}), wages ($WAGES_{it-1}$), firm-level spending on R&D ($R\&D_{it-1}$), exporting (EXP_{it-1}), and as performance variables profitability (PFT_{it-1}) and sales growth (FGR_{it-1}). A more detailed definition of the variables is provided in Table 1 above. We also include controls for the location, i.e. the city where the advertised job will take place (l_{it}), occupation at 2-digit ISCO-08 level (o_{it}), time (d_t) dummies and firm fixed effects (b^f). Using firm fixed effects allows us to control for any time-invariant unobserved variables, such as the sector in which the firm operates. This enables us to capture the effect of variables that do vary over time, as the different independent variables in our estimations.

Table 3 shows the results using a fixed effect estimation, with different model specifications. The first two columns present the regressions without the variable firm growth, with the specification presented in the first column including time and occupation dummies. The specification in the second column also includes location dummies. The third and fourth columns present the same specification of the first two columns with the exception that we extend the model including also firm growth, defined as the sales log difference over two consecutive years.

▶ Table 3: Skill Diversification Index and firm characteristics

	(1)	(2)	(3)	(4)
Log Sales	0.2231***	0.1975***	0.2316***	0.1866***
	(0.0464)	(0.625)	(0.0450)	(0.0578)
Log Age	-0.3984***	-0.3994**	-0.4093**	-0.4611**
	(0.1520)	(0.0065)	(0.1615)	(0.2169)
Wage Intensity	0.0155***	0.0123*	0.0146***	0.0128**
	(0.0053)	(0.0065)	(0.0051)	(0.0061)
Profitability	-0.2277***	-0.2797***	-0.2149***	-0.2921***
	(0.0517)	(0.0769)	(0.0535)	(0.0661)
R&D Intensity	0.0946***	0.1058***	0.1208***	0.1486***
	(0.0311)	(0.0329)	(0.0379)	(0.0392)
Export Intensity	0.1151***	0.1082***	0.1236***	0.1086***
	(0.0273)	(0.0329)	(0.0283)	(0.0359)
Firm Growth			-0.2365	-0.4328*
			(0.1577)	(0.2509)
Time dummies	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes
Location dummies	No	Yes	No	Yes
Observations	2453	2453	2397	2397
R^2	0.171	0.292	0.181	0.304
Number of firms	1748	1748	1714	1714

Notes: The table shows the results of a panel regression at the firm level, following specification (6). Standard errors were clustered at the firm level and are shown in parenthesis. *, **, and *** denote significance at the 1, 5, and 10 per cent level. Dependent Variable: Skill Diversification Index; right hand side (RHS) variables are lagged by one year.

The results we observe are robust to the various specifications. Concerning the baseline model, we observe that age shows a negative and significant relationship with skill diversification in all specifications, that is younger firms are more likely to demand more diverse skills. Wages show

a strongly positive relation with the SDI index, suggesting that firms that demand more skill diverse jobs also pay higher wages. Among the performance variables, we do not find any positive relationship between either firm growth or profitability and demand for more skill diverse jobs. At times, the relationship is negative and significant. This could be due to the fact that, in a shorter time span, firms might be investing in new (and costly) activities related to the demand for more diverse skills, activities that might give returns only in a longer time span. Note that the analyses we present here are all of short-run nature, and the long-run effect of skill diversity on performances could be different. Concerning the extended firm-level controls, increased skill diversity is positively associated to R&D and exporting, i.e., higher levels of firm spending on R&D and export market performance are linked to higher diversity in skills. This is very interesting, since these are considered to be complex firm activities (Coad et al., 2021), associated with idiosyncratic patterns of learning by firms (Dosi et al., 2015). The results we observe suggest that what workers do at workplace significantly differ depending on the activities undertaken by the respective organization.

Skill typologies and firm outcomes

In addition to estimating which firms require more diversified skill sets, we now investigate the relationship between the demand for specific typologies of skill sets and firm-level outcomes. Therefore, similarly to Deming and Kahn (2018)[9] to take into account diversified skill requirements, we construct our independent variable $SKILL_{i,t}$ as the probability that a job advertised by firm i at time t requires a skill belonging to four main categories based on the skill taxonomy developed by the ILO, namely: Cognitive, Digital, Socio-emotional and Manual, and the various combinations of the above. In particular, we set $SKILL_{i,t}$ to take values in the following 10 categories: [*Cognitive, Digital, Socio-emotional, Manual, Cognitive & Digital, Cognitive & Socio-emotional, Cognitive & Manual, Digital & Socio-emotional, Digital & Manual, Socio-emotional & Manual*]. With the aim of understanding the relationship between firm-level outcomes and these different skill requirements, we estimate the following econometric specifications:

$$Y_{it} = \beta X_{it-1} + \delta SKILL_{it-1} + b^f + l_{it} + o_{it} + d_t + \eta_{it} \quad (7)$$

where Y_{it} represents the considered dimensions of firm i's performance at time t, namely: wages, sales growth, relative profitability, exporting intensity and R&D spending; $X_{i,t-1}$ represents firm size at time t – 1; $SKILL_{it-1}$ is the skill typology vector at time t – 1 defined above; l_{it}, o_{it}, d_t, and b^f are location, occupation, time and firm fixed effects. Note that to avoid spurious correlation between the intensity variable and the variable controlling for firm size, when the dependent variable is defined in terms of intensities, such as wage intensity (total wages over total sales), we control for firm size using the log of assets and not sales, since they enter the dependent variable. These details are clearly reported in the respective tables.

Tables 4 – 8 report the relationships respectively between wages, profitability, sales growth, exporting and R&D, and different skill requirements, according to Equation 7. The first column of each table presents the baseline specification with only the four main skill categories, namely cognitive, digital, socio-emotional and manual. The specifications in columns 2-6 include also the different combinations of the four main skill categories, as detailed above.

In Table 4, with wage intensity as dependent variable, we can observe that the estimated coefficients of digital and socio-emotional skills are positive and significant in all model specifications,

[9] Unlike Deming and Kahn (2018), we also take into consideration the category of manual skills, and separate cognitive skills into digital and other cognitive skills.

albeit socio-emotional skills are significant only at 10% level when considering also the interaction of digital and socio-emotional skills in column (3) and all the skill combinations in column (6). Digital skills show a statistically significant coefficient at the less than 1% significance level between 0.009 and 0.01 in all specifications, suggesting that the demand for digital skills by firms is associated with a 0.1% increase in wage intensity. In contrast, the coefficient of manual skills is consistently negative and significant showing that firms requiring manual skills offer lower wages. However, when manual skills are complemented with digital skills, it leads to higher wage intensity. Similarly, a combination of digital and socio-emotional skills is also associated to higher wage intensity. Furthermore, we do not observe any significant relationship between cognitive skills alone and wage intensity, while socio-emotional skills are significantly and positively associated with higher wage intensity.

Concerning firm performance measures, Table 5 shows the results for firm profitability. According to our analysis, both digital and socio-emotional skills are associated to lower firm profitability. This may suggest that firms that are looking for digital skills are investing heavily in technology, in other words, are in the process of technology adoption that may lead to lower profitability in the short run. Cognitive skills report a positive coefficient that is significant at 10% level in columns (2)-(4). While manual skills show positive and significant coefficients between 1.0 and 1.2 in all model specifications, considering also the complementary role of different skill typologies, the results suggest that firms requiring manual skills and a mixture of other competencies are more profitable, especially in the case of the interaction of manual or socio-emotional skills with digital skills.

Instead, as can be observed in Table 6, digital, socio-emotional and manual skills are associated with higher firm growth, while cognitive skills show a negative and significant relationship to growth. By looking at the different skill combinations, we observe that both the combination of manual-digital skills, and cognitive-social skills are associated with higher firm growth. These findings, together with the results on profitability, suggest that firms that invest in digital skills might suffer on profitability owing to their high investments, while these investments are boosting short-run sales growth.

Table 7 presents the estimation results with export intensity as dependent variable. In this case, the demand for digital skills appear to be important for performance in the export market, where effects are statistically significant at the 1% level in all specifications. The magnitude of the coefficient is also high: firms' demand for digital skills is related to a 1.4% increase in export intensity. Similar is the case for cognitive skills, with firms demanding cognitive skills being more likely to show higher export intensity. Contrarily, manual skills do not exhibit any relationship with exporting intensity. However, when manual skills are combined with cognitive or digital skills, they are related to higher exporting intensity.[10]

Table 8 presents the results with R&D intensity as dependent variable. Similarly to the case of exporting intensity, digital skills are positively associated with the innovative activities of firms. However, here, the magnitude of their effect is the highest: demand for digital skills is associated with around a 2.4% increase in the intensity of the R&D activity undertaken by firms. Likewise, cognitive skills are associated positively with R&D intensity, and the same is true for socio-emotional skills, albeit with slightly lower magnitude. The coefficient for manual skills are negative and significant, suggesting that more R&D intense firms have a lower demand for manual skills.[11]

The results presented in this section complement our previous finding on the important role played by firm-level characteristics in the changing nature of jobs and skills. Our evidence suggests that the skills demanded by firms are indeed related to their different activities, learning

[10] We obtain similar results also with export dummy as dependent variable, that takes value 1 in case of exporting and 0 otherwise.
[11] We obtain similar results also with a R&D dummy as dependent variable, that takes value 1 if the firm spends on R&D and 0 otherwise.

processes and routines. A striking result is the positive association of digital skills with wages, firm growth, exporting and R&D intensity, suggesting that competitive firms close to the technological frontier have higher demand for digital skills. Furthermore, manual jobs are associated with lower wages but higher growth, and the combination of manual and digital skills brings higher wages and can also be positively associated to exporting. A significant element of novelty of our analysis is to explicitly look at manual skills and analyze separately cognitive and digital skills and thus to disentangle their different relationships with firm characteristics and performance.

▶ Table 4: Wage Intensity and Skill Requirements

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive	0.0002	-0.0002	0.0008	0.0001	0.0004	0.0006
	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0029)	(0.0029)
Digital	0.0102***	0.0098***	0.0093***	0.0088***	0.0089***	0.0092***
	(0.0023)	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)
Social	0.0043**	0.0044**	0.0036*	0.0045**	0.0041**	0.0037*
	(0.0019)	(0.0020)	(0.0020)	(0.0020)	(0.0019)	(0.0020)
Manual	-0.0056*	-0.0104***	-0.0106***	-0.0098***	-0.0100***	-0.0097***
	(0.0031)	(0.0034)	(0.0033)	(0.0034)	(0.0034)	(0.0034)
Cognitive & Digital		0.0012	-0.0072**			-0.0099***
		(0.0017)	(0.0034)			(0.0036)
Cognitive & Social		0.0001			-0.0015	0.0010
		(0.0027)			(0.0028)	(0.0031)
Cognitive & Manual		0.0037***		0.0005		0.0049
		(0.0011)		(0.0054)		(0.0059)
Digital & Social			0.0074**		0.0035**	0.0080**
			(0.0033)		(0.0018)	(0.0037)
Digital & Manual			0.0045***	0.0073***		0.0089***
			(0.0011)	(0.0024)		(0.0033)
Social & Manual				-0.0036	0.0035***	-0.0097*
				(0.0054)	(0.0011)	(0.0058)
Firm Size (Log of assets)	0.0112***	0.0113***	0.0113***	0.0113***	0.0113***	0.0113***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Location dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,386	7,386	7,386	7,386	7,386	7,386
R^2	0.029	0.029	0.030	0.030	0.030	0.030
Number of firms	3902	3902	3902	3902	3902	3902

Notes: The table shows the results of a panel regression at the firm level, following specification (7). Standard errors were clustered at the firm level and are shown in parenthesis *, **, and *** denote significance at the 1, 5, and 10 per cent level. Dependent Variable: Wage Intensity; right hand side (RHS) variables are lagged by one year.

▶ Table 5: Firm Profitability and Skill Requirements

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive	0.1336	0.1759*	0.1718*	0.1792*	0.1550	0.2058**
	(0.1004)	(0.1007)	(0.1006)	(0.1002)	(0.1012)	(0.1022)
Digital	-0.5382***	-0.5048***	-0.4861***	-0.4874***	-0.5509***	-0.5242***
	(0.0820)	(0.0866)	(0.0858)	(0.0838)	(0.0866)	(0.0870)
Social	-0.1967***	-0.2048***	-0.2286***	-0.1936***	-0.1794***	-0.2189***
	(0.0695)	(0.0695)	(0.0702)	(0.0694)	(0.0694)	(0.0704)
Manual	0.1075	1.1891***	1.0373***	1.1732***	1.1835***	1.1883***
	(0.1103)	(0.1227)	(0.1202)	(0.1226)	(0.1227)	(0.1231)
Cognitive & Digital		0.0230	-0.0661			-0.2076*
		(0.0619)	(0.1159)			(0.1256)
Cognitive & Social		0.0751			0.0130	-0.0431
		(0.0956)			(0.0987)	(0.1065)
Cognitive & Manual		-0.7636***		-0.5507***		-0.3211
		(0.0391)		(0.1908)		(0.2084)
Digital & Social			0.4165***		0.1079*	0.3542***
			(0.1131)		(0.0623)	(0.1282)
Digital & Manual			-0.7515***	-0.0450		-0.1747
			(0.0399)	(0.0858)		(0.1146)
Social & Manual				-0.1637	-0.7609***	-0.2891
				(0.1866)	(0.0391)	(0.2010)
Firm Size	0.0656***	0.0553***	0.0560***	0.0550***	0.0558***	0.0554***
	(0.0087)	(0.0085)	(0.0086)	(0.0085)	(0.0086)	(0.0085)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Location dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,415	7,415	7,415	7,415	7,415	7,415
R^2	0.057	0.026	0.028	0.025	0.027	0.026
Number of firms	3916	3916	3916	3916	3916	3916

Notes: The table shows the results of a panel regression at the firm level, following specification (7). Standard errors were clustered at the firm level and are shown in parenthesis *, **, and *** denote significance at the 1, 5, and 10 per cent level. Dependent Variable: Firm Profitability; RHS variables are lagged by one year

▶ **Table 6: Firm Growth and Skill Requirements**

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive	-0.1821***	-0.2058***	-0.1829***	-0.1938***	-0.2071***	-0.1968***
	(0.0229)	(0.0231)	(0.0230)	(0.0229)	(0.0230)	(0.0232)
Digital	0.0486***	0.0481***	0.0342**	0.0259	0.0431**	0.0420**
	(0.0168)	(0.0173)	(0.0172)	(0.0172)	(0.0172)	(0.0173)
Social	0.0887***	0.0863***	0.0823***	0.0937***	0.0880***	0.0796***
	(0.0147)	(0.0148)	(0.0149)	(0.0148)	(0.0147)	(0.0149)
Manual	0.1430***	0.0855***	0.0716***	0.0727***	0.0865***	0.0842***
	(0.0253)	(0.0265)	(0.0263)	(0.0264)	(0.0264)	(0.0264)
Cognitive & Digital		-0.0436**	-0.1767***			-0.1889***
		(0.0171)	(0.0335)			(0.0387)
Cognitive & Social		0.1495***			0.1330***	0.1839***
		(0.0280)			(0.0296)	(0.0323)
Cognitive & Manual		0.0787***		-0.0452		0.0013
		(0.0095)		(0.0541)		(0.0628)
Digital & Social			0.1297***		-0.0150	0.0649*
			(0.0319)		(0.0172)	(0.0392)
Digital & Manual			0.0930***	0.0651***		0.1872***
			(0.0099)	(0.0245)		(0.0342)
Social & Manual				0.0671	0.0761***	-0.0972*
				(0.0502)	(0.0095)	(0.0576)
Firm Size	-0.1051***	-0.1076***	-0.1078***	-0.1080***	-0.1077***	-0.1080***
	(0.0031)	(0.0031)	(0.0031)	(0.0031)	(0.0031)	(0.0031)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Location dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,061	2,061	2,061	2,061	2,061	2,061
R^2	0.179	0.185	0.185	0.185	0.185	0.187
Number of firms	1515	1515	1515	1515	1515	1515

Notes: The table shows the results of a panel regression at the firm level, following specification (7). Standard errors were clustered at the firm level and are shown in parenthesis *, **, and *** denote significance at the 1, 5, and 10 per cent level. Dependent Variable: Firm Growth; RHS variables are lagged by one year

▶ Table 7: Export Intensity and Skill Requirements

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive	0.0849***	0.0850***	0.0861***	0.0915***	0.0868***	0.0912***
	(0.0157)	(0.0158)	(0.0157)	(0.0159)	(0.0159)	(0.0160)
Digital	0.1467***	0.1426***	0.1420***	0.1394***	0.1407***	0.1362***
	(0.0134)	(0.0141)	(0.0140)	(0.0142)	(0.0139)	(0.0144)
Social	-0.0386***	-0.0411***	-0.0418***	-0.0433***	-0.0423***	-0.0452***
	(0.0105)	(0.0106)	(0.0107)	(0.0108)	(0.0106)	(0.0109)
Manual	-0.0145	0.0002	-0.0060	-0.0003	0.0001	-0.0006
	(0.0223)	(0.0239)	(0.0235)	(0.0239)	(0.0236)	(0.0241)
Cognitive & Digital		0.0188	-0.0010			-0.0324
		(0.0121)	(0.0204)			(0.0269)
Cognitive & Social		0.0239			0.0188	0.0399
		(0.0172)			(0.0196)	(0.0244)
Cognitive & Manual		-0.0146**		-0.0650**		0.0095
		(0.0073)		(0.0308)		(0.0479)
Digital & Social			0.0311*		0.0245**	0.0356
			(0.0183)		(0.0122)	(0.0273)
Digital & Manual			-0.0113	0.0410**		0.0386
			(0.0077)	(0.0174)		(0.0256)
Social & Manual				0.0143	-0.0162**	-0.0606
				(0.0252)	(0.0073)	(0.0418)
Firm Size (Log of assets)	0.0456***	0.0457***	0.0461***	0.0457***	0.0460***	0.0463***
	(0.0032)	(0.0032)	(0.0032)	(0.0032)	(0.0032)	(0.0032)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Location dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7415	7415	7415	7415	7415	7415
R^2	0.153	0.154	0.154	0.154	0.155	0.156
Number of firms	3.916	3.916	3.916	3.916	3.916	3.916

Notes: The table shows the results of a panel regression at the firm level, following specification (7). Standard errors were clustered at the firm level and are shown in parenthesis *, **, and *** denote significance at the 1, 5, and 10 per cent level. Dependent Variable: Export Intensity; RHS variables are lagged by one year

▶ Table 8: R&D Intensity and Skill Requirements

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive	0.2685***	0.2414***	0.2625***	0.2660***	0.2596***	0.2605***
	(0.0682)	(0.0686)	(0.0684)	(0.0683)	(0.0686)	(0.0690)
Digital	0.2475***	0.1936***	0.1928***	0.2372***	0.1882***	0.1965***
	(0.0531)	(0.0545)	(0.0543)	(0.0541)	(0.0543)	(0.0546)
Social	-0.3463***	-0.3330***	-0.3485***	-0.3472***	-0.3428***	-0.3454***
	(0.0447)	(0.0450)	(0.0451)	(0.0449)	(0.0449)	(0.0452)
Manual	-0.1758**	-0.2088***	-0.1862**	-0.2152***	-0.2117***	-0.2126***
	(0.0713)	(0.0746)	(0.0741)	(0.0744)	(0.0744)	(0.0745)
Cognitive & Digital		0.2210***	-0.0498			0.0674
		(0.0492)	(0.1011)			(0.1148)
Cognitive & Social		0.0598			-0.0349	-0.1587
		(0.0874)			(0.0915)	(0.0994)
Cognitive & Manual		0.0333		-0.3725**		0.0627
		(0.0279)		(0.1787)		(0.2036)
Digital & Social			0.3077***		0.2658***	0.3660***
			(0.0965)		(0.0493)	(0.1157)
Digital & Manual			0.0061	0.0998		-0.3181***
			(0.0288)	(0.0679)		(0.0984)
Social & Manual				0.3351**	0.0370	0.2673
				(0.1698)	(0.0279)	(0.1903)
Firm Size (Log of Assets)	0.0435***	0.0417***	0.0414***	0.0411***	0.0407***	0.0418***
	(0.0118)	(0.0118)	(0.0118)	(0.0118)	(0.0118)	(0.0118)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Location dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7415	7415	7415	7415	7415	7415
R^2	0.051	0.052	0.053	0.051	0.053	0.054
Number of firms	3,916	3,916	3,916	3,916	3,916	3,916

Notes: The table shows the results of a panel regression at the firm level, following specification (7). Standard errors were clustered at the firm level and are shown in parenthesis *, **, and *** denote significance at the 1, 5, and 10 per cent level. Dependent Variable: R&D Intensity; RHS variables are lagged by one year

▶ Concluding remarks

In this paper we investigated the heterogeneous demand for skills within occupations across firms, and the firm activities that relate to it, using online job vacancies posted in India between 2016 and 2020. By relying on the free-text data from Indian Naukri job portal, using a multi-level machine learning empirical strategy, and a novel skill taxonomy and a methodology to implement it in the vacancy data developed by the ILO (Bennett et al., 2022), we extract information on occupational skills and match them with information about the hiring firm drawn from the Prowess firm-level data-set.

Firstly, we find strong evidence of within-occupation heterogeneity in skill demand across firms, which holds true for all the analyzed occupations. We empirically quantify this heterogeneity by employing a method based on cosine similarity that allows for an in-depth distributional analysis across different skills. *Secondly*, we document that firms demanding more diversified skills are younger, pay higher wages, and are dynamic organizations with high engagement in complex activities, such as exporting and R&D investment. *Thirdly*, we observe that specific types of skills are associated with different firm activities and outcomes. Digital skills appear to be associated with higher wage intensity, while there is no significant relationship between other cognitive skills and wage intensity. Manual skills are positively linked to firm growth, and to lower wages, however, we detect a wage premium when manual skills are complemented by digital skills. Further, disentangling between digital and cognitive skills enables us to also shed light on the different roles they play in firm dynamics. Our results confirm that highly innovative and internationally competitive firms demand both digital and other cognitive skills, which, together with the combination of digital and socio-emotional skills, show positive and significant associations to higher growth, engagement in exporting and innovation. The positive association between high export intensity and higher demand for other digital and cognitive skills is expected. To remain competitive in international markets and to maintain their global value chain positioning, firms might be pushed to upgrading, adopting new digital technologies and acquiring complementary cognitive competences. However, other cognitive skills alone are non-significant for wages and profitability in the short-run, likely indicating that they might take longer to translate into positive outcomes.

Our evidence suggests that different firms require different skills, even for the same occupation. This suggests that it is not simply the pace of technological change that shapes the job skill-content, but also organizational characteristics are important drivers of the within occupational demand for skills. Therefore, when analysing the changing world of work and, especially in the debate on skilled/unskilled or routine/non-routine labour, the role of organizations, their structures and routines should be appropriately accounted for. Technological trajectories, innovation and skills co-evolve and are affected by the ways in which firms organise activities and production processes, and by the economic and institutional set-up they develop in. The skill requirements of similar occupations may not be as universal as in the common interpretation of O*NET based studies (Acemoglu and Autor, 2011; Autor, 2013), they could be time-varying, context- and organization-specific. If occupational skill requirements are indeed firm-specific, this present a compelling case for firms to play a crucial role in training their workforce, rather than relying merely on finding pre-trained and specialized candidates. In fact, while labor market policies, especially education and training policies, might certainly be beneficial for workers in overcoming skill obsolescence, these are far from sufficient. As pointed out also by Dosi and Virgillito (2019), firms should be incentivized to invest in enhancing employees' learning via on-the-job training schemes, which can be tailored to the organization's complex needs. In this scenario, the role of workers' organizations will be extremely important in pushing for organizational investments in labour force training and up- or re-skilling. This may help workers to cope with technological advancements in their firms and hence help to re-distribute benefits from technological advances and productivity gains.

Annex

A Data structure

In table A1 we present a schematic structure of the online job posts. As reported in the first column, the different fields are Job Title, Job Description, Post Date, Company, Skills, Education, Industry, Job Location, Experience and Pay-rate. The Field description, as reported in the second column describes the various fields. The sub-fields within the "Job Description" includes Role Category, Role, Key Skills and Company Profile.

B Mapping of 2-digit occupations

To create variables for occupations at the 2-digit ISCO-08 level, we start with data from 2016 and 2017 and consider the highly disaggregated "Role" sub-field. We organize all the roles into a single list with 678 unique words, taking into consideration only those roles that appear at least 10 times in the data. We then match the roles with 2 digit ISCO-08 codes following some intuitive rules: (i) From the official ISCO-08 repository[12], we consider the description of the around 6000 occupations at the 4-digit level. If the naukri role corresponds exactly to an ISCO-08 occupation title (4-digit level), a direct match is established. (ii) When we are not able to classify a 4-digit occupation but the match with a 2-digit index is clear-cut, we directly classify it at 2-digit level. (iii) Otherwise, we cross-check with the more detailed description of occupational titles provided by the SOC-2020 classification[13], which can easily be reclassifyied into ISCO-08. (iv) If we are still not able to find any correspondence, we proceed with a manual Google search and/or a more thorough manual inspection of the naukri data. Finally, we obtain our correspondence table between naukri role categories and ISCO-08 2-digit occupations.

To then refine our classification of the 2-digit occupations and extend it to years 2019 and 2020, we perform the following steps:

1. *Pre-processing of job titles*

To implement the machine learning approach, we process the sample. Firstly, we discard job vacancies that do not contain the "Job Title" field. Secondly, we clean all 2016-2017 free-text job titles by eliminating capital letters, numbers, punctuation, symbols, prepositions, stop words and commonly used sentences that we deem as noise (for instance, the initial sentence of a large number of ads begins with "we are writing for. . . "). This is also done for the subset of adverts that were classified through the computational mapping, a balanced sub-sample where all naukri role categories are univocally matched with 2-digit ISCO-08 codes, and to which we will refer to as the 2016-2017 labeled data (16-17LD). The classes of our classification are 2-digit ISCO- 08 sub-major groups and each of these is populated by a number of job titles. To avoid any bias in the classification, by assigning class weights inversely proportional to their times of appearance, to train the NLP algorithm we create the balanced 16-17LD data-set, where we have about 5,000 examples of job adverts for each 2-digit ISCO-08 class. Thirdly, by jointly considering all the words that appear in the ISCO-08 and SOC-2020 Index of occupational titles, we create a list of about 8,000 relevant words – the relevant word vector. We then further clean all the naukri job titles in the 16-17LD and remove from the balanced set all the words that do not appear in the relevant word vector.

[12] https://www.ilo.org/public/english/bureau/stat/isco/isco08/.
[13] https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020.

▶ **Table A1: Schematic structure of naurki.com online job post**

Field	*Field description*	*Sub-field*	*Sub-field description*
Job Title	Occupational title with varying degree of detail (e.g. "Large Molecule Analyst", "Testing Analyst", Occupational title with varying degree of detail (e.g. "Large Molecule Analyst", "Testing Analyst", "Applications Developer") or alternatives (e.g. "Sales Engineer (electrical)/Sr. Sales Engineers/ asst. Sales Manager")		
Job Description	Detailed free text job description of variable length that contains the sub-fields in the following column	***Role category***	Synthetic description of the role similar to ISCO-08 sub-major groups (2-digits)
		Role	More detailed description of the role, similar to ISCO-08 unit groups (4-digits)
		Key skills	Required skills description, also detailing required educational attainment (under- and post-graduate degree or Ph.D., e.g. "UG:BSc – Any specialisation, PG:M.Sc - Any Specialization, Bio-Chemistry, Chemistry") or lack of (e.g. "PG:Graduation Not Required", "Doctorate: Doctorate Not Required") and working experience
		Company Profile	Name of the hiring company or of the recruitment agency and the company for which it is looking for candidates; company description of varying length
Post Date	Year/Month/Day and posting time		
Company	Name of the hiring company or of the recruitment agency and the company for which it is looking for candidates		
Skills	Short skill category (e.g. "Medical", "Sales", "IT Software - Application Programming". . .)		
Education	Required educational attainment (or lack of, e.g. "Graduation Not Required") and field of study		
Industry	Hiring company industry, can provide different alternatives with varying degree of detail (e.g. "Pharma/Biotech/Clinical Research", "Electricals/ Switchgears", "IT-Software/Software Services)		
Job Location	Indian city or aera (e.g. "Bengaluru/Bengalore", "Mumbai suburbs")		
Experience	Required years of working experience		
Pay-rate	Wage range *per annum* in INR (may also depend on the working experience and credentials of the candidate)		

2. *Conversion to vector with word embedding*

By relying on the fastText library[14], we then transform the clean job titles in vectors by using word embedding – i.e., a representation of words as real-valued vectors that encode their meaning such that words that are closer in the vector space are expected to have similar meaning.

3. *Train-test split*

Thanks to the large size of our sample, by using an ad hoc Python library, we split the 16-17LD into two sub-sets: the training set (around the 70 per cent of the 16-17LD), that will be used to train the support vector machine, and the test set (around 30 per cent of the 16-17LD), that will be used to assess the performance of the machine.

4. *Training of Support Vector Machine classifier*

After processing and splitting the 16-17LD data, we feed the train set to the SVM classifier. The latter removes noise from the data-set and provides a more accurateclassification, that helps us discriminate cases where the ISCO-08 2-digit occupational code identified through the naukri role correspondence table actually diverges from the content of the job title.

5. *Evaluation and classification of unlabeled job posts*

For each job advert, the SVM classifier outputs an ordered list containing the probabilities that the job advert belongs to each possible class (i.e., 2-digit ISCO-08 sub-major groups), from the most to the least probable. Bearing in mind that classes with probabilities higher than 60 per cent are not to be found, a match between an online vacancy and ISCO-08 code is established only when its probability is higher than 40 per cent.

A final manual double-check is then performed, also comparing the SVM and the naukri role data classification. In fact, also when using the SVM it is possible to obtain misclassifications, especially in a number of problematic classes. For instance, often the algorithm misclassifies "Protective services workers" (ISCO-08 code 54) as armed forces (ISCO-08 sub-major group 03), for which virtually no vacancies are reported in the naukri portal. A similar misclassification issue appears for the term engineer, that is very common and may appear also in job titles where the role actually being sought is electronic operator or electrician. This is especially the case for "Electrical and electronic trades workers" (ISCO-08 sub-major group 74). Therefore, with the aim of avoiding any of such inconsistencies when running the SVM on the whole sample in the next step of the pipeline, the sub-major groups 03, 74, 83 and 93 are excluded from the 16-17LD balanced sample, and the SVM is retrained on 24 ISCO-08 sub-major groups rather than 28.

6. *Execution of the SVM algorithm for the whole naukri.com data-set*

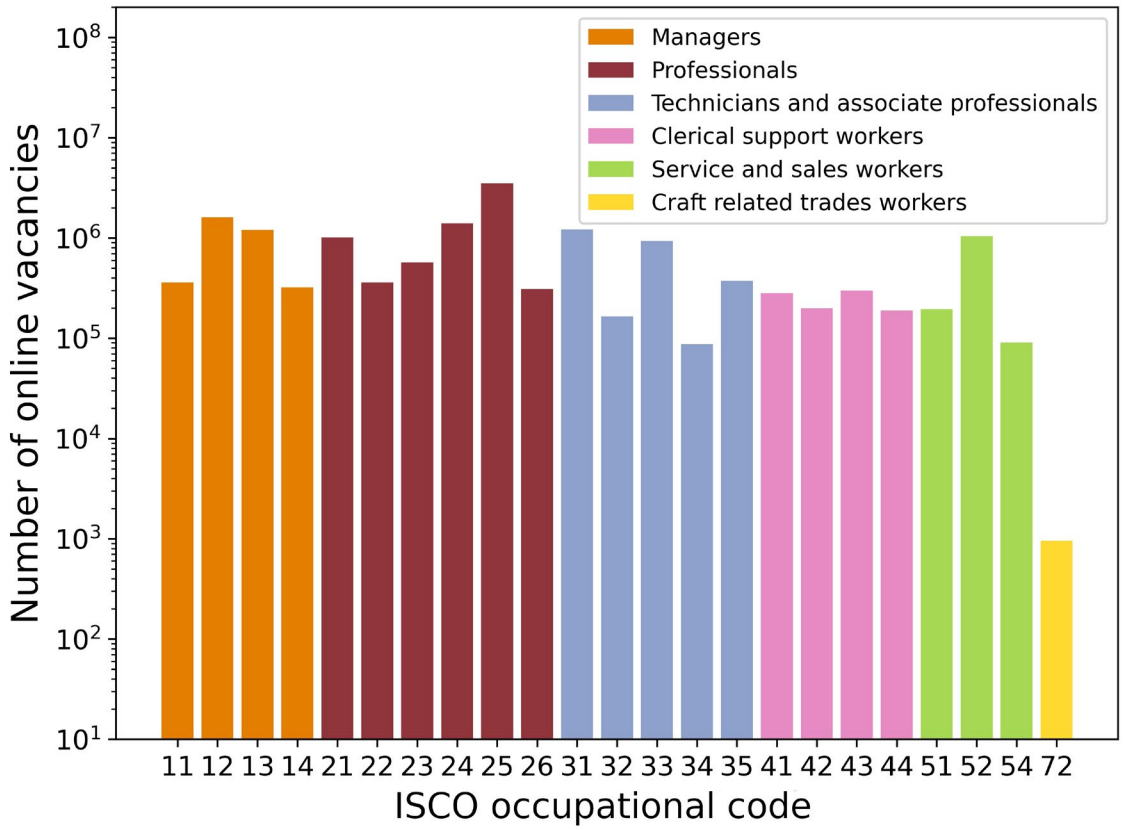
Carrying out this set of operations on the 16-17LD has allowed us to obtain a new correspondence table of job titles to ISCO-08 2-digit codes. Using this set of clean job titles univocally classified, to label the remaining vacancies: we carry out the same procedure on the whole 2016-2017 and 2019-2020 data-sets. By following these steps, we are able to classify the 71 per cent of the whole sample – 11.1 out of 15.6 million online vacancies.

[14] fastText is an open-source library created by Facebook AI Research (FAIR) lab for word embedding and text classification in over 150 languages. It can be built as a Python module and is available at: https://fasttext.cc/.

C Representativeness of the vacancy data

Figure 3 reports the number of vacancies for each 2-digit ISCO-08 occupation. For each ISCO-08 broad category the occupational shares of the online listings are the following: 57 per cent for professionals; 29 per cent for managers; 10 per cent for technicians & associate professionals; 2 per cent for service and sales workers; 1 per cent for clerical support workers; 0.8 per cent for craft and related trades workers; 0.2 per cent for elementary occupations; while no listings are found for skilled agricultural, forestry, fishery workers, plant & machine operators, assemblers and armed forces occupations. We thus observe that a large number of the vacancies cover professional occupations – e.g., in finance, consultancy, manufacturingor ICTs –, and in lower percentage managerial ones, whose hiring process probably also employs less publicly advertised channels.

▶ **Figure C1: Occupational shares at the 2-digit level (pooled)**

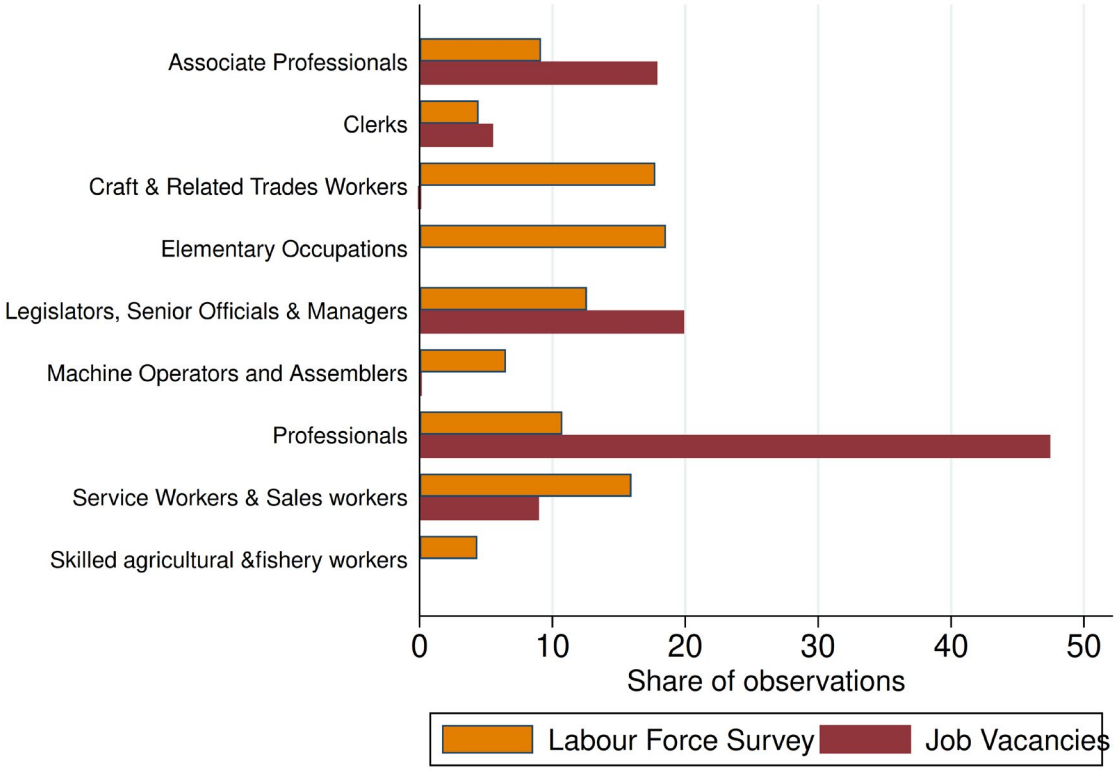


Notes: The figure shows the number of online job posts, considering the entire Naukri sample, for each 2-digit ISCO-08 sub-major group, where the y-axis is displayed in logarithmic scale. Color indicates the corresponding 1-digit ISCO-08 major groups.

To benchmark our data, we rely on The Periodic Labour Force Survey (PLFS), which is a nationally representative labour survey conducted by the Indian National Statistical Office (NSO). Given the time frame of our vacancy data, we use the 2017-18 wave. In their annual report, the Government of India (2019) provides the percentage distributions of workers by broad occupational divisions, namely 1-digit level, which are equivalent to the ISCO 1-digit level occupational classes. They show that the share of the workers in division 1 (legislators, senior officials and managers), division 2 (professionals) and division 3 (technicians and associate professionals), increased from 7.9 per cent to 9.1 per cent for rural males, from 5 per cent to 8.75 per cent for rural females, decreased from 31.15 per cent to 30.45 per cent for urban males and increased from 31.95 per cent to 34.65 per cent for urban females (Government of India, 2019).

In Figure C2 we show the occupational shares at the 1-digit level for both PLFS data and job vacancy data. As expected, the skilled categories of workers have a higher representation in online vacancy listings contrary to the elementary occupations.

▶ Figure C2: Occupational shares in the labour force survey compared with the vacancy data at the 1-digit level (percent)



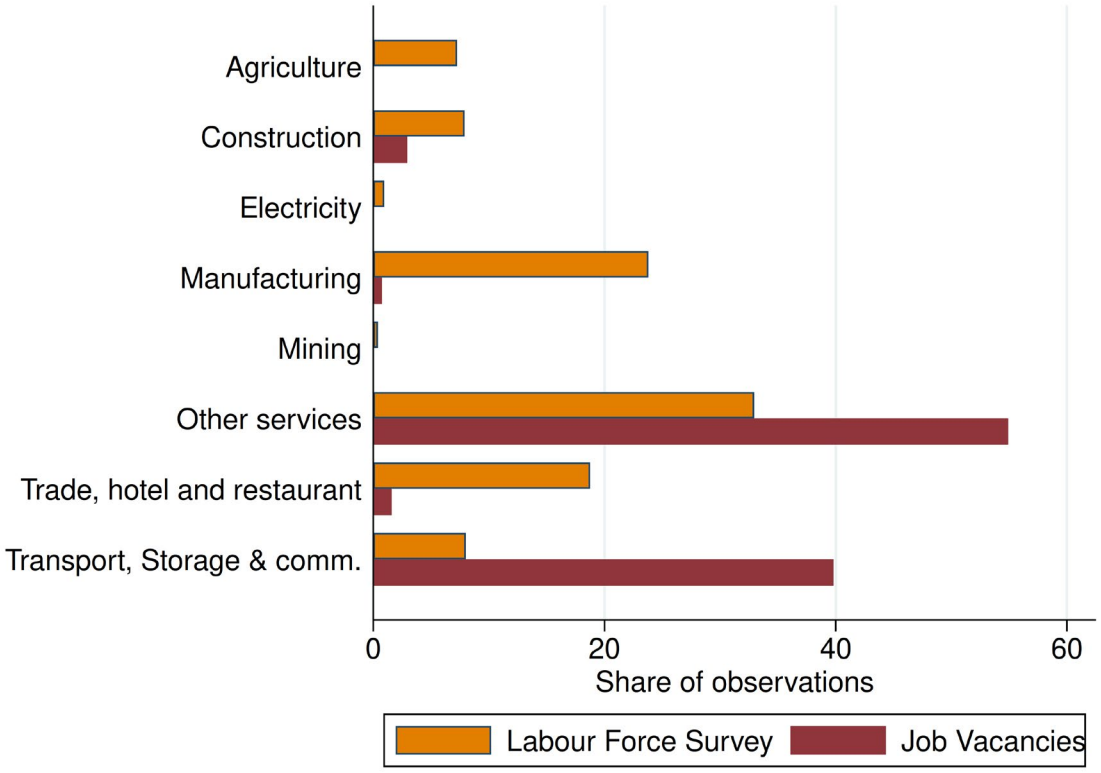
Notes: Occupational shares in the 2017-18 Periodic Labour Force Survey (PLFS) (orange) and Naukri vacancy data (burgundy). The PLFS data are obtained from Government of India (2019), which provides statistics by gender and rural and urban categories. Here, we compute the average of the shares of urban males and urban female categories.

We also look at the sectoral shares of jobs. The PLFS uses the National Industrial Classification (NIC) 2008 and provides data for broad industrial categories. In Figure C3, we show the sectoral shares in the job vacancy data (burgundy) as opposed to the employment figures reported in the Periodic Labour Force Survey (orange) by broad industry division, taking into consideration only urban female and male workers. Figure C3 confirms the absence of agriculture and a very low share of manufacturing jobs in the naukri data. PLFS data shows higher shares also in trade, hotel and restaurant, while for other services (including finance and insurances), transport, storage and communications the number of job ads is more than 20 percentage points higher than the employment share reported in the national survey. It is worth noting that the national survey includes public sector workers, with the vast majority of government jobs still following traditional advertisement channels, such as Government gazettes or newspapers, whereas naukri vacancies include mostly private firms.

The regional distribution of the naukri job listings reveals that over 85 per cent of vacancies are concentrated in large urban areas. Figure C4 reports all cities with above 150,000 inhabitants in which respectively at least 500 vacancies were advertised in 2020. We observe a higher representation of metropolitan cities, with the vast majority of vacancies being located in Bangalore, Delhi and its satellite cities Gurgaon and Noida, Mumbai, Chennai and Pune, all showing over 100,000 advertised vacancies, followed by Kolkata, Chandigarh, Ghaziabad and Hyderabad, with

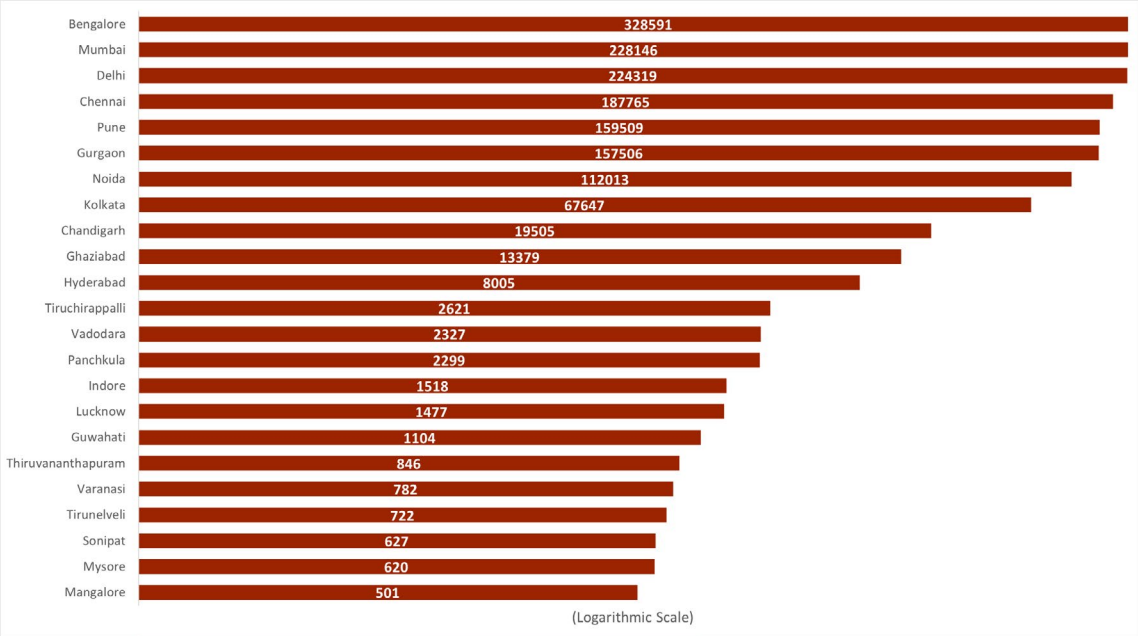
figures of one or two orders of magnitude lower. Overall, these figures indicate that there is substantial regional variation across major cities in India.

▶ **Figure C3: Sectoral shares in the labour force survey compared with the vacancy data at the 1-digit level (percent)**



Notes: Sectoral shares in the 2017-18 Periodic Labour Force Survey (PLFS) (orange) and Naukri vacancy data (burgundy). The PLFS data are obtained from Government of India (2019), which provides statistics by gender and rural and urban categories. Here, we compute the average of the shares of urban males and urban female categories.

▶ Figure C4: Number of vacancies at the city level



Notes: The figure shows the number of job vacancies by Indian city for cities with above 150,000 inhabitants and at least 500 advertised vacancies, 2020 (x-axis in logarithmic scale).

References

Acemoglu, D. and D. Autor (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of Labor Economics*, Elsevier, vol. 4, 1043–1171.

Acemoglu, D. and P. Restrepo (2020): "The wrong kind of AI? Artificial intelligence and the future of labour demand," *Cambridge Journal of Regions, Economy and Society*, 13, 25–35.

Alekseeva, L., J. Azar, M. Giné, S. Samila, and B. Taska (2021): "The demand for AI skills in the labor market," *Labour Economics*, 71, 102002.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2020): "The Evolution of Work in the United States," *American Economic Journal: Applied Economics*, 12, 1–34.

Autor, D. (2013): "The "task approach" to labor markets: an overview," *Journal for Labour Market Research*, 46, 185–199.

Autor, D. H. (2015): "Why Are There Still So Many Jobs? The History and Future of Workplace Automation," *Journal of Economic Perspectives*, 29, 3–30.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006): "The polarization of the US labor market," *American Economic Review*, 96, 189–194.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008): "Trends in US wage inequality: Revising the revisionists," *The Review of Economics and Statistics*, 90, 300–323.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The skill content of recent technological change: An empirical exploration," *The Quarterly Journal of Economics*, 118, 1279–1333.

Autor, H. and D. Dorn (2013): "The growth of low-skill service jobs and the polarization of the US labor market," *American Economic Review*, 103, 1553–97.

Bennett, F., V. Escudero, H. Liepmann, and A. Podjanin (2022): "Using Online Vacancy and Job Applicants' Data to Study Skills Dynamics," *ILO Working Paper 75*.

Blair, P. Q. and D. J. Deming (2020): "Structural increases in skill demand after the great recession," Tech. Rep. 26680, *National Bureau of Economic Research*.

Bottazzi, G., E. Cefis, and G. Dosi (2002): "Corporate Growth and Industrial Structure. Some Evidence from the Italian Manufacturing Industry," *Industrial and Corporate Change,* 11, 705–723.

Bottazzi, G. and A. Secchi (2011): "A new class of asymmetric exponential power densities with applications to economics and finance," *Industrial and Corporate Change*, 20, 991–1030.

Cetrulo, A., D. Guarascio, and M. E. Virgillito (2020): "Anatomy of the Italian occupational structure: concentrated power and distributed knowledge," *Industrial and Corporate Change*, 29, 1345–1379.

Cetrulo, A., A. Sbardella, and M. E. Virgillito (2022): "Vanishing social classes? Facts and figures of the Italian labour market," *Journal of Evolutionary Economics*, 1–52.

Ciarli, T., M. Kenney, S. Massini, and L. Piscitello (2021): "Digital technologies, innovation, and skills: Emerging trajectories and challenges," *Research Policy*, 50, 104289.

Cirera, X., D. Comin, M. Cruz, K. M. Lee, and A. Soares Martins-Neto (2021): "Firm-level technology adoption in Vietnam," *Policy Research Working Paper*, World Bank, Washington, DC, 9567.

Coad, A., N. Mathew, and E. Pugliese (2020): "What's good for the goose ain't good for the gander: heterogeneous innovation capabilities and the performance effects of R&D," *Industrial and Corporate Change*, 29, 621–644.

Coad, A., N. Mathew, E. Pugliese, et al. (2021): "Positioning firms along the capabilities ladder," *UNU-MERIT Working Paper Series*, 2021-031.

Copestake, A., A. Pople, and K. Stapleton (2021): "AI, firms and wages: Evidence from India," Manuscript available at SSRN 3957858.

Costa, S., S. De Santis, G. Dosi, R. Monducci, A. Sbardella, and M. E. Virgillito (2021): "From organizational capabilities to corporate performances: at the roots of productivity slowdown," Tech. rep., *LEM Working Paper Series*, No. 2021/21.

Deming, D. and L. B. Kahn (2018): "Skill requirements across firms and labor markets: Evidence from job postings for professionals," *Journal of Labor Economics*, 36, S337–S369.

Deming, D. J. (2017): "The growing importance of social skills in the labor market," *The Quarterly Journal of Economics*, 132, 1593–1640.

Dosi, G., M. Grazzi, and N. Mathew (2017): "The cost-quantity relations and the diverse patterns of "learning by doing": Evidence from India," *Research Policy*, 46, 1873–1886.

Dosi, G., M. Grazzi, and D. Moschella (2015): "Technology and costs in international competitiveness: From countries and sectors to firms," *Research Policy*, 44, 1795–1814.

Dosi, G. and L. Marengo (2015): "The dynamics of organizational structures and performances under diverging distributions of knowledge and different power structures," *Journal of Institutional Economics*, 11, 535–559.

Dosi, G. and R. R. Nelson (2010): "Technical change and industrial dynamics as evolutionary processes," *Handbook of the Economics of Innovation*, 1, 51–127.

Dosi, G., R. R. Nelson, and S. G. Winter (2000): The nature and dynamics of organizational capabilities, Oxford (UK): Oxford University Press.

Dosi, G. and M. E. Virgillito (2019): "Whither the evolution of the contemporary social fabric? New technologies and old socio-economic trends," *International Labour Review*, 158, 593–625.

Dwivedi, Y. K., L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, J. Edwards, A. Eirug, et al. (2021): "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *International Journal of Information Management*, 57, 101994.

Fernández-Macías, E. and J. Hurley (2017): "Routine-biased technical change and job polarization in Europe," *Socio-Economic Review*, 15, 563–585.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010): "Multiproduct firms and product turnover in the developing world: Evidence from India," *The Review of Economics and Statistics*, 92, 1042–1049.

Goos, M., A. Manning, and A. Salomons (2009): "Job polarization in Europe," *American Economic Review*, 99, 58–63.

Government of India (2019): "Period Labour Force Survey, technical report," National Statistical Office.

Hershbein, B. and L. Kahn (2018): "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings," *American Economic Review*, 108, 1737–72.

Kureková, L. M., M. Beblavý), C. Haita, and A.-E. Thum (2016): "Employers' Skill Preferences across Europe: Between Cognitive and Non-Cognitive Skills," *Journal of Education and Work*, 29, 662–87.

Martins-Neto, A., N. Mathew, P. Mohnen, and T. Treibich (2021): "Is there job polarization in developing economies? A review and outlook," *CESifo Working Paper No. 9444.*

Michaels, G., A. Natraj, and J. Van Reenen (2014): "Has ICT polarized skill demand? Evidence from eleven countries over twenty-five years," *Review of Economics and Statistics*, 96, 60–77.

Mishel, L. and J. Bivens (2017): "The zombie robot argument lurches on: There is no evidence that automation leads to joblessness or inequality," *Economic Policy Institute Working Papers*.

Modestino, A. S., D. Shoag, and J. Ballance (2020): "Upskilling: Do employers demand greater skill when workers are plentiful?" *Review of Economics and Statistics*, 102, 793–805.

Spitz-Oener, A. (2006): "Technical Change, Job Tasks, and Rising Educational Demands: Looking Outside the Wage Structure," *Journal of Labor Economics*, 24, 235–70.

Acknowledgements

This paper has been prepared as part of the "Skills and transitions" research project by the Research Department of the International Labour Organization (ILO). We thank Verónica Escudero, Hannah Liepmann and two anonymous reviewers for their insightful comments. This work was supported also by the Comprehensive Innovation for Sustainable Development (CI4SD) Research Programme of the United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT), Maastricht, the Netherlands. Nanditha Mathew acknowledges the support by the European Union's Horizon 2020 research and innovation program under grant agreement No. 101004703 - PILLARS (Pathways to Inclusive Labour Markets).

Advancing social justice, promoting decent work

The International Labour Organization is the United Nations agency for the world of work. We bring together governments, employers and workers to improve the working lives of all people, driving a human-centred approach to the future of work through employment creation, rights at work, social protection and social dialogue.

Contact details

Research Department (RESEARCH)

International Labour Organization
Route des Morillons 4
1211 Geneva 22
Switzerland
T +41 22 799 6530
research@ilo.org
www.ilo.org/research

ISBN 9789220392706

9 789220 392706