



Can Machines Learn Creativity Needs? An Approach Based on Matrix Completion

Giorgio Gnecco¹ · Sara Landi¹ · Massimo Riccaboni¹

Received: 14 December 2021 / Accepted: 16 June 2022

© The Author(s) 2022

Abstract

Technological progress has been recently associated with a crowding-out of cognitive-skill intensive jobs in favour of jobs requiring soft skills, such as ones related to social intelligence, flexibility and creativity. The nature of soft skills makes them hardly replaceable by machine work and among subsets of soft skills, creativity is one of the hardest to define and codify. Therefore, creativity-intensive occupations have been shielded from automation. Given this framework, our study contributes to a nascent field on interdisciplinary research to predict the impact of artificial intelligence on work activities and future jobs using machine learning. In our work, we focus on creativity, starting from its possible definitions, then we get significant insights on creativity patterns and dynamics in the Italian labour market, using a machine learning approach. We make use of the INAPP-ISTAT Survey on Occupations (ICP), where we identify 25 skills associated with creativity. Then, we apply matrix completion—a machine learning technique which is often used by recommender systems—to predict the average importance levels of various creative skills for each profession, showing its excellent prediction capability for the specific problem. We also find that matrix completion typically underestimates the average importance levels of soft skills associated with creativity, especially in the case of professions belonging to the major

The authors thank INAPP for providing access to the dataset used for the analysis. They also acknowledge two anonymous Reviewers for their suggestions to improve an earlier version of the manuscript. The authors acknowledge partial support from the PAI 2018 project “Technological change, soft skills and future high skilled jobs”, and from the Industry 4.0 competence center ARTES 4.0. They also wish to thank Sotiris Blanas, Andrea Canidio, Massimiliano Franceschetti, and Marco Grazzi for their helpful comments. A preliminary version of the work has been presented at the 62nd Scientific Annual Meeting of the Italian Society of Economics (RSA 2021), October 26th–29th, 2021, online.

✉ Giorgio Gnecco
giorgio.gnecco@imtlucca.it

Sara Landi
sara.landi@imtlucca.it

Massimo Riccaboni
massimo.riccaboni@imtlucca.it

¹ IMT-School for Advanced Studies, Piazza S. Francesco 19, 55100 Lucca, Italy

group of legislators, senior officials and managers, as well as intellectual professionals. Conversely, overestimates are typically obtained for other professions, which may be associated with a higher risk of being automated.

Keywords Creativity and soft skills · Counterfactual analysis · Matrix completion · Labor market · Automation

JEL Classification C44 · C53 · C63 · J24 · J44

1 Introduction

Since the early 2000s, technological progress has been associated with a crowding-out of cognitive-skill intensive jobs in favour of jobs requiring soft skills, such as ones related to social intelligence, flexibility and creativity (Acemoglu and Autor 2011). The growth of occupations characterized by cognitive tasks has been sluggish both in terms of employment and wages (Acemoglu and Autor 2011; Beaudry et al. 2016; Deming 2017a). Moreover, the returns to cognitive skills have not risen (Castex and Dechter 2014).

Some authors suggest that this phenomenon could be ascribed to a progressive extension of the perimeter of the tasks at risk of automation toward those traditionally performed by high-skill workers (Levy and Murnane 2012; Brynjolfsson and McAfee 2014; Remus and Levy 2016; Autor 2014). Indeed, Deming (2017a) shows that it is especially the decrease in Science, Technology, Engineering and Mathematics (STEM) jobs, requiring high technical abilities but low social interactions, which has determined the shrinkage of high-skill/cognitive share of employment and wage bill. Instead, the cognitive occupations that have fared much better are those involving complex human interaction skills (related to communication, collaboration, teamwork, flexibility, creativity, etc.).

Those “social skills” are also referred to in the literature as “soft skills” or “non-cognitive skills” (Deming 2017b). According to Rainsbury et al. (2002), one can define soft skills as either behavioural, human, or interpersonal skills that an individual is required to possess with the aim of applying technical skills and knowledge in her/his workplace. Indeed, James and James (2004) claim that the term “soft skills” represents a novel way to refer to a set of either abilities or talents that a person can bring to the workplace. According to several authors, soft skills can be described as “micro social” skills, further categorized, e.g., as: (1) intrapersonal versus interpersonal skills; (2) personal versus social skills; (3) cognitive skills (Muzio et al. 2007). For what concerns the comparison between hard and soft skills, it is worth remarking that: (1) most people are able to distinguish between hard skills (e.g., working with equipment or software) and soft skills (e.g., interpersonal/intrapersonal focus) with quite relative ease; (2) there is quite a significant difference in learning for the two different cases of hard and soft skills; (3) the majority of positions inside an organization need—for a successful execution of the related work tasks—not only the acquisition of hard skills, but also a significant degree of proficiency in soft skills (Laker and Powell 2011).

This change of paradigm toward forms of organization of work characterized by an increased importance of social interaction skills has been noticed across several industries and occupations. Bessen et al. (2020) show that the adoption of automation technologies extends beyond manufacturing firms, while Katz and Margo (2014) suggest that the capability of combining the possibilities offered by automation and Information and Communication Technologies (ICTs) for product customization with creativity and communication skills in the relationship with customers is the key factor to explain and sustain the revival of artisanal jobs observed in the United States.

The rapidity and pervasiveness with which economic and organizational changes occur determine an increasing need in reading promptly and efficiently the evolutionary dynamics of all skills in general and soft skills in particular. Indeed, the nature of the soft skills makes them hardly replaceable by machine work, because they are a sort of tacit knowledge that human beings apply spontaneously and even unconsciously, built generation after generation, and not codifiable in “closed-form solutions”. Instead, cognitive tasks, though complex, are often much easier to codify.

Among subsets of soft skills, creativity is assuming increasing importance in shaping the future of employment. Easton and Djumalieva (2018) find that there is a strong demand for creativity in the online advertising job market and, following the work by Windsor and Bakhshi (2015), they are able to claim that creativity itself is consistently the most significant predictor for the chance of an occupation of growing (expressed as a percentage of the workforce), by the year 2030. A global study by Adobe, run in 2016, found that businesses which invest in creativity report an increase in employee productivity (78%) and customer satisfaction (80%). Silvestri et al. (2018) found that creative inventions (i.e., unconventional patents) are associated with a higher technological impact and market value during recessions.

As the European Union recently stated in the Entrepreneurship Competence Framework, creativity refers to a set of soft skills characterized by a strong strategic component, and goes together with other skills such as learning, innovation, project management, proactivity, and adaptability. Hence, focusing on such skills allows one to reflect upon the possibilities that workers have to be active parts inside firms. Creativity is, in fact, a pillar of the varied concept of entrepreneurship, and represents one of the fifteen competences that articulate the entrepreneurship framework (Venkuté et al. 2020). According to the empirical findings obtained by the European Working Condition Surveys (Eurofound 2005, 2015), every second workplace involves some form of “creative work”, which is not significantly threatened by automation, whereas every fourth worker carries out “routine” tasks that are likely to be automated in the near future. Moreover, the EU National Reports on the labor market soft skills (Pérez et al. 2020) have recently addressed some relevant issues for the Italian labor market that require further investigation and policy maker attention: in Italy, job quality is often low and there is a high skill mismatch, implying that the future occupational trends will see an increase of low and medium level jobs. This evident skill emergency and the consequent job instability makes it crucial to understand which are the strongest needs in the soft skill endowments of Italian workers. It is worth remarking that the Italian labor market suffered a huge crisis after the Great Recession in 2008–2011, followed by a sovereign debt crisis that worsened an already unstable situation. Moreover, starting from the early 2000s, Italy’s labor market has undergone substantial

reforms aimed at improving flexibility of job contracts. Still, one observes high unemployment, low productivity, high skill mismatch and a trend that favors low-skilled over high-skilled jobs (Hoffman et al. 2021). All these characteristics, combined with the richness of data available, makes Italy a perfect candidate for our analysis, which is described in the following.

In this paper, we exploit similarities in the Italian occupational structure and implement matrix completion—a recently developed machine learning technique which is often used by recommender systems (Hastie et al. 2015)—in order to predict the average importance levels of creative soft skills (i.e., soft skills associated with creativity) employed in each occupation and to identify peculiarities of such occupations and such skills. For the purpose of our analysis, we describe creativity along four dimensions and we identify in our data 25 skills associated with creativity. We then formulate a matrix completion optimization problem for the prediction of the average importance levels of creative soft skills (or, in short, for creativity levels prediction), and we solve it through a state-of-the-art algorithm (the Soft Impute algorithm). In more detail, the objective of our analysis is twofold: on the one hand, we investigate if matrix completion is able to predict with good accuracy elements belonging to selected columns (associated with skills) of a suitable occupation matrix¹ (which is derived from a recent survey), when such elements are artificially obscured (i.e., when only a portion of that matrix is made available as input to matrix completion); on the other hand, after providing a positive answer to the first question, we employ matrix completion to compare expected creativity levels (i.e. the elements of the original matrix that are predicted by matrix completion), with the associated actual creativity levels (i.e., with the corresponding elements of the original matrix, in the same positions that have been previously hidden artificially to matrix completion). This is done with the aim of identifying “outlier” professions whose “creativity needs” (as encoded by elements belonging to the selected columns of the matrix) are, respectively, much larger/much smaller than expected (according to their data-driven predictions, which are produced as outputs by matrix completion).

In other words, we apply matrix completion to identify a counterfactual value (or expected value, based on the information coming from the observed portion of the matrix) for the average importance level of each creative skill across professions. Then we get from the results of our analysis that such counterfactual value is typically in good agreement with the actual average importance level. However, for some professions, the observed level deviates positively/negatively from the counterfactual (i.e., the counterfactual provides an underestimate/overestimate), hence such professions are associated with a surplus/deficit in creativity levels. In this regard, the methodological approach used in this work resembles the one employed in the analysis of total factor productivity, in which a set of actual outputs is compared versus the corresponding set of predicted outputs according to given inputs and a specific production function.² In

¹ As detailed in Sect. 3, each entry of the occupation matrix to which matrix completion is applied in this work represents the average importance level (expressed as a percentage between 0% and 100%) of a skill/competence/working condition/working style for a specific worker type. As discussed in that section, a subset of the columns of that matrix is related to creativity.

² See Hidalgo (2021) for a comparison between total factor productivity, the analysis of economic complexity, and matrix factorization, which is related to matrix completion.

this case, the emergence of outliers may be explained, e.g., by the absence of relevant inputs in the model or by its lack of additional nonlinear terms.

To the best of the authors' knowledge, this is the first article in which the matrix completion technique mentioned above is applied to analyze workers' skills in a whole economy, with the aim of predicting suitable elements of an occupation matrix as functions of the other elements of that matrix. Our analysis shows that: (a) matrix completion demonstrates an excellent prediction capability on the specific dataset, making it meaningful to analyze cases for which the actual average importance levels of skills associated with creativity is significantly larger/smaller than the predicted one. Indeed, given the extremely good fit typically obtained, a large error in the prediction of the average importance levels of skills associated with creativity for a specific profession may be imputed to peculiarities of that profession, which make it less similar to other professions in the dataset and, as a consequence, more/less "hungry" of creative skills with respect to those professions³; (b) soft skills associated with creativity are not neatly separated among intellectual and technical workers; (c) some professions show a surplus of creativity with respect to its level predicted by their counterfactuals. In this case, their risk of being replaced by machines is deemed to be low; (d) on the other hand, we get a deficit of creativity for other professions. In this case, the risk for such professions of being replaced by artificial intelligence techniques is larger. In order to reduce this risk, training on creative soft skills might be tailored on these specific professions.

The paper is structured as follows. Section 2 reports a review of related literature. Section 3 describes the dataset used in our analysis. Section 4 summarizes the machine learning approach adopted in the work (more technical details about it are reported in the Appendix). Section 5 provides the main results of the analysis. Section 6 concludes the article with a discussion.

2 Related Literature

Our work builds on four main streams of existing literature dealing respectively with: (1) the relationship between ICT and soft skills; (2) skill relatedness, (3) creativity; (4) matrix completion. In the work, these issues are connected as follows. Our analysis allows the assessment of patterns and peculiarities of creative soft skills in the Italian labor market, as the endowments and formation of creativity among Italian workers could mitigate the negative effects of digitalization and job automation on low-skilled professions. More specifically, our analysis relies on the application of matrix completion to an occupation matrix whose entries represent, e.g., average importance levels of various soft skills (including ones associated with creativity) in different jobs. As detailed in the work, such an occupation matrix is quite well approximated by a low-

³ Another possible related explanation for the emergence of "outlier" professions is that such professions may come from "small" groups of similar professions that are under-represented in the training set employed by matrix completion, whereas other "larger" groups of similar professions may be more represented in that set.

rank matrix (an essential assumption for its successful application⁴), which could be justified by the presence of skill relatedness across industries.

(1) Relationship between ICT and soft skills

Numerous theoretical and empirical studies investigate the effect of technological progress on jobs, paying particular attention to the growth in importance of social skills in the labour market as an increased interest in social interactions skills has been noticed across all industries and occupations. The diffusion of ICT, and more broadly of “digital technologies”, has in fact led to significant transformations in several organizational aspects of consumption models, firms, industries, labour activities, and production processes. In the seminal paper by Autor et al. (2003), the authors claim that digitalization increases the possibility of making tasks automated, when these are characterized by a high degree of routineness. This is due to the fact that routine tasks are typically more easily codified, as compared to non-routine tasks.

In a very recent work, Bachmann et al. (2022) studied the impact of robot adoption on workers’ transitions (from employment to unemployment and viceversa) in Europe. They found small negative effects on job separations and small positive effects on job findings, with these effects being stronger for intensive routine manual or routine cognitive tasks occupations. In a similar vein, Domini et al. (2021b) found that automation spikes are associated with an increase in firms’ simultaneous net employment growth rate, and that these can be jointly explained by a larger hiring rate and a smaller separation rate, while Harrigan et al. (2016) highlighted how automation polarized French labor market in France. According to the authors of the latter work, such a polarization was mostly driven by changes in the composition of firms in each industry. Specifically, they found that “techies” (i.e., technology-related occupations) were an important force driving aggregate polarization in France at the time, as firms with more techies grew faster. Moreover, by focusing on French data, Domini et al. (2021a) found that the adoption of automation- or AI-related capital goods was followed by wage increase (1%) 3 years after these events, and that this increase was due to the hiring of new workers. It is worth mentioning that the concept of routineness of a task does not apply only to the cases of low skilled and unqualified (manual) functions, but also to those cognitive tasks that are carried out principally by managers and occupational workers. Therefore, one can distinguish between routine and non-routine tasks (e.g., in the case of manual tasks): while the latter exhibit a complementary relationship to digital technologies, cognitive routine tasks (which are typical of clerical workers and administrative jobs) are potentially highly exposed to automation, due to the introduction of digital technologies (Autor et al. 2003).⁵ To conclude, the literature on the task-based approach is rapidly expanding, and many papers have appeared recently, including recent contributions related to Italy as a case study (see, e.g., Bonacini et al. 2021; Carbonero and Scicchitano 2021; Caselli et al. 2021; Cetrulo et al. 2020; Esposito and Scicchitano 2022; Vannutelli et al. 2021).

⁴ See the Appendix for details.

⁵ Nevertheless, in this regard, it is worth remarking that, according to the literature, the disruptive effects of automation on labour substitution are quite controversial (Fernández-Macías et al. 2021).

(2) Skill relatedness

The term “skills” includes a wide range of qualitatively different capabilities that could encompass a huge variety of more specific competences, used in particular occupations and industries (Fitjar and Timmermans 2017). This implies that, to some extent, skills are industry-specific and the stronger the relatedness between industries, the smaller human capital destruction occurring when labor flows from one industry to another (Galetti et al. 2021).

Skills acquired in a specific industry, in fact, are often useful also in the context of other industries. Given the extremely relevant role assumed by human capital in a firm’s strategic asset stocks, a firm is expected to concentrate its efforts of diversification in areas that require skills which are already part of the background of its current workforce. In their pioneering work, Neffke and Henning (2013) proposed a quantitative way to define the similarity of industries’ human capital or skill requirements (i.e., the industries’ skill relatedness), by taking into account information associated with cross-industry labor flows. They showed that firms are significantly more likely to achieve diversification into industries that are linked to the firms’ core activities in terms of some skill-relatedness measure than into industries that do not present such ties.

(3) Creativity

The central importance of creativity as one of the main engines of scientific and social progress has been recently witnessed by a large number of related research efforts, all having the common goal of understanding its origin and nature. Among the various approaches followed by researchers to understand creativity, it is worth mentioning explanations of creativity from several points of view related, e.g., to mysticism, pragmatism, psychodynamics, psychometrics, cognition, socio-personality, and also multidisciplinary approaches (Sternberg and Lubart 1999). These different perspectives have provided valuable insights into creativity, but they have also been seen as having hindered some serious psychological research on this research topic. The diffused focus of research has also limited any possible conceptual agreement on what creativity exactly is. Indeed, in the literature, creativity has been variously described, e.g., as a process, product or personal trait. Some researchers have interpreted creativity as a staged process (e.g., Amabile 1983; Basadur et al. 1982), or as a cognitive process associated with the production of diverging ideas (De Bono 1992; West and Farr 1990).

The literature review conducted by Walia (2019) shows the existence of quite a general agreement among scholars about the fact that the concept of creativity is associated with the production of original and useful ideas and products (Mumford 2003). Runco and Jaeger (2012) observed that the standard definition of creativity (which requires the presence of some elements of originality and effectiveness) has quite a long history in the related literature. Hennessey and Amabile (2010) argued that the requirement to implement a creative idea can be considered as the origin of the innovation process. Most recent research on creativity focused on novelty and usefulness of ideas as its peculiar features (Mumford 2003). More in detail, novelty refers to originality, i.e., to the production of something which can be considered as new, whereas usefulness is related to the suitability of an idea for solving a specific problem under investigation (Amabile and Pratt 2016; Hennessey and Amabile 2010).

Some proposals of a definition of creativity include more stringent criteria, such as high quality level (Sternberg and Lubart 1995), absence of conformity (Niu and Sternberg 2002), surprise (Boden 2004), non-obviousness (Simonton 2012), and aestheticism accompanied by authenticity (Kharkhurin 2014).

Mumford et al. (2002) reported two different groups of processes, both related to creative work: (a) activities that lead to the generation of a novel idea; (b) activities that are required to implement such ideas (implementation). As an example, the identification of a still unsolved problem (conceptualization) certainly requires creativity and, according to Csikszentmihalyi (1988), could be even interpreted itself as a creative task. Taking simultaneously into consideration various heterogeneous elements that may jointly interact during the related processes of conceptualization, ideation, and implementation, creativity is sometimes modeled in the literature as a socio-psychological phenomenon (Amabile and Pillemer 2012), i.e., as a phenomenon in which several distinct characteristics of an individual interact, e.g., with her/his culture and environment (Lebuda and Csikszentmihalyi 2018). Even though the acts of creation may be considered to be individualistic at a first sight, their origin can be actually social (Glăveanu 2013). Finally, at the level of an individual, creativity could be interpreted as a construct of one's imagination (Lindqvist 2003); nevertheless, such a process is actually shaped by the creator's everyday interactions with her/his specific environment, taking into account, e.g., of the specific cultural, historical, or ideological context (Thibodeaux 2014).

A possible analytical definition of creativity comes from the seminal work of Edward De Bono and Efrim Zimbalist (De Bono and Zimbalist 1970). The authors defined lateral thinking as a set of different processes associated with a deliberate, systematic way of thinking in a creative way, i.e., related to an innovative way of thinking that is able to occur in a repeatable manner. According to their pioneering research in the field, lateral thinking is strictly related to creativity and can be described along four dimensions: (1) fluidity, as the ability of a subject to give the highest possible number of answers to a certain question; (2) flexibility, as the number of categories to which one can bring back these questions; (3) originality, as the ability of expressing new and innovative ideas; (4) processing, as the ability of realizing concretely one's ideas. Following this path, many researchers defined creativity as a set of strategic soft skills, which allow workers to find result-oriented solutions to open problems (Tucciarelli 2014; Ciappei and Cinque 2014; Pellerey 2016). Therefore, creativity goes hand in hand with other competences such as learning, innovation, project management, proactivity and adaptability to change.

Teamwork and creativity require a set of abilities that have been defined in the related literature as communication and mutual understanding of each other's preferences, motivations and comparative advantages, i.e., a kind of ability that has been conceptualized within the theory of mind (Premack and Woodruff 1978; Baron-Cohen 2000; Camerer et al. 2005). Starting from the work by Autor et al. (2003), tasks requiring creativity have been considered problematic to computerize. Indeed, recombining existing features to create new ideas is easily automatable, but recognizing whether these new combinations make sense and are valuable requires encoding an objective set of creative values (Frigotto and Riccaboni 2011). Moreover, Windsor and Bakhshi (2015) and Frey and Osborne (2017) suggested that the creative process is hard to

codify and, therefore, that creativity-intensive occupations have been shielded from automation. This remark extends to other soft skills different from the ones associated with creativity.

(4) *Matrix completion*

We conclude this literature review by addressing its last point, i.e., the application of machine learning to the specific context of the analysis of professions (this is only one of the many applications machine learning has found in the last years). Recent works devoted to this specific application are, e.g., Cockburn et al. (2018), Goldfarb et al. (2020), Iansiti and Lakhani (2020), and Lanzolla et al. (2021). In this framework, as already mentioned, in the present work creativity levels of different professions—measured through a recent survey, and collected in a suitable occupation matrix—are analyzed based on a recently developed machine learning technique, which is called matrix completion (Hastie et al. 2015). The choice of matrix completion for our analysis is due to the fact that this technique is often used by recommender systems (Ricci et al. 2011) to infer users' preferences, then suggest items to them, based on the inferred preferences. This is motivated by the ability of matrix completion to capture possibly hidden relationships among elements of a rating matrix (e.g., in the case of a movie-rating matrix, to capture the similarity between movies, or between users' preferences). A similar argument can be applied in the case of our occupation matrix. In this application, a low/high risk of automation for a profession may be associated with an underestimate/overestimate by a state-of-the-art machine learning algorithm of the creativity needs of that profession, based on a suitable training set. In the present context of the analysis of professions, it has to be mentioned that recommender systems are adopted, e.g., for job recommendation (Al-Otaibi and Ykhlef 2012). In the economic context, matrix completion can be exploited, e.g., to construct synthetic controls (Amjad et al. 2018), for their possible successive use in counterfactual analysis (Athey et al. 2021). Another possible application is in recommender systems for education. Variations of matrix completion have also been developed in the literature to discover relationships among different databases, assuming that each of them is modeled by a matrix (Bouchard et al. 2013). In the present work, matrix completion is applied for two purposes: in order to assess its prediction capability of the creativity needs of different professions, and with the aim of generating counterfactuals. It has to be remarked, indeed, that in the context of job market analysis, matrix completion has not been applied yet to define theoretical counterfactuals of expectations of average importance levels of skills. Such counterfactuals can be useful, e.g., to identify specific variations in the creativity needs of the various professions.

3 Data

ICP (*Indagine Campionaria sulle Professioni*, which can be translated into English as *Sample Survey on Professions*) is a survey on workers, promoted by the Italian National Institute of Statistics (ISTAT) and by the National Institute for Public Policies Analysis (INAPP). It was last run in 2013 by INAPP (INAPP 2013) on nearly 16.000 Italian workers in about 800 occupations, according to the 5-digit CP2011

classification (which is the Italian equivalent of the ISCO-08 ILO's classification).⁶ The survey reports more than 400 variables related to skills, attitudes and tasks, and is based on 20 workers on average per each Italian occupation. The sample is stratified so as to make it possible to achieve a high level of representativeness with respect, e.g., to firm size, geographical domain (at the level of macro-regions), occupation, and sector. The ICP survey investigates several characteristics of the occupations, based on a particularly rich and articulated questionnaire, which is structured in six main sections. These are the expressions of a content model able to provide simultaneously information from both a job-oriented and a worker-oriented perspective: (1) worker characteristics (enduring abilities and work style of workers); (2) worker requirements (skills and education); (3) occupational requirements (organisational and work context); (4) experience requirements (training, cross functional skills); (5) workforce characteristics (labour market information); and (6) occupation-specific information (generalised activities and work context).⁷ In doing this, several descriptors are considered by the survey, making it possible to discriminate, for instance, between inner individual abilities and competences acquired for the specific job.

Information in the survey is reported by entrepreneurs and Human Resources (HR) responsables. Respondents are asked to report the skill needs of their employed workforce to be satisfied in the short run, as future trends and new training needs are mainly related to competences and soft skills. For almost every question, two rating scales are provided: e.g., in the case of a skill, participants are asked to assess the importance level of that skill in their job and, provided their answer exceeds a suitable threshold, the intensity level in the use of such skill.

The ICP directly asks workers, instead of experts, to answer the questionnaire. In this way, the focus of the survey is given on the viewpoint of those who exercise the analyzed daily occupational activities, since they are able to provide a direct and concrete assessment of the importance and the intensity level of use of certain characteristics, which, depending on the answer, may turn out to be either essential or inessential to carry out one's job (Barbieri et al. 2022).

The ICP is a unique source of information on the work content. In fact it is the only survey replicating the US O*NET, the most comprehensive structure containing qualitative and quantitative information on job tasks, work context and workplaces at a very detailed level. As well as the US O*NET, also the Italian ICP focuses on occupations, so variables are built based on worker-level surveys and a post-survey validation made by focus groups. It is worth noticing that tasks and skills are specific to the Italian economy. So, the ICP can be used to define the structure of the labour market, the level of technology and the industrial relations, characterizing the Italian economy. Overall, the ICP dataset has a double advantage: it is accurate, granular, and rich, and most importantly, it is specific to the Italian productive system. No

⁶ The International Standard Classification of Occupations (ISCO) is an International Labour Organization (ILO) classification structure having the aim of organizing information related to labour and jobs. ISCO is defined by ILO itself as a tool for classifying jobs into a properly defined set of subgroups, according to the specific tasks and duties involved in each job. The updated classification was adopted at the end of 2007, and is known in the literature as ISCO-08.

⁷ The relationship between some of these characteristics and creativity is discussed later (see the end of this section).

international crosswalk (based for instance on US data) is needed. Thus, one possibly avoids biases arising when information referring, e.g., to the US occupational structure are linked to labour market data referring to different economies such as the European ones. More information on that can be found in Bonacini et al. (2021) and Vannutelli et al. (2021).

The ICP follows the Classification of Occupations (CP2011) produced by ISTAT in 2011 and, along this line, it divides jobs into eight major groups:

1. Legislators, managers and senior officials;
2. Intellectual, scientific and highly specialized professionals;
3. Technicians and associate professionals;
4. Clerical support workers;
5. Service and sales workers;
6. Artisans, specialized and agricultural workers;
7. Plant and machine operators and assemblers;
8. Elementary occupations.

Table 1 reports some descriptive statistics of the sectors in our available sample, related to the ICP survey performed in 2013. As the sample is representative of the Italian occupational structure, 4% of the observations belong to the primary sector (agriculture and farming), while the majority of workers (65%) is employed in one among services, other services, healthcare system, and cultural industry. More specifically, in the manufacturing sector we included those jobs that are specifically associated with goods transformation (i.e., related to metallurgical, siderurgical, chemical, pharmaceutical, mechanical, automotive, aerospace, railway, electronic, food, textile industries) while the term “other industries” refers to all the other industrial activities, such as agricultural industry, retail and distribution, energetic industry, transports and logistics, constructions. In the services sector, we included all those activities related to accommodation and catering, tourism, education, finance and banking, professional and scientific professions, real estate, public administration; while the term “other services” is related to services to the person and associations. Given the large availability of information included in the ICP, we decided to consider in our successive analysis not only skills, but also those variables that refer to working attitudes and styles as well as generalized work activities. We ended up identifying 25 items in the ICP survey as being associated with creativity, as highlighted in Table 2. Our choice of the set of 25 items was based on the opinion of an expert from INAPP (Riccaboni et al. 2021), taking into account the items available in the survey and the various definitions of creativity reported in the literature review of Sect. 2. Our resulting occupation matrix, coming from the ICP survey, is given by $m = 796$ rows which refer to professional units and $n = 255$ columns, of which 55 denote skills (25 directly associated with creativity, and the other 30 not directly associated with it), whereas the other 200 columns refer to competences, working conditions, and working styles. Each entry in position (i, j) represents the average importance level⁸ (expressed as a percentage between 0% and 100%) of skill/competence/working condition/working

⁸ As already mentioned, the ICP dataset provides actually also some additional information, as the survey explores, e.g., also the average intensity level in the use of each skill, conditional on a sufficiently high importance level expressed by the participant for that skill. In order to simplify our application of matrix

Table 1 Descriptive statistics of the sectors covered by the ICP dataset (INAPP 2013)

Variables	n. of professions	% of professions
Primary sector	34	4
Manufacturing sector	161	20
Other industries	97	12
Service sector	315	40
Other services	112	14
Healthcare system	40	5
Cultural industry	37	5

style j by worker type i . The 25 columns associated with the creative skills, on which our successive analysis is focused, form a subset J of columns of our occupation matrix.

In Table 3 we report the (empirical) mean and standard deviation (with respect to the professions) for the 25 creative soft skills we consider in our analysis. The values shown are percentages, meaning that workers answer on a scale from 0 to 100 to the question “How important is this soft skill in performing your job daily?”. We can clearly see that all these skills are relevant for the entire population, in particular, listening actively, adaptability and coordination with others are required the most.

4 Matrix Completion

In the following, we apply Matrix Completion (MC) to predict average importance levels of skills associated with creativity (or, shortly, creativity levels) for the various professions in our occupation matrix, exploiting the similarity patterns detected automatically by that machine learning technique from its application to the specific dataset.

MC can be defined as the task of filling in the missing entries of a partially observed matrix (Recht 2011). One well-known example of such a matrix is the rating matrix in a recommender system representing users’ tastes on products (e.g., movies). Given a subset of elements of a rating matrix, in which each entry in position (i, j) represents the rating of movie j by customer i , if customer i has watched movie j and is otherwise missing, one would like to predict the other entries of the matrix in order to make good recommendations to customers on which movie to watch next. In order to do that, one exploits similarities between users (rows) and between products (columns), as one expects that users assigning the same or similar ratings on different products will share similar interests on new products, resulting in a low-rank structure of the rating matrix. It is worth remarking that, in this application to recommender systems, MC is used with the final goal of predicting the values assumed by a subset of entries of

completion, only the average importance levels have been taken into account, ending up in a matrix of smaller size, and whose entries are also simpler to interpret.

Table 2 Soft skills associated with creativity in the ICP dataset

Indicator	Skill	ICP item code	Acronym
<i>Fluidity—Flexibility</i>	Solving complex problems	C17A	COMPL
	Solving unexpected problems	C27A	UNEX
	Listening actively	C2A	LIST
	Adaptability	F10	AD
	Service orientation	C16A	SERV
	Classification	D11A	CLASS
	Comprehension	D16A	COMPR
<i>Originality—Processing</i>	Critical sense	C7A	CRIT
	Analytical skills	C18A	ANSK
	Decision making	C31A	DMAK
	Originality	D6A	OR
	Production of ideas	D5A	IDEAS
	Learning strategies	C9A	LEAST
	Active learning	C8A	ALEARN
<i>Learning—Innovation</i>	Teaching	C15A	TEACH
	Creative thinking	G11A	CREAT
	Innovation	F15	INN
	Persuading	C13A	PERS
	Understanding others	C11A	UNDOTH
	Negotiating	C14A	NEG
<i>Planning—Proactivity</i>	Time management	C32A	TIME
	Financial resources management	C33A	FRM
	Material resources management	C34A	MRM
	Human resources management	C35A	HRM
	Coordination with others	C12A	COORD

the rating matrix for which no ground truth is available at the time of prediction.⁹ Still, in order to perform a validation of an MC algorithm (for instance, with the aim of selecting an optimal value for one hyperparameter of its associated optimization problem), the MC predictions are typically compared with the ground truth coming from another subset of entries that can be actually observed, but that are artificially obscured to the MC algorithm. In this way, one has the possibility of assessing in a fair way the prediction capability of the MC algorithm. A similar use is made in our present application (for which the occupation matrix is actually known as a whole). Moreover, one reason for which we expected MC to be successful for our application is that, in a similar way as in its use for recommender systems, our occupation matrix turned out to be quite well approximated by a low-rank matrix (see the next paragraphs for details). A possible explanation for this is that workers in different industries acquire specific

⁹ Of course, the customer may still evaluate a-posteriori the quality of her/his personalized recommendation, which is produced as a by-product of the MC application.

Table 3 Mean and standard deviation for the 25 creative soft skills

Skill	Mean	Std dev
Listening actively	71.74	13.23
Adaptability	69.53	10.32
Coordination with others	64.57	13.44
Time management	63.94	12.79
Critical sense	59.50	19.24
Active learning	56.63	17.84
Innovation	56.10	18.40
Solving complex problems	55.82	18.80
Service orientation	54.84	15.54
Teaching	51.09	18.11
Understanding others	50.48	18.22
Comprehension	50.43	13.73
Decision making	47.61	18.84
Classification	46.69	15.22
Negotiating	46.65	18.07
Originality	46.60	20.99
Analytical skills	46.42	15.57
Material resources management	44.88	15.01
Creative thinking	44.17	22.68
Learning strategies	43.96	18.86
Production of ideas	41.77	16.57
Human resources management	39.89	21.25
Financial resources management	39.32	24.25
Solving unexpected problems	39.09	18.15
Persuading	38.46	16.73

skills, and when they move to other professions, they apply such skills in the new context. As a consequence, jobs in different industries (i.e., rows in our occupation matrix) may share similar importance and intensity levels of soft skills.

In order to apply MC to our occupation matrix (or, for a successive analysis, to its suitable submatrix, see Sect. 5.1), we generate artificially partially observed matrices from it, by artificially obscuring randomly 10%, 25% and 50% of the entries in the 25 columns associated with the creative skills, focusing each time on the prediction capability of matrix completion on each single row (occupation). The procedure is repeated several times (details are in the Appendix). This sampling procedure is justified by the fact that we want to apply MC to predict entries of the occupation matrix that are related to creativity (i.e., that belong to one of the 25 selected columns). The three choices 10%, 25% and 50% are made in order to assess the accuracy (and robustness) of MC for quite different percentages of obscured entries. Finally, the motivation itself behind artificially obscuring known elements of the occupation matrix is that, in this way, it is possible to validate/test the predictions obtained by MC (i.e., a ground truth is available for comparison purposes).

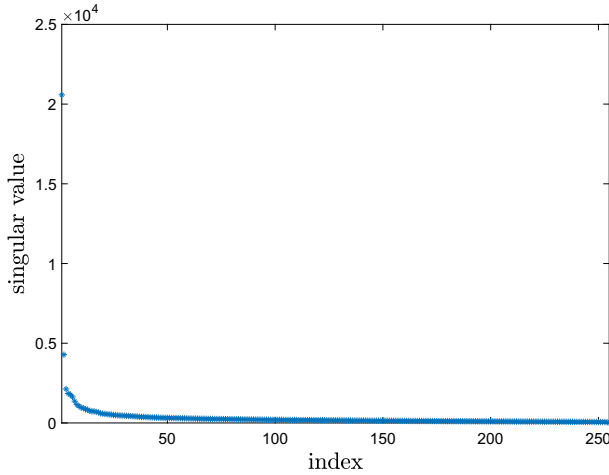


Fig. 1 Distribution of the singular values of the occupation matrix considered in the analysis

In summary, we consider the following nuclear-norm regularized MC optimization problem:

$$\underset{\mathbf{Z} \in \mathbb{R}^{m \times n}}{\text{minimize}} \left(\frac{1}{2} \sum_{(i,j) \in \Omega^{\text{tr}}} (M_{i,j} - Z_{i,j})^2 + \lambda \|\mathbf{Z}\|_* \right), \tag{1}$$

where Ω^{tr} is a training set of positions (i, j) corresponding to the known entries of the partially observed matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$, $\mathbf{Z} \in \mathbb{R}^{m \times n}$ is the completed matrix, $\|\mathbf{Z}\|_*$ is its nuclear norm, and $\lambda \geq 0$ is a regularization constant. Then, we solve it by applying the Soft Impute algorithm (Mazumder et al. 2009). This is proved therein to converge to an optimal solution of the optimization problem (1). The optimization problem itself is solved by the Soft Impute algorithm several times, for different choices of the set of obscured entries. For each such repetition, the best value of λ is found by minimizing a suitable error on a validation subset of missing entries, whereas the final performance is evaluated on the remaining test set of other missing entries. Technical details about the choice of the various training/validation/test sets and of the regularization parameter of the optimization problem (1) are reported in the Appendix. Additional details about the optimization problem (1) and the Soft Impute algorithm are reported also in Metulini et al. (2022).

As a preliminary check for the applicability of MC, we compute the singular value decomposition of our occupation matrix. Figure 1 shows that its singular values decay quite fast to 0, hence that the matrix can be well approximated by a low-rank one. As reported in the Appendix, this is a necessary condition for an effective application of MC, which is satisfied by the dataset under analysis.

In order to better illustrate the approach used in our analysis, we consider a specific instance of the optimization problem (1), then we visualize its results. Figure 2 shows the original matrix, the locations of the observed and the missing entries in the original

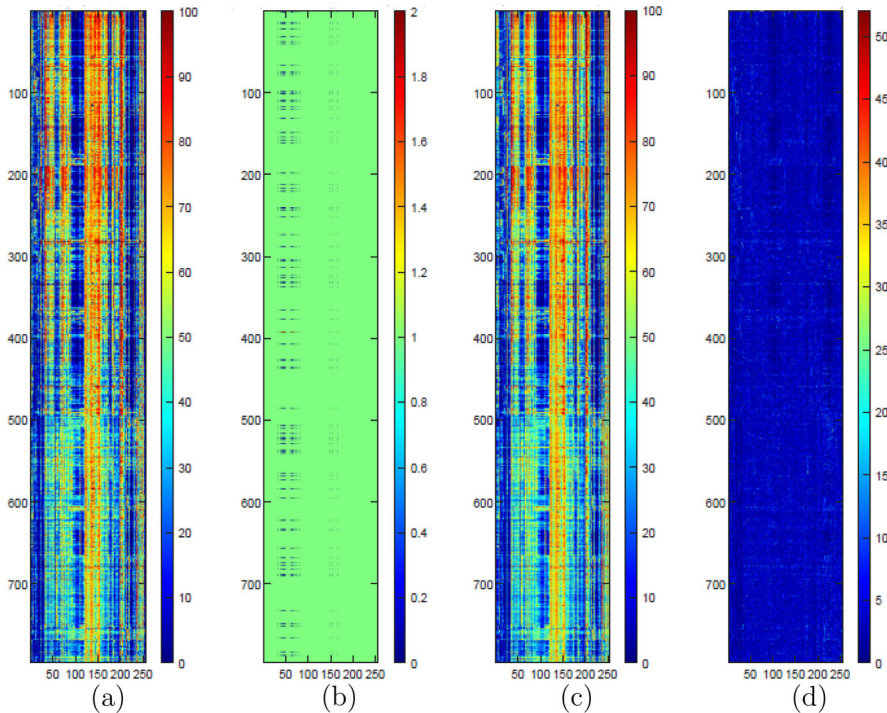


Fig. 2 Visual representation of **a** the original matrix, **b** the locations of its observed (green) and missing entries (blue: validation; red: test), **c** the reconstructed matrix for a specific repetition, and **d** the absolute value of the prediction error for the same repetition (color figure online)

matrix (where the green cells denote the elements in the training set, the blue cells refer to the validation set and the red ones are related to the test set), the reconstructed matrix, and the error in absolute value for each cell. As the second subfigure shows, all the obscured entries are at the intersection between specific rows (one of which is associated with the test set) and the set of columns associated with creativity.

Then, Fig. 3 presents a visual representation of the (empirical) mean and standard deviation of the Root Mean Square Error (RMSE) of MC prediction on the test set (see the Appendix for details on its definition) per profession, where each row (profession) refers to the mean and standard deviation computed with respect to the repetitions having as test set elements belonging only to that specific row of the original matrix. As the figure shows, both the mean and standard deviation per profession are typically small (taking into account that the entries of the original matrix are numbers between 0 and 100).

Figure 4 reports more detailed results for all the repetitions for which the elements of the test set refer to a specific profession. In particular, the figure shows how the RMSE of MC prediction (evaluated, respectively, on the training set, validation set, and test set) changes as a function of the regularization parameter λ . The sizes of the three sets are also reported in the figure. As the figure reveals, the optimal value of λ (which is computed based on the RMSE of MC prediction on the validation set) is associated with a quite small RMSE of MC prediction for both the validation and

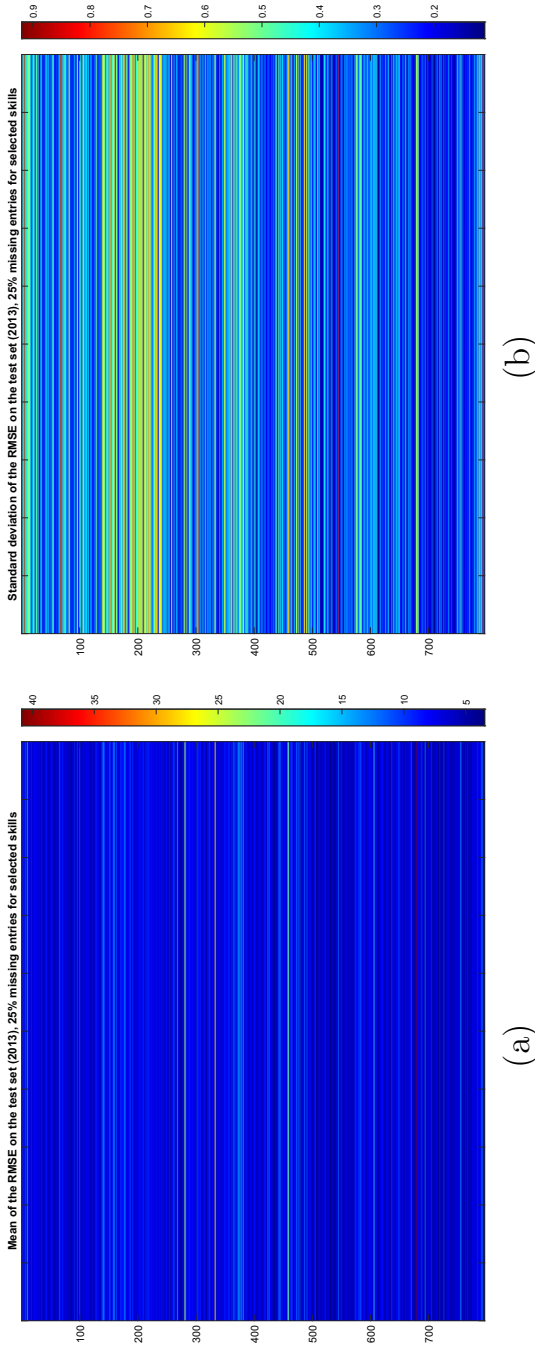
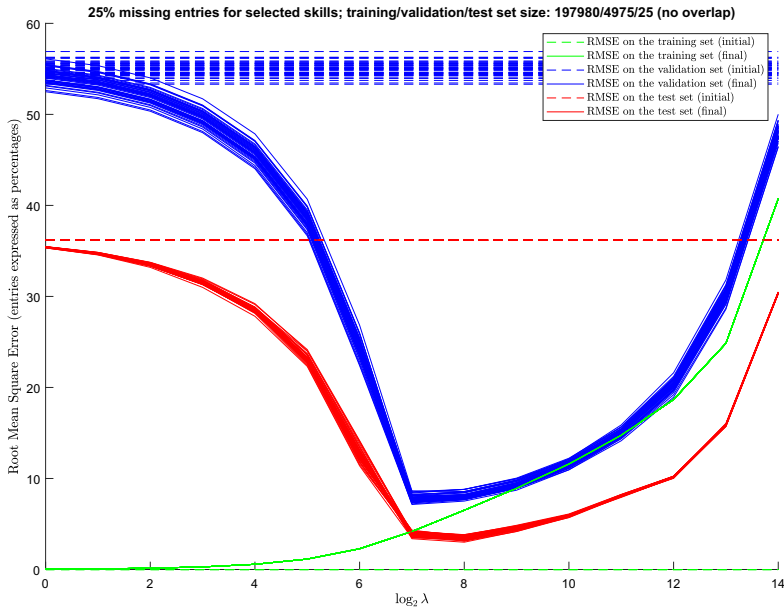


Fig. 3 **a** Mean and **b** standard deviation of the RMSE of prediction on the test set per profession, for a specific choice of the percentage of missing entries in the columns associated with creativity (color figure online)



Year: 2013; profession: 7.3.1.2.0; number of simulations: 44; mean of the RMSE on the test set: 3.8766; standard deviation of the RMSE on the test set: 0.26004

Fig. 4 RMSE of prediction on the training, validation and test sets as functions of the regularization parameter λ , for all the repetitions associated with the same test set related to a specific profession (color figure online)

test sets. Moreover, for each of the two cases, the variability of the RMSE curve with respect to changing the training set (i.e., performing another repetition of the analysis, for the same choice of the test set) is very small, whereas the variability of the RMSE curve of MC prediction for the training set itself is negligible. It is worth mentioning that, for the specific case reported in Fig. 4, the performance of the method on the test set is slightly better than the one on the validation set. This is likely due to the fact that the elements of the two sets are sampled from different portions of the occupation matrix. Moreover, the test set has a much smaller cardinality (since its elements come from a specific row of the occupation matrix). In any case, the behaviour of the RMSE curve as a function of the regularization parameter λ is similar for both sets, being also the minima of these two curves achieved for almost the same values of λ . Although Figs. 3 and 4 refer only to the percentage 25% of missing entries in the selected columns associated with creativity, similar results have been obtained for the other two percentages.

Finally, Fig. 5 reports, for each of the three percentages of missing values in the 25 selected columns, the boxplots of the Mean Relative Error per profession, denoted as MRE_i and defined later in Eq. (2), for 4 subsets of professions i . Such subsets correspond to the 4 quartiles of the average true value $\frac{1}{25} \sum_{j \in J} true_{i,j}$, where J is the set of the 25 skills directly associated with creativity, in such a way that the average is limited to such skills. The figure clearly shows that the surplus tends to be larger for the professions with a larger average true value of creative skill levels. This holds for each of the three percentages of missing values.

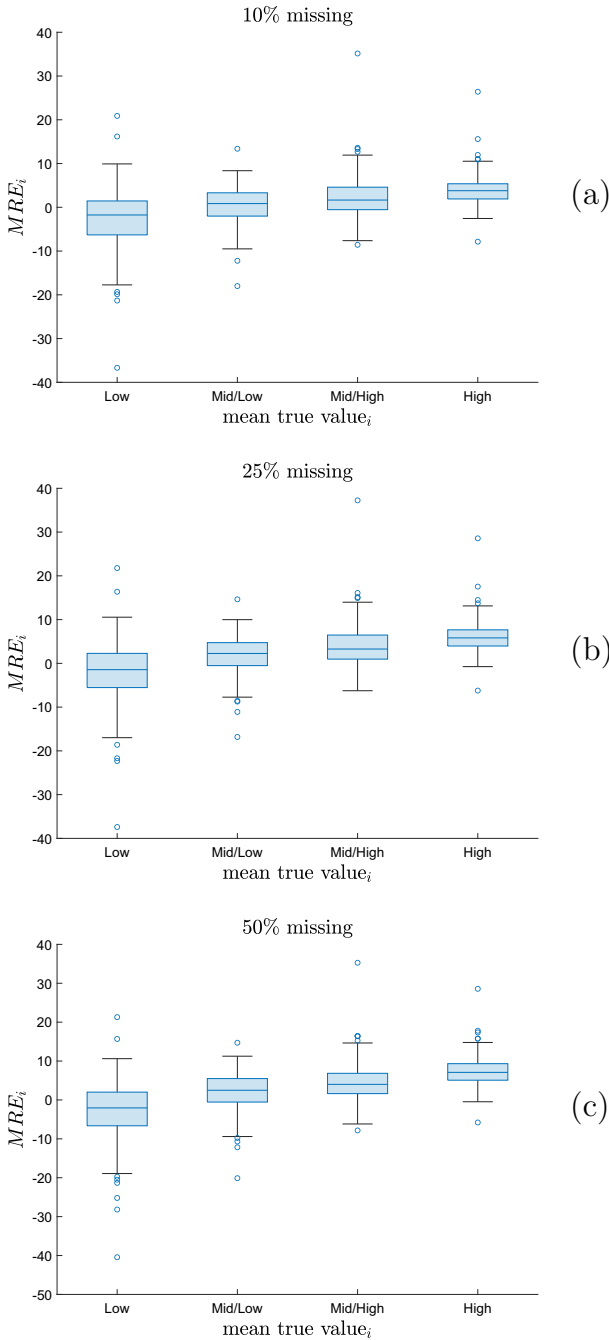


Fig. 5 Boxplots of MRE_i for 4 subsets of professions i corresponding to the 4 quartiles of the average true value $\frac{1}{25} \sum_{j \in J} true_{i,j}$ of the skill importance levels, for the skills directly associated with creativity: **a** 10% missing values in the selected columns; **b** 25% missing values in the selected columns; **c** 50% missing values in the selected columns

5 Results

For each profession i , we predict (via the MC estimate $predicted_{i,j,r}$) the missing entry $true_{i,j}$ of each of the 25 skills $j \in J$ associated with creativity, in each repetition r belonging to the set of r_i repetitions in which such elements of row i are in the test set. Then, we compute the associated Mean Relative Error, defined as follows (and expressed as a percentage):

$$MRE_i = \frac{1}{r_i} \sum_{r=1}^{r_i} \left[\frac{\sum_{j \in J} (true_{i,j} - predicted_{i,j,r})}{\sum_{j \in J} true_{i,j}} \right] \times 100 \quad (2)$$

To avoid burdening the notation, the dependence on the percentage of missing elements in the 25 selected columns is not reported in Eq. (2). We say that for a specific profession i there is a surplus (deficit) of creativity when $MRE_i > 0$ (when $MRE_i < 0$).

What we find (see Fig. 6 for the histograms of the MRE_i in the three cases) is that around 65% of occupations deviate positively (i.e., they have positive MRE_i). Moreover, deviations are concentrated in the interval $[-1\%, +5\%]$. For instance, in the case of 10% missing entries associated with creativity, about 52% of the observations of the MRE_i fall in this interval, meaning that the prediction error is typically very small (see also Table 4 for summary statistics, and Table 5 for results at the level of each major group).

Referring to the list reported in Sect. 3, the professional groups experiencing most of the positive deviations are: the second (intellectual, scientific, highly specialized professions), the third (technical professions), the sixth (artisans, specialized workers, agriculturalists) and the seventh (plant operators, machinery workers, drivers of vehicles), while the main skills explaining these results (i.e., which contribute to the summation in Eq. (2) with the largest positive terms) are: listening, human resources management and creative thinking.

The remaining professions whose predictions deviate negatively are mostly in the first (legislators, entrepreneurs, senior management), third (technical professions), sixth (artisans, specialized workers, agriculturalists), seventh (plant operators, machinery workers, drivers of vehicles) group, while the associated skills that motivate the negative differences are analytical skills and solving unexpected problems.

In summary, there is a need to reinforce competences such as solving unexpected problems, listening actively, critical sense, adaptability and service orientation, which concerns transversely all the professional groups, including technicians and associate professionals as well as service and sales workers, artisans, specialized workers and agriculturalists and, to a smaller extent, plant and machine operators and assemblers.

The most intense skill upgrading needs are observed in the high-tech manufacturing sector (chemical and pharmaceutical, electronic, energetic, engineering industry) but they are quite relevant also in the service sector when related to education, health, communication, financial activities and other services.

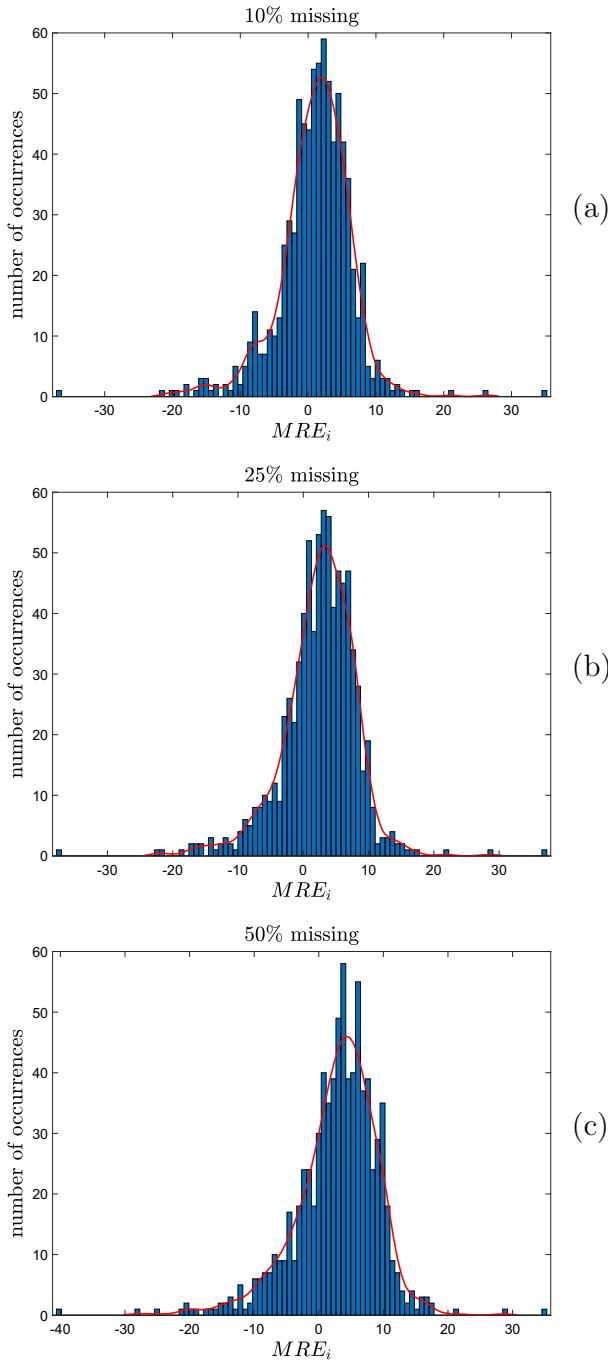


Fig. 6 Histograms (and kernel-fitted histograms) of MRE_i in the full dataset of professions, for different percentages of missing entries in the columns associated with creativity: **a** 10% missing entries in the selected columns; **b** 25% missing entries in the selected columns; **c** 50% missing entries in the selected columns

Table 4 Summary statistics for the distribution of MRE_i in the full dataset of professions, for different percentages of missing entries in the columns associated with creativity

	10% missing		25% missing		50% missing	
	# professions	%	# professions	%	# professions	%
$MRE_i > 5\%$	151	18.97	264	33.17	313	39.32
$MRE_i \in [-1\%; +5\%]$	419	52.64	367	46.10	308	38.69
$MRE_i < -1\%$	226	28.39	165	20.73	175	21.98
<i>Total</i>	796	100	796	100	796	100

To address the issue above in more detail, in Table 6 we report 4 professions for each major group whose MRE_i is among the largest in absolute value, and the relative skills in which they deviate the most, both positively and negatively.

5.1 Results Related to Specific Economic Sectors

In this subsection, we focus our analysis on two economic sectors, the health system and the cultural industry, by applying MC to the respective occupation submatrices associated with professions belonging to each of the two sectors. These two sectors are chosen for the following reasons. In the case of the health system, we are interested to identify professions at a high risk of being automated, due to actual creativity levels smaller the ones predicted by the counterfactuals generated by matrix completion. On the one hand, the healthcare sector is a working context in which creativity is often assumed to play a limited role (at least in the cases for which routine tasks are common). On the other hand, the analysis of the cultural industry is performed to get a further validation of the method, since the creativity levels for the professions in this sector are typically expected to be higher than the ones of the professions belonging to other sectors. So, in this case, we check that the predictions of MC are consistent with this expectation.¹⁰

Figure 7 shows that the creative behaviour of professions in the healthcare sector is highly heterogeneous: categories of health workers dealing with patients on a daily basis are used to interact more with people, thus the need for soft skills in general and

¹⁰ Another reason for which we focused our attention on the occupation submatrices related to the health system and to the cultural industry is that they contain a much smaller number of rows (40 and 37, respectively) than, e.g., the one related to manufacturing (161). Indeed, in our application the MC optimization problem had to be solved a quite large amount of times (it was needed to solve such problem not only for different choices of the regularization parameter, but also for various choices of the partition of the occupation matrix in training/validation/test sets). This restricted the number of different additional analyses that could be practically performed. Another motivation for our choices is that it is hard to associate some of the occupations available in our dataset (which are grouped according to the 5-digit CP2011 classification) to a specific sector. Indeed, many professions are actually practised in different sectors. For instance, some of those professions that we included in Table 1 under the “other industries” label (e.g., agricultural industry, mining industry, energetic industry and service industry, such as finance, banking and insurance) might as well be practised inside the manufacturing industry. Moreover, the same profession could be practised either in a private or in a public sector. Hence, an integration to the current dataset would be needed to extend the current analysis, e.g., to other economic sectors such as manufacturing, or to compare private/public sectors.

Table 5 Distribution of MRE_i in the interval $[-1\%; +5\%]$ by major group, for different percentages of missing entries in the columns associated with creativity. The total number of professions for each major group is also reported in the table

Major groups Professions with MRE_i in $[-1\%; +5\%]$	10% missing % professions	25% missing % professions	50% missing % professions	Observations total # professions
1. Legislators, managers and senior officials	64.18	49.25	37.31	67
2. Intellectual, scientific and highly specialized professionals	68	48.57	37.14	175
3. Technicians and associate professionals	53.14	48.75	45	160
4. Clerical support workers	30	33.34	36.67	30
5. Service and sales workers	39.68	38.09	30.16	63
6. Artisans, specialized and agricultural workers	44.70	45.88	39.41	170
7. Plant and machine operators and assemblers	47.57	48.54	41.74	103
8. Elementary occupations	32.14	32.14	21.43	28

Table 6 Professions, means of the observed values of creative soft skills, MRE_i (in the case of 10% missing observations in the selected columns) and skills which contribute most to the positive/negative deviation

Professions	Mean true values	MRE_i 10%	Largest pos. dev. skills	Largest neg. dev. skills
Police chiefs	82.52	7.15	ANSK, INN, CREAT	LEAST, UNDOOTH, UNEX
PA managers	79.20	8.21	AD, DES	NEG, SERV, UNEX
Firm managers (manufacture)	55.05	- 3.16	SERV, MRM, DES	CRIT, NEG, DMAK
Firm managers (building)	59	- 1.82	SERV, ANSK, CREAT	LEAST, FRM, HRM
Cartographers	46.05	- 6.16	UNDOOTH, PERS, SERV	ANSK, CLASS, INN
Dietologists	57.80	- 6.17	CRIT, UNDOOTH, NEG	PERS, TEACH, DES
Lawyers	62.05	11.04	ANSK, SERV, CREAT	PERS, NEG, DMAK
Dialogists	51.87	11.29	PERS, FRM, HRM	AD, INN, CREAT
IT technicians	63	26.40	CRIT, TIME, HRM	LIST, LEAST, UNDOOTH
Interviewers	16.50	- 36.67	CRIT, AD, SERV	LIST, TIME, AD
Athletes	29	- 15.22	MRM, DES, CREAT	ALEARN, NEG, DMAK
Presentators	53.84	13.60	CRIT, SERV, COMPL	DES, CREAT
Typists	28.20	- 21.31	UNDOOTH, PERS, COMPL	MRM
Data entry officers	35.90	15.55	ALEARN, UNDOOTH, DES	TIME
Purchasing managers	49.82	6.18	CRIT, CREAT	DMAK, FRM, DES
Post office workers	34.82	6.44	UNDOOTH, AD, ANSK	TIME, DES

Table 6 continued

Professions	Mean true values	MRE_i 10%	Largest pos. dev. skills	Largest neg. dev. skills
Shop assistants	19.90	- 19.87	CRIT, ALEARN, UNDOOTH	LIST, LEAST, TIME
Remote sellers	35.60	20.89	TEACH, UNEX, FRM	PERS, NEG, DES
Waiters	32.31	- 14.97	UNDOOTH, NEG, FRM	UNEX, HRM, AD
Hairdressers	52.20	13.31	AD, NEG, TEACH	ALEARN, UNDOOTH, PERS
Tilers	27.75	13.38	UNDOOTH, NEG, DES	LIST, CRIT, AD
Tunnel owners	46.62	- 19.33	TEACH, COMPL	UNDOOTH, NEG, ANSK
Planes mechanics	45.87	- 18.01	UNDOOTH, NEG, SERV	AD, INN
Engravers	54.35	11.92	ALEARN, PERS, TIME	CRIT, NEG, FRM
Electrochemical plant workers	52	35.14	LEAST, UNDOOTH, NEG	LIST, CRIT, AD
Machinery operators	34.36	16.18	UNEX, DMAK, CREAT	ALEARN, LEAST, SERV
Bus conductors	24	- 14.64	LEAST, AD, DMAK	UNDOOTH, FRM, DES
Boiler conductors	29	- 15.98	ALEARN, LEAST, AD	DES, INI
Cleaning staff	25.10	- 15.08	CRIT, ALEARN, AD	TIME, MRM
Forestry personnel	52.33	11.75	LIST, CRIT, UNEX	NEG, SERV, DES
Fishing/hunting personnel	65	15.58	LIST	ALEARN, UNDOOTH, NEG
Industrial activity personnel	30.35	- 13.78	PERS, NEG, SERV	LIST, AD, CLASS

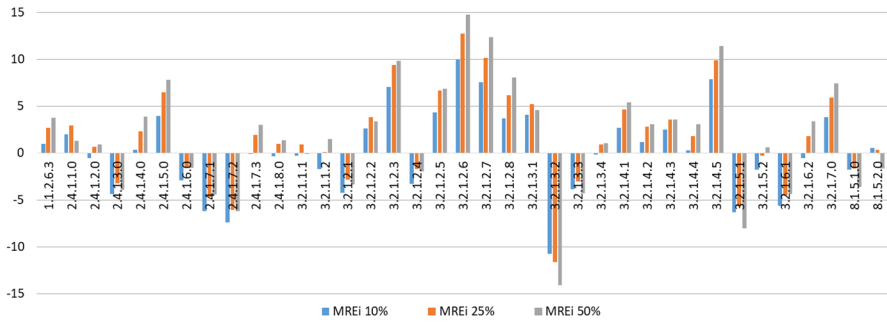


Fig. 7 MRE_i in the Health Sector-ADA 19 (ADA stands for *Area di Attività*, or *Area of Activity*). The correspondence between codes and professions is reported in Table 7 in the Appendix (color figure online)

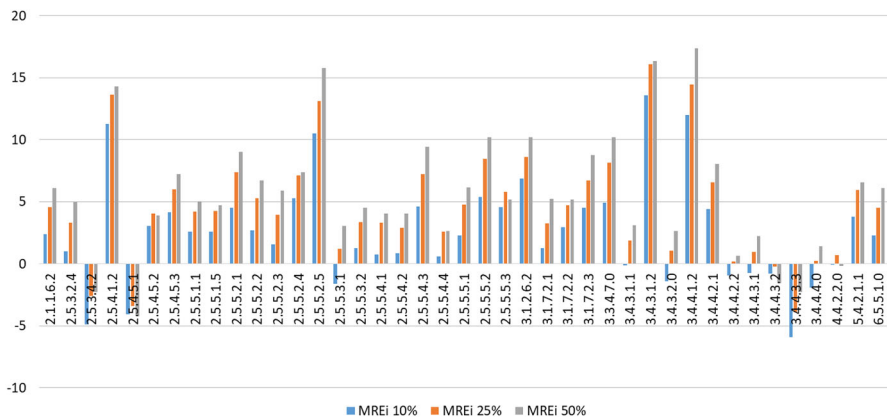


Fig. 8 MRE_i in the Cultural Industry-ADA 22 (ADA stands for *Area di Attività*, or *Area of Activity*). The correspondence between codes and professions is reported in Table 8 in the Appendix (color figure online)

creativity in particular is higher. This also implies that MC negative deviations (i.e., underestimates of the true creativity level) are likely to occur.

That is exactly the case for jobs like Therapists in neuropsychomotricity (3.2.1.2.5), Therapists in psychiatric rehabilitation (3.2.1.2.6), Social educators (3.2.1.2.7) and Occupational therapists (3.2.1.2.8). On the other end of the distribution we find jobs that are mainly dealing with technical diagnostic, such as Diagnostic imaging and radiotherapy specialists (2.4.1.6.0) and Biomedical laboratory assistants (3.2.1.3.2).

In the case of the cultural sector, instead, the distribution is smoother and creativity patterns are more homogeneous among professions (see Fig. 8). In particular, the MC predictions for almost all kinds of workers in this sector deviate positively and the magnitude of these deviations is quite large.

The most creative occupations, according to the results of our analysis above, are Dialogists (2.5.4.1.2), Set designers (2.5.5.2.5), Artistic performances presenters (3.4.3.1.2) and Stage builders (3.4.4.1.2), while Painters and sculptors (2.5.5.1.1), Directors (2.5.5.2.1), Actors (2.5.5.2.2), Artistic directors (2.5.5.2.3), Scriptwriters (2.5.5.2.4) take values in the middle of the distribution.

5.2 The Evolution of Creativity Needs

Recent developments in the MC literature extend the range of applications of this method to causal inference with panel data (Athey et al. 2021). They show that MC outperforms other synthetic control methods, as its simulations are based on real data. Thanks to the availability of the first edition of the *Survey on Occupations*, which was run in 2007, in this subsection we extend our analysis to this year, in order to assess whether our model has some predictive power over time, when the results of the MC analysis of the 2007 dataset are used to predict creativity levels observed later in 2013.

The first edition of the ICP adopted a slightly different occupation classification (*Classificazione delle Professioni 2001* or *Classification of Professions 2001*, CP2001), as this was subject to changes operated by ISTAT in 2011. Even though occupations are clustered into the same major groups in both editions of the survey, the two different classification criteria adopted do not allow to get a perfect one-to-one matching of professions at the 5-th digit occupational level. As a consequence, we reduce the rows of our two datasets (related to the years 2007 and 2013, respectively) to those professions for which there was either no change in the classification code, or a negligible change in the job description occurred. In this way, the two occupation matrices are reduced to 510 closely matching professions. Then, we apply to each of the reduced 2007 and 2013 occupation matrices the same kind of MC analysis described in the previous parts of Sect. 5 and in the Appendix. As before, the analysis is performed for each of the three percentages of missing observations in the columns associated with creative skills (10%, 25%, 50%). In particular, for each pair (i, j) , where i denotes a profession in the reduced subset of 510 professions, and j one of the selected creative skills, we compare the mean MC prediction of the average importance level of skill j for the profession i based on the 2013 and 2007 datasets, respectively, i.e., we compute¹¹

$$\Delta \overline{predicted}_{i,j} = \frac{1}{r_i} \sum_{r=1}^{r_i} \left(predicted_{i,j,r}^{(2013)} - predicted_{i,j,r}^{(2007)} \right) \quad (3)$$

(again, to avoid burdening the notation, the dependence on the percentage of missing elements in the selected 25 columns is not reported in Eq. (3)).

Figures 9, 10, and 11 report, for the three cases, and for each creative soft skill j , the resulting histograms of the variations $\Delta \overline{predicted}_{i,j}$ (in each histogram, j is fixed, whereas i varies). These figures show that the creative soft skills for which the largest positive variations are obtained are persuading, creative thinking, financial resource management and classification, while the creative soft skills with the largest negative variations are listening actively, teaching, time management and solving unexpected problems.

¹¹ The same seed for pseudo-random number generation has been used for the simulations based on the reduced 2007 and 2013 datasets. So, for each profession i considered, the number r_i is the same for the two analyses.

As a further comparison, Fig. 12 reports the boxplots of the averages

$$\Delta \overline{predicted}_i = \frac{1}{25} \sum_{j \in J} \Delta \overline{predicted}_{i,j} \quad (4)$$

over the creative soft skills, for the three cases considered in the analysis. The figure also shows that the [first quartile, third quartile] intervals of $\Delta \overline{predicted}_i$ are, in the three cases, $[-7.64\%, 1.00\%]$, $[-7.51\%, 0.96\%]$, $[-7.13\%, 0.55\%]$, respectively. Hence, the predictions—based on the MC analysis of the reduced 2007 dataset—of the creativity levels for the professions considered in the two analyses are quite similar to the corresponding predictions based on the reduced 2013 dataset.

The professions i in the reduced datasets for which the average $\Delta \overline{predicted}_i$ assumes the largest negative values are: Shop assistants (5.1.2.2.0), Meter readers (8.1.1.2.0), Remote sellers (5.1.2.5.2), Biomedical laboratory technicians (3.2.1.3.2) and Plant operators in sugar refining (7.3.2.5.0). Similarly, the ones for which the average $\Delta \overline{predicted}_i$ assumes the largest positive values are: Plant operators in pottery production (7.1.3.3.1), Flame welders (6.2.1.2.0), Electric welders (6.2.1.7.0), Assemblers of composite industrial articles (7.2.7.9.0) and Hunters (6.4.5.4.0). In the period 2007–2013, the professions that experienced the largest increase in the predictions of the average importance levels of creative soft skills belong mainly to the 6-th (Artisans, specialized and agricultural workers) and 7-th (Plant and machine operators and assemblers) major groups. These are major groups characterized by a high level of routinary (manual and cognitive) tasks (Gualtieri et al. 2018). Thus, these results suggest that technological change, advances in ICTs, the advent of Industry 4.0 and the last developments in artificial intelligence, which might lead to a substitution of mechanical and repetitive jobs, are forcing these same jobs to adapt and reshape their soft skills provision.

6 Final Discussion

The recent economic literature is animating a lively debate on the effects of digitalization on employment, ending up claiming that routinary jobs are more exposed to being substituted by machines (Cetrulo et al. 2020). In this framework, the existence and the magnitude of the substitution effect depend on the tasks involved in each job: in particular, routine tasks that are easier to codify are more likely to be automated, while non-routine tasks that comprise a larger use of soft skills might be less exposed to this risk (Cirillo et al. 2021). Among subsets of soft skills, creativity is one of the hardest to be encoded, thus highly creative occupations tend to suffer less from the introduction of machines and digital systems. In this paper, we exploit similarities in the Italian occupational structure and implement a recently developed machine learning technique to predict the importance levels of creative skills employed in each occupation and to identify the creativity needs of occupations. We find that professions belonging to the major group of legislators, senior officials and managers, as well as intellectual professionals have a greater surplus of creative soft skills. Nevertheless,

10% missing

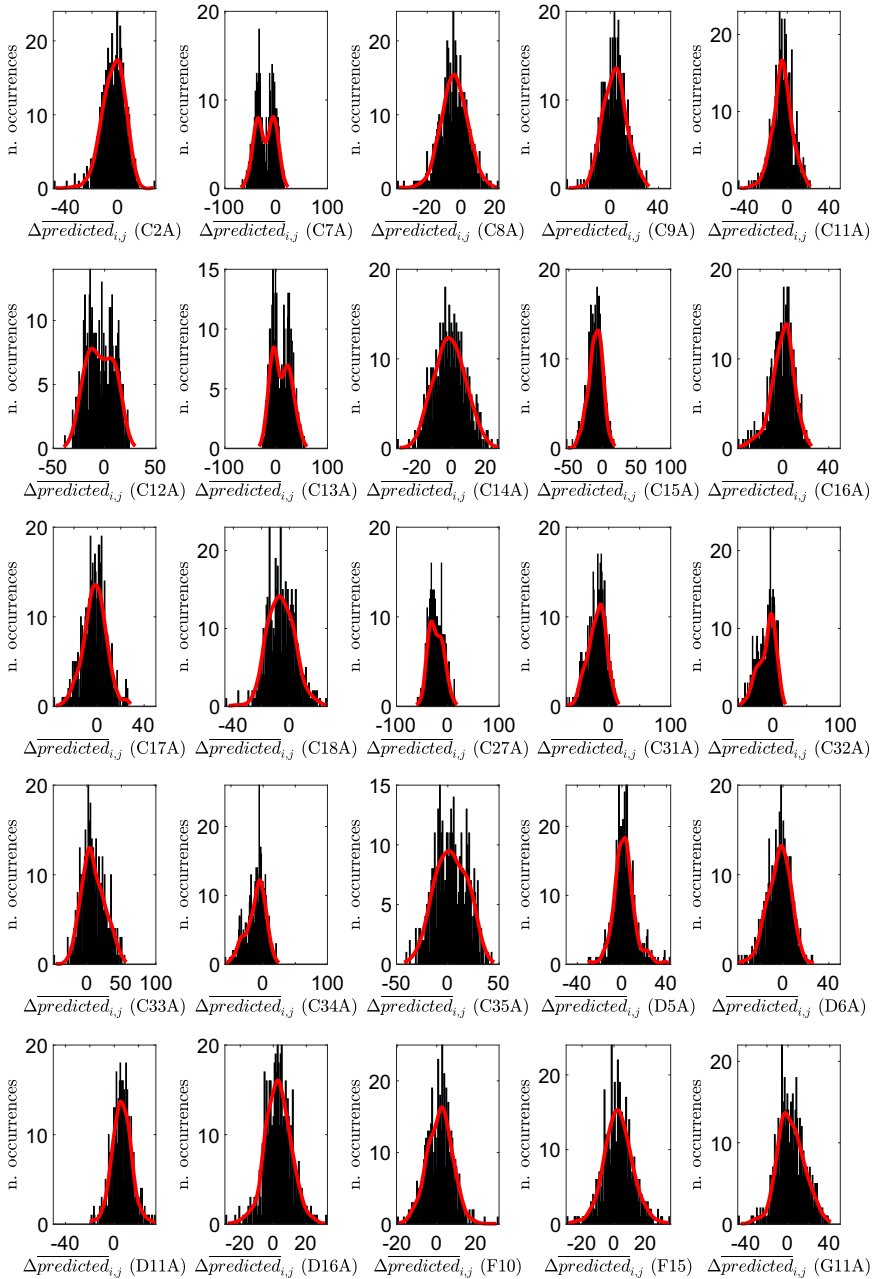


Fig. 9 Histograms of the variations $\Delta \overline{\text{predicted}}_{i,j}$ for the various creative soft skills j : case of 10% missing entries in selected columns

25% missing

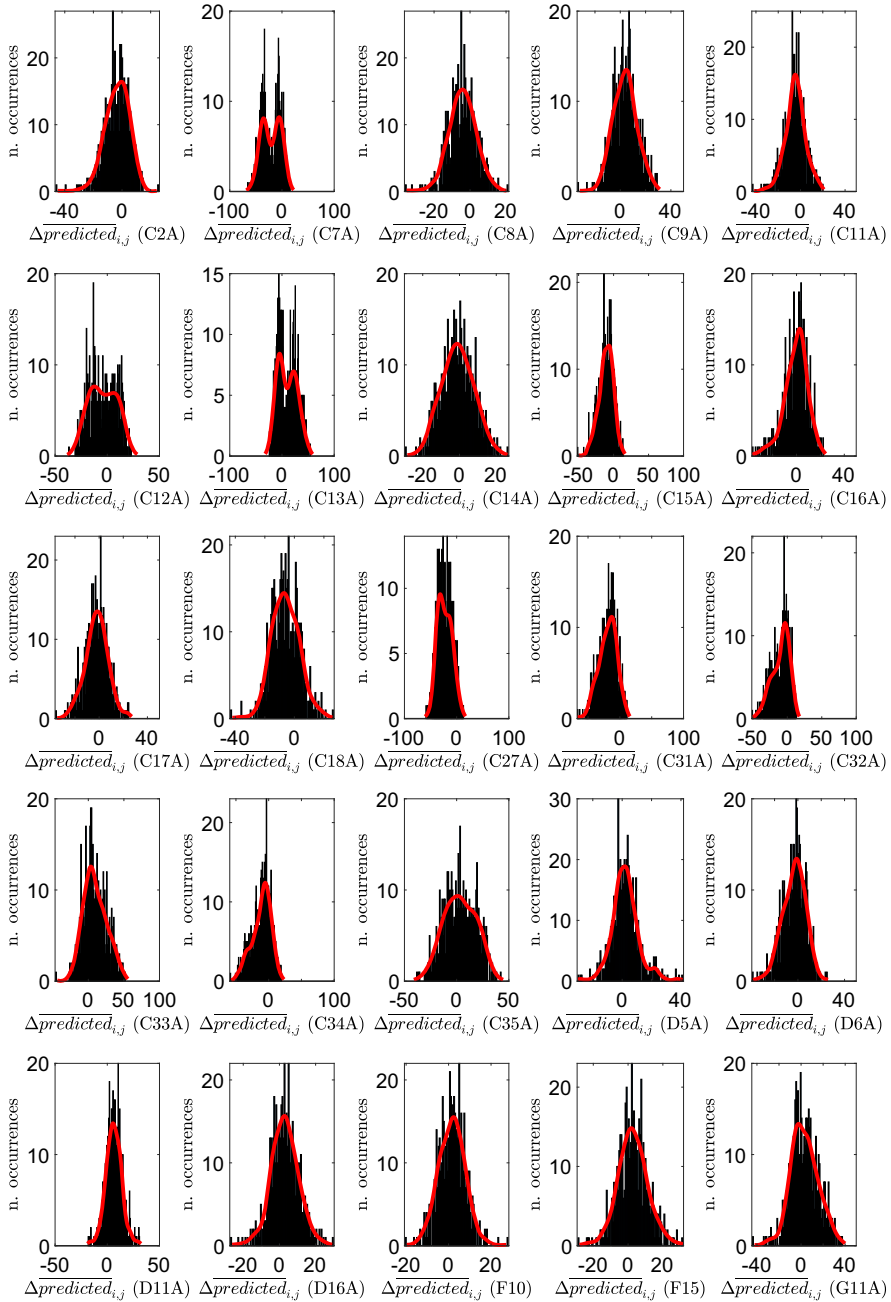


Fig. 10 Histograms of the variations $\Delta \overline{\text{predicted}}_{i,j}$ for the various creative soft skills j : case of 25% missing entries in selected columns

50% missing

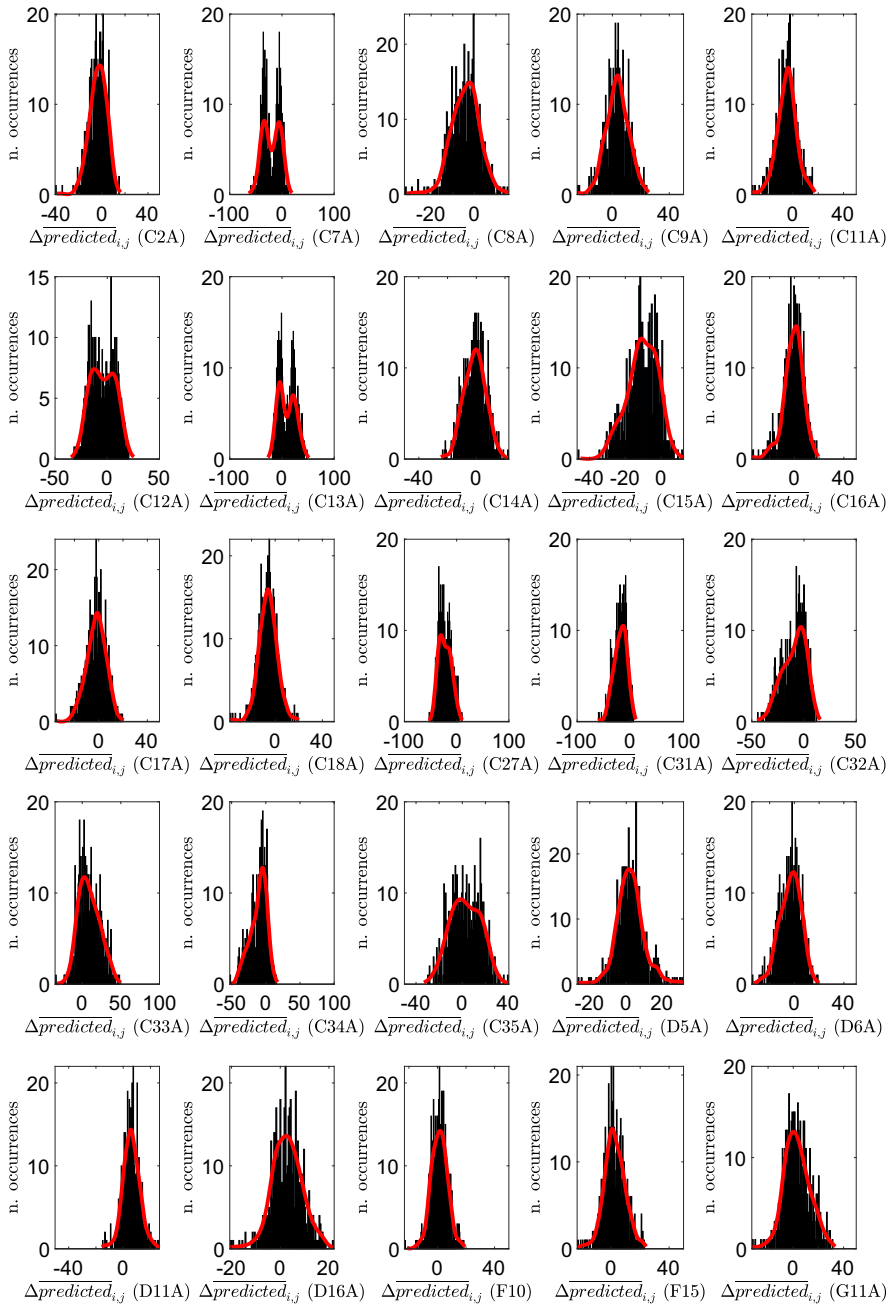


Fig. 11 Histograms of the variations $\Delta \overline{\text{predicted}}_{i,j}$ for the various creative soft skills j : case of 50% missing entries in selected columns

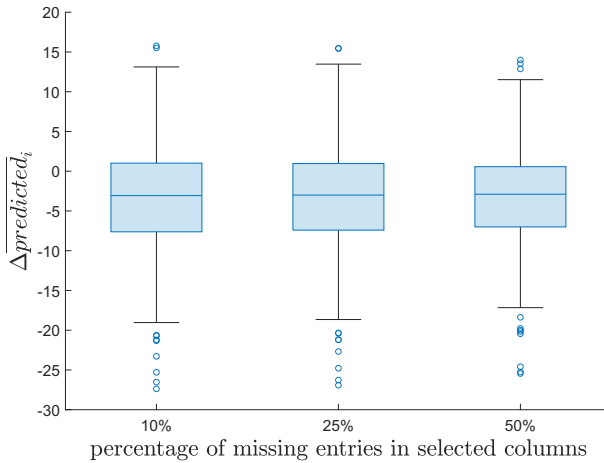


Fig. 12 Boxplots of the variations $\Delta \overline{\text{predicted}}_i$ for the three cases considered in the analysis

creativity patterns and trends are extremely heterogeneous and idiosyncratic and this means that there is no clear separation among technical and intellectual skills with respect to creative soft skills. Our study contributes to a nascent field on interdisciplinary research to predict the impact of AI on work activities and future jobs using machine learning (see also Poulakias 2021). Our results suggest that creativity gaps in the labor market are so peculiar that training might be tailored specifically for every occupation, to reduce its risk of being automated. In Italy, the policy response to the challenges posed by digitalization has been the extension of employment protection and short-time work to small firms that were not previously covered (Basso et al. 2022). Looking at the future, especially after recovering from the current Covid-19 pandemic, policy makers should take into account the possibility of better targeting the policies to sectors, occupations and firms as their patterns of sets of soft skills, creativity in particular, are highly heterogeneous between and within sectors. Skills demand in the country is increasing and changing rapidly, yet Italy is struggling to make the transition to a new digitalized economy where soft skills play a central role (OECD 2017). Making progresses will require that the country fosters both the demand and the supply of skills: for example, improving the quality of schooling that will, in time, translate into high-qualified workers, or encouraging and facilitating the transition to digital technologies for firms, so they can compete globally. In general, a wider strategy and a coordinated approach to skills policies is desirable to get Italy back on a growing path.

In our application, matrix completion has demonstrated an excellent prediction capability, making it meaningful to further analyze cases for which the actual creativity levels are significantly larger/smaller than the predicted ones, as the errors are possibly related to specific features of professions that distinguish them from other professions in the dataset. As a possible extension of our analysis, as soon as the data collected in the last edition of ICP (conducted in 2019) will be made publicly available, the method used in this article could also be applied to predict the most recent trends in

the labour market, and possibly also to make a forecast analysis for the future trends after the current COVID-19 pandemic (Barbieri et al. 2022).

Funding Open access funding provided by Scuola IMT Alti Studi Lucca within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Further Details on Matrix Completion and on Its Application to the Prediction of Creativity Levels for Different Professions

As already mentioned in the main text, in this work we adopt the following formulation for the Matrix Completion (MC) optimization problem, which was studied theoretically in Mazumder et al. (2010):

$$\underset{\mathbf{Z} \in \mathbb{R}^{m \times n}}{\text{minimize}} \left(\frac{1}{2} \sum_{(i,j) \in \Omega^{\text{tr}}} (M_{i,j} - Z_{i,j})^2 + \lambda \|\mathbf{Z}\|_* \right). \quad (5)$$

In the above, the regularization parameter controls the trade-off between the first data-fitting term and the second regularization term, which favors a small nuclear norm of the reconstructed matrix. The latter requirement is often related to getting a small rank of the obtained optimal solution \mathbf{Z}_λ^o of the optimization problem (5),¹² which follows by geometric arguments similar to the ones typically exploited to motivate the use of the classical LASSO (Least Absolute Shrinkage and Selection Operator) penalty term (Tibshirani 1996).

The connection between MC and singular values can be illustrated as follows. Given a nonzero matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$, and denoting its rank by $\text{rank}(\mathbf{M})$, it is well-known—by Eckart–Young theorem, see Theorem 2.1.2 in Moitra (2018)—that, for any $k \in \{1, \dots, \text{rank}(\mathbf{M})\}$, the best rank- k approximation of \mathbf{M} according to the Frobenius norm is provided by its truncated singular value decomposition, in which one keeps only the k -largest singular values of \mathbf{M} , and zeroes all the others. So, for an effective application of MC, one needs a quite rapid decay to 0 of the singular values of the matrix \mathbf{M} to be completed. This is an important necessary condition for such an effective application. However, such a condition is not sufficient when the

¹² In the notation, the dependence of that optimal solution on λ has been highlighted.

matrix is only partially observed, because finding the singular value decomposition itself requires the knowledge of the whole matrix.

In our application of the Soft Impute algorithm to an occupation matrix, we combine the original MATLAB implementation of Soft Impute from Mazumder et al. (2010) with the MATLAB function `svt.m` from Li and Zhou (2017). First, a specific row (profession) is fixed. All its entries in 25 given columns related to creativity are obscured, and form the test set (their set of positions is denoted by Ω^{test}). Then, one selects a percentage (respectively, 10%, 25%, 50%) of the rows of the matrix, excluding the given row. The entries in such rows are obscured in correspondence of the 25 given columns. These obscured entries form the validation set (their set of positions is denoted by Ω^{val}). All the remaining entries of the matrix form the training set. It follows from the construction above that the training, validation, and test sets do not overlap. To avoid overfitting, we select the regularization constant λ by means of the following validation method. The optimization problem (5) is solved for several choices λ_k for λ , exponentially distributed as $\lambda_k = 2^{k-1}$, for $k = 1, \dots, 15$. For each λ_k , the Root Mean Square Error (RMSE) of matrix reconstruction on the validation set is defined as

$$RMSE_{\lambda_k}^{\text{val}} := \sqrt{\frac{1}{|\Omega^{\text{val}}|} \sum_{(i,j) \in \Omega^{\text{val}}} (M_{i,j} - Z_{\lambda_k,i,j})^2}, \quad (6)$$

then the choice λ_k° that minimizes $RMSE_{\lambda_k}^{\text{val}}$ for $k = 1, \dots, 15$ is obtained. Finally, the RMSE of matrix reconstruction on the test set is evaluated in correspondence of the optimal value λ_k° as

$$RMSE_{\lambda_k^\circ}^{\text{test}} := \sqrt{\frac{1}{|\Omega^{\text{test}}|} \sum_{(i,j) \in \Omega^{\text{test}}} (M_{i,j} - Z_{\lambda_k^\circ,i,j})^2}. \quad (7)$$

The RMSE of matrix reconstruction on the training set is defined in a similar way. The whole procedure is repeated a sufficiently large number of times for each profession. This number is not the same for each profession due to the following method employed to reduce the total simulation time. Since the MC optimization problem (5) depends only on the training set, 200 random permutations of the rows of the matrix are generated. Each permutation is associated with a single training set. At the same time, it is associated with various different choices for the validation and test sets. More precisely, for each permutation, the first $h\%$ among such rows (with $h = 10, 25, 50$), plus the successive row in the permutation, form a set \mathcal{S} of rows, which is used to generate the union of the validation and test sets, by including in such union all the elements at the intersection between such rows and the selected columns. All the remaining elements of the matrix constitute the training set (one for each permutation). Then, $|\mathcal{S}|$ different test sets are constructed by including in each of them all the elements at the intersection between the selected columns and one specific row belonging to \mathcal{S} , in all possible $|\mathcal{S}|$ ways. The corresponding $|\mathcal{S}|$ validation sets are constructed by including in them all the elements at the intersection between the selected columns

and the remaining rows of \mathcal{S} , in all possible $|\mathcal{S}|$ ways. As already mentioned, in all these $|\mathcal{S}|$ cases, the training set is always the same, so the MC optimization problem (5) is solved once for each choice of \mathcal{S} . It is also clear that, in this way, the number of times a specific row appears in \mathcal{S} is not guaranteed to be the same for each row, because it depends on the specific set of random permutations generated. The number of such permutations (200) is selected in order to include each row in the sets \mathcal{S} with high probability for a sufficiently large number of times. Indeed, the expected number of times this occurs is about 20, 50, 100, respectively for $h = 10, 25, 50$. In this way, it is possible in principle to assess how the prediction on the test set associated with each row changes by varying the training and corresponding validation sets (see Fig. 4 in the main text for an example of this kind of analysis).

It is worth mentioning that in our application of MC to occupation matrices, all the very few missing entries of such matrices in the training set (when they are present) are replaced by zeros before running the Soft Impute algorithm. The value of the tolerance parameter of the algorithm is chosen as $\varepsilon = 10^{-7}$. Moreover, when convergence of the algorithm is not achieved, in order to reduce the simulation time, the algorithm is stopped after $N^{\text{it}} = 50$ iterations. An additional post-processing step is included, thresholding to 0 any negative element (when present) of the completed matrices, and to 100 any element (when present) larger than 100. This is justified by the fact that the elements of the occupation matrices represent percentages. In order to avoid introducing new notation, in all the previous equations the expression \mathbf{Z}_{λ_k} can be also redefined so as to refer to each post-processed MC output. The results reported in the main text refer actually to such post-processed MC output.

A final remark can be made about the trade-off between the prediction capability and the biasedness of the MC method (see also the discussion on this issue presented in the Supplementary Material of Gnecco et al. 2022). According to Foucart et al. (2017) and Ma and Chen (2019), biasedness in MC depends, for instance, on how unobserved entries are selected (in our case, only entries associated with selected skills are obscured, and for the professions for which this occurs, all the entries associated with such skills are actually obscured). Nevertheless, in general biasedness can be beneficial to prediction capability, because of the trade-off between bias and variance (Hastie et al. 2009).

Appendix B: Correspondence Between Codes and Selected Professions in Figs. 7 and 8

See Tables 7 and 8.

Table 7 Correspondence between codes and professions; healthcare sector, Fig. 7

Code	Profession
1.1.2.6.3	Health care directors
2.4.1.1.0	General doctors
2.4.1.2.0	Medical care specialists
2.4.1.3.0	Surgical care specialists
2.4.1.4.0	Laboratorians and clinical pathologists
2.4.1.5.0	Dentists and oral surgeons
2.4.1.6.0	Diagnostic imaging and radiotherapy specialists
2.4.1.7.1	Dietologists and hygienists
2.4.1.7.2	Social and occupational health specialists
2.4.1.7.3	Epidemiologists
2.4.1.8.0	Anaesthetists and critical care specialists
3.2.1.1.1	Nursing professions
3.2.1.1.2	Obstetrical professions
3.2.1.2.1	Podiatrists
3.2.1.2.2	Physiotherapists
3.2.1.2.3	Logopedists
3.2.1.2.4	Optometrists
3.2.1.2.5	Therapists in neuropsychomotricity
3.2.1.2.6	Therapists in psychiatric rehabilitation
3.2.1.2.7	Social educators
3.2.1.2.8	Occupational therapists
3.2.1.3.1	Audiologists
3.2.1.3.2	Biomedical laboratory assistants
3.2.1.3.3	Radiographers
3.2.1.3.4	Neurophysiopathologists
3.2.1.4.1	Orthopaedic technicians
3.2.1.4.2	Hearing care professionals
3.2.1.4.3	Dental hygienists
3.2.1.4.4	Cardiovascular pathophysiology and perfusion technicians
3.2.1.4.5	Dieticians
3.2.1.5.1	Technicians in prevention in the environment and the workplace
3.2.1.5.2	Health visitors
3.2.1.6.1	Opticians
3.2.1.6.2	Dental technicians
3.2.1.7.0	Popular medicine technicians
8.1.5.1.0	Janitors and assimilated
8.1.5.2.0	Aides and assimilated

Table 8 Correspondence between codes and professions; cultural sector, Fig. 8

Code	Profession
2.1.1.6.2	Paleontologists
2.5.3.2.4	Archaeologists
2.5.3.4.2	Art experts
2.5.4.1.2	Dialogists
2.5.4.5.1	Archivists
2.5.4.5.2	Librarians
2.5.4.5.3	Curators
2.5.5.1.1	Painters and sculptors
2.5.5.1.5	Conservators
2.5.5.2.1	Directors
2.5.5.2.2	Actors
2.5.5.2.3	Artistic directors
2.5.5.2.4	Screenwriters
2.5.5.2.5	Scenographers
2.5.5.3.1	Choreographers
2.5.5.3.2	Dancers
2.5.5.4.1	Composers
2.5.5.4.2	Conductors
2.5.5.4.3	Instrumentalists
2.5.5.4.4	Singers
2.5.5.5.1	Folk artists
2.5.5.5.2	Variety artists
2.5.5.5.3	Acrobats
3.1.2.6.2	Radio or TV broadcasters
3.1.7.2.1	Audio-video equipment and video shoots technicians
3.1.7.2.2	Sound technicians
3.1.7.2.3	Audio-video editors
3.3.4.7.0	Agents and representatives of artists and athletes
3.4.3.1.1	Radio or TV announcers
3.4.3.1.2	Presenters
3.4.3.2.0	Radio, television, cinema and theaters production technicians
3.4.4.1.2	Set designers
3.4.4.2.1	Museum technicians
3.4.4.2.2	Library technicians
3.4.4.3.1	Art estimators
3.4.4.3.2	Philatelic and numismatic experts
3.4.4.3.3	Calligraphy experts
3.4.4.4.0	Restoration technicians
4.4.2.2.0	Library officers

Table 8 continued

Code	Profession
5.4.2.1.1	Cinema and theatre merchants
6.5.5.1.0	Stagehands and property workers

References

- Acemoglu D, Autor D (2011) Skills, tasks and technologies: implications for employment and earnings. In: Handbook of labor economics, vol 4. Elsevier, pp 1043–1171
- Al-Otaibi ST, Ykhlef M (2012) A survey of job recommender systems. *Int J Phys Sci* 7(29):5127–5142
- Amabile TM (1983) The social psychology of creativity: a componential conceptualization. *J Pers Soc Psychol* 45(2):357–376
- Amabile TM, Pillemer J (2012) Perspectives on the social psychology of creativity. *J Creative Behav* 46(1):3–15
- Amabile TM, Pratt MG (2016) The dynamic componential model of creativity and innovation in organizations: making progress, making meaning. *Res Organ Behav* 36:157–183
- Amjad M, Shah D, Shen D (2018) Robust synthetic controls. *J Mach Learn Res* 19:15–51
- Athey S, Bayati M, Doudchenko N, Imbens G, Khosravi K (2021) Matrix completion methods for causal panel data models. *J Am Stat Assoc*. <https://doi.org/10.1080/01621459.2021.1891924>
- Autor DH (2014) Skills, education, and the rise of earnings inequality among the “other 99 percent”. *Science* 344(6186):843–851
- Autor DH, Levy F, Murnane RJ (2003) The skill content of recent technological change: an empirical exploration. *Q J Econ* 118(4):1279–1333
- Bachmann R, Gonschor M, Lewandowski P, Madoń K (2022) The impact of robots on labour market transitions in Europe. *Ruhr economic paper* 2022/933. <https://www.rwi-essen.de/publikationen/ruhr-economic-papers/1177/>
- Barbieri T, Basso G, Scicchitano S (2022) Italian workers at risk during the COVID-19 epidemic. *Ital Econ J* 8:175–195
- Baron-Cohen S (2000) Theory of mind and autism: a fifteen year review. In: Understanding other minds: perspectives from developmental cognitive neuroscience, vol 2, pp 3–20
- Basador M, Graen GB, Green SG (1982) Training in creative problem solving: effects on ideation and problem finding and solving in an industrial research organization. *Organ Behav Hum Perform* 30(1):41–70
- Basso G, Boeri T, Caiumi A, Paccagnella M (2022) Unsafe jobs, labour market risk and social protection. *Econ Policy*. <https://doi.org/10.1093/epolic/eiac004>
- Beaudry P, Green DA, Sand BM (2016) The great reversal in the demand for skill and cognitive tasks. *J Labor Econ* 34(S1):S199–S247
- Bessen J, Goos M, Salomons A, van den Berge W (2020) Automation: a guide for policymakers. *Economic Studies*, Brookings
- Boden MA (2004) *The creative mind: myths and mechanisms*. Psychology Press, New York
- Bonacini L, Gallo G, Scicchitano S (2021) Working from home and income inequality: risks of a ‘new normal’ with COVID-19. *J Popul Econ* 34:303–360
- Boucharad G, Yin D, Guo S (2013) Convex collective matrix factorization. In: *JMLR workshop and conference proceedings*, vol 31, pp 144–152
- Brynjolfsson E, McAfee A (2014) *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company, New York
- Camerer C, Loewenstein G, Prelec D (2005) Neuroeconomics: how neuroscience can inform economics. *J Econ Lit* 43(1):9–64
- Carbonero F, Scicchitano S (2021) Labour and technology at the time of Covid-19. Can artificial intelligence mitigate the need for proximity? GLO Discussion Paper Series 765, Global Labor Organization (GLO). <https://ideas.repec.org/p/zbw/glodps/765.html>
- Caselli M, Fracasso A, Scicchitano S, Traverso S, Tundis E (2021) Stop worrying and love the robot: an activity-based approach to assess the impact of robotization on employment dynamics. *GLO Discus-*

- sion Paper Series 802, Global Labor Organization (GLO). <https://ideas.repec.org/p/zbw/glodps/802.html>
- Castex G, Dechter E Kogan (2014) The changing roles of education and ability in wage determination. *J Labor Econ* 32(4):685–710
- Cetrulo A, Guarascio D, Virgillito ME (2020) Anatomy of the Italian occupational structure: concentrated power and distributed knowledge. *Ind Corp Change* 29(6):1345–1379
- Ciappei C, Cinque M (2014) Soft skills per il governo dell'agire. Franco Angeli (in Italian)
- Cirillo V, Evangelista R, Guarascio D, Sostero M (2021) Digitalization, routineness and employment: an exploration on Italian task-based data. *Res Policy* 50(7):104079
- Cockburn IM, Henderson R, Stern S (2018) The impact of artificial intelligence on innovation. NBER Working Paper no. 24449. National Bureau of Economic Research
- Csikszentmihalyi M (1988) Motivation and creativity: toward a synthesis of structural and energetic approaches to cognition. *New Ideas Psychol* 6(2):159–176
- De Bono E (1992) Six thinking hats for schools. Hawker Brownlow Education, Cheltenham
- De Bono E, Zimbalist E (1970) Lateral thinking. Penguin, London, pp 1–32
- Deming DJ (2017a) The growing importance of social skills in the labor market. *Q J Econ* 132(4):1593–1640
- Deming DJ (2017b) The value of soft skills in the labor market. NBER Rep 4:7–11
- Domini G, Grazzi M, Moschella D, Treibich T (2021a) For whom the bell tolls: the firm-level effects of automation on wage and gender inequality. LEM working paper 2021/25. <https://www.lem.sssup.it/WPLem/2021-25.html>
- Domini G, Grazzi M, Moschella D, Treibich T (2021b) Threats and opportunities in the digital era: automation spikes and employment dynamic. *Res Policy* 50(7):104137
- Easton E, Djumalieva J (2018) Creativity and the future of skills. Nesta. https://media.nesta.org.uk/documents/Creativity_and_the_Future_of_Skills_v6.pdf
- Esposito P, Scicchitano S (2022) Educational mismatch and labour market transitions in Italy: is there an unemployment trap? *Struct Change Econ Dyn* 61:138–155
- Eurofound—European Foundation for the Improvement of Living and Working Conditions (2005) Fourth European Working Conditions Survey. <https://www.eurofound.europa.eu/surveys/european-working-conditions-surveys/fourth-european-working-conditions-survey-2005>
- Eurofound—European Foundation for the Improvement of Living and Working Conditions (2015) Sixth European Working Conditions Survey. <https://www.eurofound.europa.eu/surveys/european-working-conditions-surveys/sixth-european-working-conditions-survey-2015>
- Fernández-Macias E, Klenert D, Antón J (2021) Not so disruptive yet: characteristics, distribution and determinants of robots in Europe. *Struct Change Econ Dyn* 58:76–89
- Fitjar RD, Timmermans B (2017) Regional skill relatedness: towards a new measure of regional related diversification. *Eur Plan Stud* 25:516–538
- Foucart S, Needell D, Plan Y, Wootters M (2017) De-biasing low-rank projection for matrix completion. In: Proceedings of SPIE, 10394, Wavelets and Sparsity XVII, 1039417
- Frey CB, Osborne M (2017) The future of employment: how susceptible are jobs to computerisation? *Technol Forecast Soc Change* 14:54–80
- Frigotto ML, Riccaboni M (2011) A few special cases: scientific creativity and network dynamics in the field of rare diseases. *Scientometrics* 89(1):397–420
- Galetti JRB, Tessarin MS, Morceiro PC (2021) Skill relatedness, structural change and heterogeneous regions: evidence from a developing country. *Pap Reg Sci* 100:1355–1376
- Glăveanu VP (2013) Rewriting the language of creativity: the five a's framework. *Rev Gen Psychol* 17:69–81
- Gnecco G, Nutarelli F, Riccaboni M (2022) A machine learning approach to economic complexity based on matrix completion. *Sci Rep* 12:9639
- Goldfarb A, Taska B, Teodoridis F (2020) Could machine learning be a general-purpose technology? Evidence from online job postings. SSRN. <http://dx.doi.org/10.2139/ssrn.3468822>
- Gualtieri V, Guarascio D, Quaranta R (2018) Does routinization affect occupation dynamics? Evidence from the 'Italian O* Net' data. MPRA Paper 89585, University Library of Munich
- Harrigan J, Reshef A, Toubal F (2016) The march of the techies: technology, trade, and job polarization in France, 1994–2007. NBER Working Paper no. 22110. National Bureau of Economic Research
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction. Springer, New York
- Hastie T, Tibshirani R, Wainwright M (2015) Statistical learning with sparsity: the lasso and generalizations. CRC Press, Boca Raton

- Hennessey BA, Amabile TM (2010) Creativity. *Annu Rev Psychol* 61:569–598
- Hidalgo CA (2021) Economic complexity theory and applications. *Nat Rev Phys* 3:92–113
- Hoffman EB, Malacrino D, Pistaferri L (2021) Labor market reforms and earnings dynamics: the Italian case. IMT Working paper no. 21/142. <https://www.imf.org/-/media/Files/Publications/WP/2021/English/wpiea2021142-print-pdf>
- Iansiti M, Lakhani KR (2020) Competing in the age of AI: strategy and leadership when algorithms and networks run the world. Harvard Business Press, Boston
- INAPP (2013) Indagine Campionaria sulle Professioni (ICP) dataset. Access to the dataset is possible only for research purposes and can be requested to INAPP according to the procedure detailed at https://inapp.org/it/dati/accesso_micro_file
- James RF, James ML (2004) Teaching career and technical skills in a “mini” business world. In: Business Education Forum, vol 59. National Business Education Association, pp 39–41
- Katz LF, Margo RA (2014) Technical change and the relative demand for skilled labor: the United States in historical perspective. In: Human capital in history: the American record. University of Chicago Press, pp 15–57
- Kharkhurin AV (2014) Creativity. 4in1: four-criterion construct of creativity. *Creativity Res J* 26:338–352
- Laker DR, Powell JL (2011) The differences between hard and soft skills and their relative impact on training transfer. *Hum Resour Dev Q* 22(1):111–122
- Lanzolla G, Pesce D, Tucci CL (2021) The digital transformation of search and recombination in the innovation function: tensions and an integrative framework. *J Prod Innov Manag* 38(1):90–113
- Lebuda I, Csikszentmihalyi M (2018) All you need is love: the importance of partner and family relations to highly creative individuals’ well-being and success. *J Creative Behav* 54(1):100–114
- Levy F, Murnane RJ (2012) The new division of labor: how computers are creating the next job market. Princeton University Press, Princeton
- Li C, Zhou H (2017) Svt: singular value thresholding in MATLAB. *J Stat Softw* 81(2). <https://doi.org/10.18637/jss.v081.c02>.
- Lindqvist G (2003) Vygotsky’s theory of creativity. *Creativity Res J* 15:245–251
- Ma W, Chen GH (2019) Missing not at random in matrix completion: the effectiveness of estimating missingness probabilities under a low nuclear norm assumption. In: Advances in neural information processing systems, vol 32, Proceedings of the 33rd conference on neural information processing systems (NeurIPS 2019), Vancouver
- Mazumder R, Hastie T, Tibshirani R (2009) Regularization methods for learning incomplete matrices. arXiv preprint [arXiv:0906.2034](https://arxiv.org/abs/0906.2034)
- Mazumder R, Hastie T, Tibshirani R (2010) Spectral regularization algorithms for learning large incomplete matrices. *J Mach Learn Res* 11:2287–2322
- Metulini M, Gnecco G, Biancalani F, Riccaboni M (2022) Hierarchical clustering and matrix completion for the reconstruction of input-output tables. *ASTA Adv Stat Anal*. <https://doi.org/10.1007/s10479-022-04653-z>
- Moitra A (2018) Algorithmic aspects of machine learning. Cambridge University Press, Cambridge
- Mumford MD (2003) Where have we been, where are we going? Taking stock in creativity research. *Creativity Res J* 15:107–120
- Mumford MD, Scott GM, Gaddis BH, Strange JM (2002) Leading creative people: orchestrating expertise and relationships. *Leadersh Q* 13(6):705–750
- Muzio E, Fisher DJ, Thomas ER, Peters V (2007) Soft skills quantification (SSQ) For project manager competencies. *Project Manag J* 38(2):30–38
- Neffke F, Henning M (2013) Skill relatedness and firm diversification. *Strateg Manag J* 34(3):297–316
- Niu W, Sternberg R (2002) Contemporary studies on the concept of creativity: the east and the west. *J Creative Behav* 36(4):269–288
- OECD (2017) OECD skills strategy diagnostic report: Italy. <https://www.oecd.org/italy/oecd-skills-strategy-diagnostic-report-italy-2017-9789264298644-en.htm>
- Pellerey M (2016) Orientamento come potenziamento della persona umana in vista della sua occupabilità: il ruolo delle soft skills, o competenze professionali personali generali. *Rassegna Cnos* 1:41–49
- Pérez LD, Nebot AP, Capmany AB, Dima G, Mula A, Gonzalez P, Piligrimiè Z, Pilinkiene V, Riccio A, Fattore L (2020) National reports on the labour market soft-skill. <https://skills4employability.eu/wp-content/uploads/2020/08/National-Reports-on-Soft-Skills-current-scenarios-by-Skills4Employability-Erasmus-Project-IO1.A3-.pdf>

- Poulakias K (2021) Artificial intelligence and job automation: an EU analysis using online job vacancy data. CEDEFOP Working paper series no. 6. https://www.cedefop.europa.eu/files/6206_en.pdf
- Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? *Behav Brain Sci* 1(4):515–526
- Rainsbury E, Hodges DL, Burchell N, Lay MC (2002) Ranking workplace competencies: Student and graduate perceptions. *Asia Pac J Coop Educ* 3(2):8–118
- Recht B (2011) A simpler approach to matrix completion. *J Mach Learn Res* 12(12)
- Remus D, Levy FS (2016) Can robots be lawyers? Computers, lawyers, and the practice of law
- Riccaboni M, Franceschetti M, Landi S (2021) Competenze creative e sistemi di raccomandazione. Technical report, INAPP
- Ricci F, Rokach L, Shapira B (2011) Introduction to recommender systems handbook. *Recommender systems handbook*. Springer, pp 1–35
- Runco MA, Jaeger GJ (2012) The standard definition of creativity. *Creativity Res J* 24(1):92–96
- Silvestri D, Riccaboni M, Della Malva A (2018) Sailing in all winds: technological search over the business cycle. *Res Policy* 47(10):1933–1944
- Simonton DK (2012) Taking the U.S. patent office criteria seriously: a quantitative three-criterion creativity definition and its implications. *Creativity Res J* 24(2–3):97–106
- Sternberg RJ, Lubart TI (1995) *Defying the crowd: cultivating creativity in a culture of conformity*. Free Press, New York
- Sternberg RJ, Lubart TI (1999) The concept of creativity: prospects and paradigms. *Handb Creativity* 1:3–15
- Thibodeaux G (2014) Three versions of constructionism and their reliance on social conditions in social problems research. *Sociology* 48(4):829–837
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B (Methodol)* 58(1):267–288
- Tucciarelli M (2014) Coaching e sviluppo delle soft skills. *La Scuola, Brescia (in Italian)*
- Vannutelli S, Scicchitano S, Biagetti M (2021) Routine biased technological change and wage inequality: do workers' perceptions matter? GLO Discussion Paper Series 763, Global Labor Organization (GLO). <https://ideas.repec.org/p/zbw/glodps/763.html>
- Venckutė M, Berg Mulvik I, Lucas B (2020) Creativity—a transversal skill for lifelong learning: an overview of existing concepts and practices. In: Bacigalupo M, Cachia R, Kampylis P (eds) Publications Office of the European Union, Luxembourg
- Walia C (2019) A dynamic definition of creativity. *Creativity Res J* 31(3):237–247
- West MA, Farr JL (eds) (1990) *Innovation and creativity at work: psychological and organizational strategies*. Wiley, New York
- Windsor G, Bakhshi H (2015) The creative economy and the future of employment. Nesta. https://media.nesta.org.uk/documents/the_creative_economy_and_the_future_of_employment.pdf

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.