Workplace Skills as Regional Capabilities: Relatedness, Complexity and Industrial Diversification of Regions

Duygu Buyukyazici, Leonardo Mazzoni, Massimo Riccaboni & Francesco Serti

Papers in Evolutionary Economic Geography

22.10

Utrecht University Human Geography and Planning

Workplace Skills as Regional Capabilities: Relatedness, Complexity and Industrial Diversification of Regions

Duygu Buyukyazici[1], Leonardo Mazzoni[1], Massimo Riccaboni[1], and Francesco Serti[1]

[1]IMT School for Advanced Studies, Lucca, Italy

Abstract

The literature reaches a unanimous agreement that industrial diversification is path-dependent because new industries build on preexisting capabilities of regions that are partly embodied and reflected in the skills of regions' workforce. This paper explicitly accounts for regional capabilities as workforce skills to build skill relatedness and complexity measures, skill-spaces, for 107 Italian regions for the period 2013-2019. Data-driven techniques we use reveal that skill-spaces form two highly polarised clusters into social-cognitive and technical-physical skills. We show that industries have a higher (lower) probability of developing comparative advantage if their required skill set is (not) similar to those available in the region regardless of the skill type. We find evidence that similarity to technical-physical skills and higher complexity in social cognitive skills yields the highest probabilities of regional competitive advantage.

Keywords: Skill relatedness; Economic complexity; Industrial specialisation; Regional capabilities; Regional diversification.

JEL Classification: J24; O18; R10; R23.

1 Introduction

A large strand of the evolutionary economic geography (EEG) literature has shown that regions build on existing capabilities and diversify into related activities (Boschma, 2017; Hidalgo et al., 2018; Balland et al., 2019; Balland and Boschma, 2019). *Existing capabilities* is an extensively used, broad term for productive inputs that range from tangible assets, *i.e.*, physical capital, labour force, to intangible assets, *i.e.*, norms, institutions, knowledge and skills. What the literature has been somewhat reluctant to do is to explicitly address capabilities, despite diversification is seen as a process in which new activities emerge from new combinations of existing capabilities in regions (Antonietti and Boschma, 2021). In this context, the meaning and measurement of relatedness have been questioned (Tanner, 2014; Boschma, 2017) in the sense that relatedness stands for the similarity of entities that share common capabilities, yet the measurement of relatedness seldom include capabilities.

The mainstream practice, co-occurrence based relatedness, uses co-location numbers of entities as input data to compute location quotients to further use them to calculate the minimum of conditional probabilities that two entities co-occur together. Then, these two entities are assumed to share similar capabilities if they have a high relatedness score, yielding an outcome-based assessment and interpretation of regional capabilities. This approach has generated criticism in the literature. Kogler (2017, pp. 2) underlines that *"co-location of activities is frequently an a priori assumption of connectedness."* Whittle and Kogler (2019, pp. 8) ask *"Related to what?"* and as a solution, they emphasise the potential of skill relatedness to embrace capabilities at micro level: *"... this approach has the benefit of capturing relatedness at a micro level via occupational data and labour mobility, something that the previous relatedness measures have either failed to do or done so inadequately."*.

Skill relatedness concept is introduced by Neffke and Henning (2013) who developed the revealed skill relatedness method (RSR) that assumes two industries to be skill-related if they exhibit intensive labour mobility. Skill similarity is, therefore, an a priori assumption and relatedness is indirectly observed (Neffke and Henning, 2008). In this paper, we argue that regional capabilities are partly embodied and reflected in the knowledge, abilities, and skills of a region's workforce employed by the region's industry mix. Based on this argument, we aim to further elaborate the existing practice by explicitly accounting for regional capabilities as workforce skills in the measurement process of relatedness. We also consider the complexity approach, along with relatedness, to adequately account for different aspects of capabilities.

Relying on a unique data set on workplace skills, the Italian Sample Survey on Professions (ICP), we construct skill relatedness and complexity matrices for 534 industries for each of the 107 Italian NUTS 3 regions[1] over the period 2013-2019, based on the 161 workplace skills' intensities. First, we provide an overview of these measures, formalised as *the skill space* of the industrial profile of Italian regions. By employing data-driven methods, we investigate the similarity between skill types in terms of their effective use by industries and how their degree of relatedness and complexity differ. The results indicate that workplace skills are clustered into two main communities: social-cognitive skills and technical-physical skills. We find that social-cognitive skills have the highest complexity scores while technical-physical skills are generally below average. We then deploy econometric models to estimate the impact of relatedness and complexity of different skill

[1]NUTS 3 regions known as *provinces* in Italy.

types on the process of industrial specialisation of regions. The results indicate that, regardless of the skill type, higher skill relatedness and complexity enhance the probability of building a new comparative advantage, whilst lower skill relatedness and complexity increase the probability of losing an established comparative advantage. However, similarity to technical-physical skills yields a higher (lower) probability of entry (exit) than social-cognitive skills. Conversely, higher complexity in social-cognitive skills is associated with a slightly higher probability of entry. We also find that social-cognitive skill relatedness does not enhance or mitigate specialisation process of service industries, while technical-physical skill relatedness does. On the contrary, specialisation process of manufacturing industries is affected by the relatedness of both types of skills.

The present work contributes to the literature in four ways. First, to the best of our knowledge, it is the first empirical study that explicitly analyses the relationship between skill relatedness based on workplace skills and industrial diversification, *i.e.,* entry and exit probability, at the regional scale in a developed country.

Second, this is the first study that considers the role of different skill types in industrial and regional diversification literature. So far, skill relatedness research for regional diversification has employed the RSR method, forming *implicitly* skill-related industries (Elekes et al., 2021). The presence of detailed and reliable data on workplace skills enables us to use skill scores in input matrices to *explicitly* define skill relatedness measures for different skill types.

Third, this study provides the first empirical attempt to consider the skill complexity of industries and to jointly analyse skill relatedness and skill complexity in explaining path-dependencies of industrial diversification of regions. The present study shows that the skill complexity of industries, along with skill relatedness, plays a role in diversification process.

Fourth, we contribute to the relatedness literature by providing more geographical wisdom to relatedness as the literature tends to treat relatedness as a global phenomenon (Boschma, 2017). Nevertheless, region specific capabilities might cause the degree of relatedness between activities to differ across regions. The high granular analysis we conduct shows that skill relatedness of a particular industry dramatically changes across regions, indicating that relatedness is not independent of its spatial context. Differences in regional workforce skills of the same industry might conceptualise a convincing answer to the question of why some industries are related in one spatial context and not related in another.

The structure of the paper is as follows. Section 2 briefly overviews the related literature. Section 3 describes data sources. Section 4 introduces skill relatedness and skill complexity measures. Section 5 presents econometric analyses. Section 6 provides sensitivity analyses. Section 7 overviews the main findings and concludes.

2 Literature

2.1 Regional Capabilities, Diversification, and Skill Relatedness

The main argument behind regional diversification is that new activities emerging in a region are prone to be related to the region's preexisting activities in terms of capabilities, knowledge, and skills, which is called *relatedness* (Hidalgo et al., 2007; Neffke et al., 2011; Hidalgo et al., 2018). As Neffke and Henning (2013) pointed out, there are three different approaches to measure relatedness:

(1) hierarchical measure based on standard industry classification systems such as NACE and SIC (Chang, 1996; Farjoun, 1998; Lee and Lieberman, 2010), (2) outcome-based co-occurrence methods (Neffke and Henning, 2008, Hidalgo et al., 2007), (3) resource-based measures such as technological resources (Breschi et al., 2003), human-capital resources (Farjoun, 1994)[2].

Among the three measures the literature has identified thus far, co-occurrence based relatedness, only *relatedness* hereafter, has gained prominence. Notably, after the pioneering work of Hidalgo et al. (2007), scholars have developed relational networks using the relatedness concept to shed light on the diversification capabilities of a variety of activities such as trade (Hidalgo et al., 2007; Boschma et al., 2013; Guo et al., 2020), technologies (Kogler et al., 2013; Boschma et al., 2015; Tanner, 2016; Montesor and Quatraaro, 2017; Balland et al., 2019), jobs (Muneepeerakul et al., 2013; Farinha et al., 2019), skills (Neffke and Henning, 2013; Neffke et al., 2017) and industries (Neffke et al., 2011; Xiao et al., 2018) at multiple spatial scales, including countries, regions or cities (Whittle, 2019). In this literature, capabilities have generally been captured by co-location patterns of different entities. If two entities co-exist, their spatial units are assumed to have similar capabilities. For instance, Hidalgo et al. (2007) measure relatedness between trade products based on their co-location patterns, arguing that two trade products demand similar capabilities if product pairs happen to be frequently present in the same location. Neffke et al. (2011) capture regional capabilities by product relatedness between manufacturing plants. In Rigby (2015), capabilities are addressed by technological relatedness that is derived from the co-occurrence of patent classes on patent documents. Accordingly, capabilities are seen as an enabling but implied source of regional diversification whose exact nature is not directly observed (Boschma, 2017). This partly blind approach has generated criticism in the literature (Kogler, 2017), while some relatedness types are seen as more capable of providing a more explicit measure, such as skill relatedness (Whittle and Kogler, 2019).

Studies on skill relatedness, still in its infancy, can be divided into two major approaches with respect to their methodology. The first approach employs *the revealed skill relatedness* (RSR) method developed by Neffke and Henning (2013) and applied by Timmermans and Boschma (2014); Boschma et al. (2014); Diodato and Weterings (2015); Fitjar and Timmermans (2017); Neffke et al. (2017); Neffke et al. (2018); Elekes et al. (2021). Neffke and Henning (2013) underline that different industries in the economy seem to be interconnected by linkages of skill similarity, and firms are much more likely to diversify into the activities related to their core competencies. They take a micro perspective and assume that, when switching jobs, individuals tend to remain in industries related to their previous job in terms of skill content. Based on this assumption, they develop a method that employs co-occurrence techniques to assess skill-relatedness between industries by using cross-industry labour flows in Sweden. They predict expected flows between industries to compare them to actual labour flows. Industries that exhibit mobility more than the expected flows are assumed to be skill-related. The results show that the RSR index is predictive of firm diversification. Fitjar and Timmermans (2017) propose an improvement to the RSR method with a measure at the industry and region level.

Inter-industry labour flows might unravel exciting insights, yet such data are not available for many countries, making the RSR method challenging to employ. Moreover, the RSR method provides an implied measure of skill similarity between industries as it does not use workplace skills

[2]See Neffke and Henning (2013) and Whittle (2019) for brief reviews.

data, deducing skill relatedness from labour mobility. Excessive labour mobility between industries does not necessarily indicate high skill similarity. Neffke and Henning (2013) emphasise that the RSR method is an indirect measure of skill relatedness, and intense labour flows might also stem from similar corporate culture and social network effects. Hence the RSR index may also capture other effects than skill similarity introducing a substantial risk of overvaluation of skill relatedness.

Furthermore, the RSR method does not allow horizontal differentiation of skills and employs an agnostic attitude to different skill types due to the lack of workplace skills data. Previous studies, therefore, have been limited to *industry space*, an industry-to-industry formalisation of labour-flow based skill relatedness, not being able to provide a skill-to-skill analysis of industries. Consequently, the impact of different skill types on the industrial diversification of regions has not been considered so far. In addition, these drawbacks have prevented previous studies from incorporating skill complexity into the analyses, leading to an either/or approach towards skill relatedness and skill complexity. However, sophisticatedness of skills may unravel valuable insights into the regional diversification process.

The second methodology for skill-relatedness uses data on workplace skills, employs co-occurrence methods à la Hidalgo, and combines them with network analyses to construct a skill space, a network formalisation of skill-to-skill relatedness matrix. This approach can differentiate between skill types and does not require labour-flow data, unlike the RSR method. Nevertheless, only a couple of studies hitherto applied this method to focus on the relationship between different skill types. Anderson (2017) uses online freelance job market data with network methods to categorise worker skills based on their relationship with other skills in the market. She argues that workers with diverse and synergistic skills earn higher wages than others. Alabdulkareem et al. (2018) construct the skill space of occupations using skill relatedness based on co-occurrence methods combined with network analyses. They apply a community detection algorithm to the skill space and find two highly polarised skill clusters associated with the polarisation of wages and the hollowing out of the middle-wage occupations. However, these studies apply skill relatedness to occupations globally; none considers skill relatedness at the industry and region level in the context of regional diversification.

The present study combines and elaborates on these two approaches. We use workplace skills to compute relatedness as in the second approach and consider skill relatedness in the industrial diversification process of regions as in the first approach. To our knowledge, no empirical study has analysed the relationship between skill relatedness, based on workplace skills, and industrial diversification, *i.e.*, entry and exit probability, at the regional scale in a developed country.

2.2 Regional Capabilities, Diversification, and Skill Complexity

In their pioneering work, Hidalgo and Hausmann (2009) introduce a measure of economic complexity that draws valuable information from the structure of a bipartite country -product network, *i.e.*, product space. Their measure is based on the method of reflections (MOR), which iteratively combines two variables: products' ubiquity and countries' diversity. Consequently, countries with non-ubiquitous, untradable capabilities enjoy an exclusive source of comparative advantage, thereby producing more complex and privileged goods that only a small fraction of countries can produce. On the contrary, countries experiencing less exclusive, ordinary capabilities produce ubiquitous goods that many countries can produce; thus, they tend to have low complexity scores. The economic complexity approach based on MOR has been replicated by many scholars[3] using either trade data and product space or different relational networks, including knowledge space (Balland and Rigby, 2017), digital proximity (Rahmati et al., 2021), technological complexity (Whittle, 2019; Mewes and Broekel, 2020) and skill complexity (Caines et al., 2017; LoTurco and Maggioni, 2020; Antonietti et al., 2021).

The empirical work related to skill complexity is still scarce. LoTurco and Maggioni (2020) construct an occupational complexity measure by exploiting the knowledge and skill requirements of occupations to investigate the labour content of complex products. They find that the new measure of occupational complexity is more predictive of GDP growth in the USA than other measures. Antonietti et al. (2021) build three complexity measures, occupation, task, and skill, to explore the impact of key enabling technologies (KEY) on workforce demand. They show that a larger share of KEY increases the demand for more complex occupations, tasks, and skills.

To our knowledge, no study has yet investigated skill complexity in industrial diversification literature, nor with skill relatedness. The present study aims to fill this gap by arguing that skill complexity is highly relevant in the diversification process.

Regional research has come to a consensus that regional resources which underpin the diversification process are often localised, non-tradable and non-ubiquitous (Neffke et al., 2018), implying that a region's workplace skills are mostly region-specific and spatially dependent. The complexity approach introduces an effective method to quantify these aspects of regional skills in terms of their diversity, ubiquity, and sophisticatedness in a comparable way to other regions' skills. Therefore, complexity is complementary to relatedness rather than being rival. In this paper, we consider relatedness and complexity as two different mechanisms that are beneficial to unravelling the path-dependent nature of the diversification process.

3 Data

The data for the present study derives from a couple of sources. The primary occupational and industrial data source is the Italian labour Force Survey (ILFS) provided by the National Institute of Statistics of Italy (Istat). The data on workplace skills, the Italian Sample Survey on Professions (ICP)[4], is obtained from the National Institute for Public Policies Analysis (INAPP). ICP collects detailed information on the characteristics of all professions existing in the Italian labour market, particularly on the content of the work, knowledge and skills it requires, and the organisational structure where the work takes place. The ICP data provide extremely granular and valuable information on workplace skills and reflect the Italian labour market structure. The concept and survey questions are borrowed from the Occupational Information Network (O*net)[5] that is run by the Bureau of Labour Statistics in the USA. Up to date, there are two ICP waves: 2007 and 2013. In each wave, almost 16.000 workers are interviewed, representing approximately 800 occupational units at the five-digit level in the context of the Classificazione delle Professioni (CP), which is the Italian version of ISCO classification[6]. We use only one wave, the ICP 2013, to match the

[3] See Balland et al. (2022) for an overview.
[4] See https://www.inapp.org/it/dati/ICP for more details.
[5] See https://www.onetcenter.org for details.
[6] The ICP 2007 uses CP 2001, while the ICP 2013 uses CP 2011.

occupational classification scheme of ILFS.

Each wave consists of seven sections, each of which captures one aspect of occupations:

- Knowledge (33 questions, both importance and complexity level);
- Skills (35 questions, both importance and complexity level);
- Attitudes (52 questions, both importance and complexity level);
- Generalised working activities (41 questions, both importance and complexity level);
- Values (21 questions);
- Working styles (16 questions);
- Working conditions (57 questions).

The major ICP sections can further be broken down into more specific, homogeneous, and task-related sub-categories, which are presented in Table A1. The sub-categories, or the ICP descriptors, are adapted from the O*net data descriptors. Each competence in sub-categories is cross-checked against the ICP questionnaire. The codes in parentheses stand for the ICP question number that we use to identify workplace skills throughout the paper.

The four major sections of ICP data appear to be well suited for this study: knowledge, skills, attitudes, and generalised working activities, which sum up to 161 skill variables. Mainly because they use the same question design that accounts for both the importance and level of the competence in question [7]. The remaining sections have various question designs and scales, including only importance, frequency, time, agreement, et cetera. Hereafter, we use the term *skills* in a broader sense to refer to these four ICP sections.

ICP data refer only to occupational categories at the five-digit level. Therefore, we use ILFS data to connect workplace skills to spatial and industrial information. We constructed the main data set as follows. First, we transform ICP data from the five-digit occupational level (796) to be at the four-digit level (507), given that ILFS is available at the four-digit level. Then we generate skill intensity variables for each workplace skill by multiplying importance scores with level scores in a similar fashion as it was done for O*net variables in the works of Feser (2003), Gabe and Abel (2011), and Krenz (2014). This multiplicative approach increases the skill variation across occupations (Feser, 2003).

Second, we merge ICP and ILFS data sets on the four-digit occupational level. By doing so, we end up with 497[8] occupational categories and 161 workplace skills. Occupations match on a one-to-one basis, given that both data sets are classified according to CP 2011.

Lastly, we compute average skill intensity scores for each industry. Recall that the merged ICP/ILFS data set provides skill distributions of occupations (from ICP), occupational distributions

[7]Importance question: *How important is competence in carrying out your current profession?* Level question: *Among those indicated below, at what level is this competence necessary for the development of your current profession?* Importance questions rated on a scale from 1 (not important) to 5 (extremely important), while complexity level questions rated on a scale from 1 (least complex) to 7 (most complex). Then they are rescaled to be between 0 and 100.

[8]The ICP sample does not contain Armed Forces, we thus excluded these occupational categories. Legislators and Senior Officials are recoded due to aggregation differences between ICP and ILFS.

of industries (from ILFS) and industrial portfolios of regions in each year (from ILFS). We use these distributions to generate average skill intensity variables for each industry in each region and year. After excluding part-time workers, and individuals out of the 15-64 age range, the resulting sample consists of 161 average skill intensity variables for 534 industries and 107 regions for the period 2013-2019.

For the econometric analyses, we use employment (*Local Units and Persons Employed: Size Class of Persons Employed, Economic Activities, Geographical Areas*) and business register (*ASIA Business Register Database*) data from Istat to construct dependent variables. The data for control variables are extracted from both Eurostat (GDP per capita, business growth, churn) and Istat (population density, education, industrial ubiquity, regional diversity).

4 Construction of Skill Space

There are four main entities in this study: workplace skills ($N_s = 161$), regions ($N_p = 107$), industries ($N_i = 534$), and time ($N_t = 7$). We first construct industry-skill input matrices, $\mathbf{C}_{is}$, for each region in each year, summing up to 749 matrices. Each input matrix consists of 161 skills and 534 industries whose each cell $x_{i,s}$ contains the skill intensity of industry i for skill s ($i = 1, ..., n; s = 1, ..., m$) in region p in year t. We then compute skill relatedness matrices and skill complexity vectors by using these input matrices as defined below.

4.1 Skill Relatedness of Industries

We estimate skill relatedness between skill pairs based on the framework proposed by Hidalgo et al. (2007). The first step is to define the effective use of skills. Skill $s \in S$ is effectively used by industry $i \in I$ if its relative skill advantage (RSA) is greater than 1. RSA is the share of the relative importance of skill s to industry i (the numerator in equation 1) to the relative importance of skill s to all industries I (the denominator). RSA is a measure based on the Balassa index, also known as location quotient (LQ) and revealed comparative advantage (RCA). A higher value of RSA indicates a higher level of importance of skill s for industry i compared to the overall importance of skill s for all other industries. We apply the RSA formula to the industry-skill matrix $\mathbf{C}_{is}$. The result is (NxM) two-mode adjacency matrix $\mathbf{M} = (M_{i,s})$ where $M_{i,s} = 1$ if RSA is greater than 1, *i.e.,* industry i effectively uses skill s, and $M_{i,s} = 0$ otherwise. Figure 2 presents several realisations of matrix $\mathbf{M}$.

$$RSA(i,s) = \frac{icp(i,s) \setminus \sum_{s' \epsilon S} icp(i,s')}{\sum_{i' \epsilon I} icp(i',s) \setminus \sum_{i' \epsilon I, s' \epsilon S} icp(i',s')} \tag{1}$$

After we identify effectively used skills by each industry, we compute the skill relatedness of industries between each pair of effectively used skills based on the minimum conditional probability of their co-occurrences in industry classes as formulated in equation 2[9].

[9]The standardisation method we used is 'cosine', and implemented with Econ Geo R package by Balland (2016). See VanEck and Waltman (2009) for the standardisation methods of co-occurrence data.

$$R(s,s^{'}) = \frac{\sum_{i \epsilon I} e(i,s).e(i,s^{'})}{max(\sum_{i \epsilon I} e(i,s), \sum_{i \epsilon I} e(i,s'))} \tag{2}$$

where effective use of skills denoted as $e(i,s) = 1$ if $RSA > 1$, and $e(i,s) = 0$ otherwise. The resulting matrix is the skill relatedness index (MxM) of N industries which contains proximities between two skill types s and $s^{'}$. Each cell $(s,s^{'})$ represents the probability that a random industry effectively uses skill $s(s^{'})$ effectively uses skill $s^{'}(s)$ as well.

The skill relatedness index is an adjacency matrix of 161 skills based on the co-occurrence analysis of 534 industries in region p in year t. Applying equations 1 and 2 to all input matrices leaves us with 749 skill-to-skill relatedness indexes. We first analyse these indexes at the national level, averaged for regions and years.

Figure 1 displays the skill relatedness matrix of Italy's industrial portfolio formalised as a one-mode network that we coin as *the skill space*[10] for the period 2013-2019. The individual nodes represent 161 skills and are coloured by the skill sub-categories presented in Table A1. The edges between skills indicate their degree of relatedness. The visualised network is fully connected, *i.e.*, we do not allow for any isolates. All skills are presented in the graph, although the network is thresholded (relatedness≥0.45) for visualisation purposes.

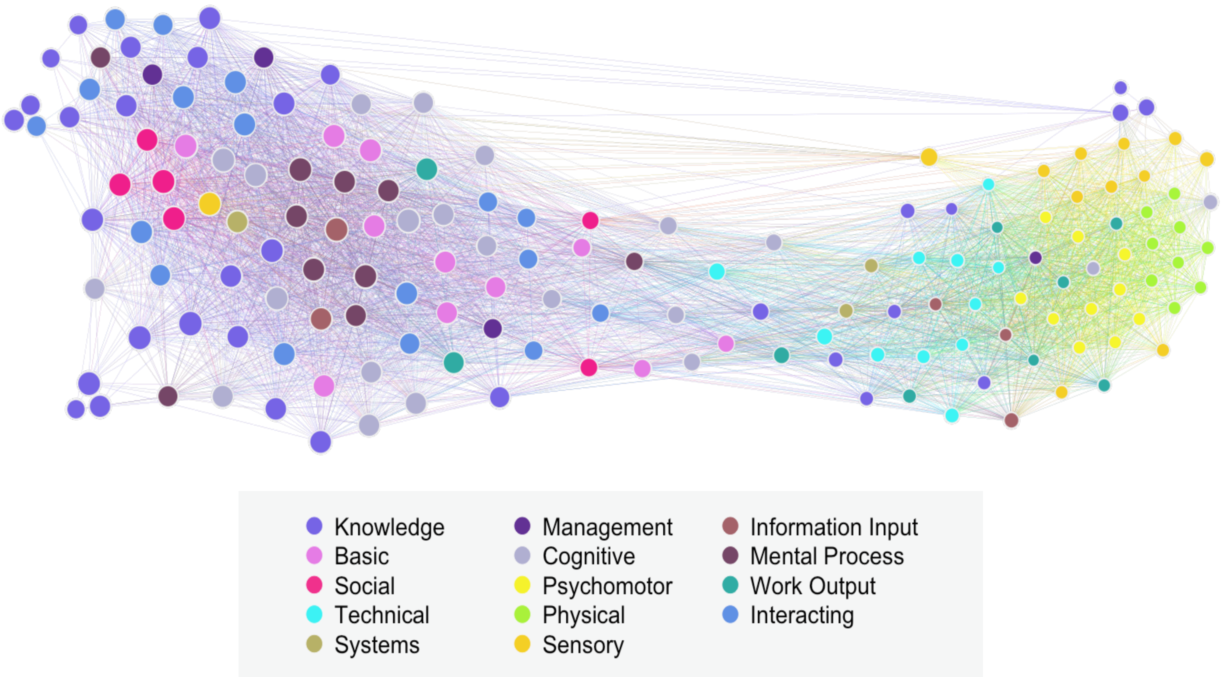
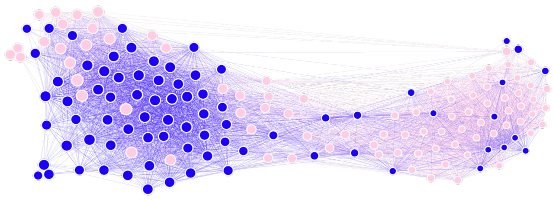
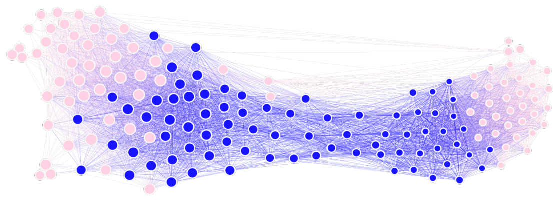


Figure 1. The Skill Space (2013-2019). Nodes represent skills. The size of each node is proportional to the complexity level of the skill the node represents. Nodes are coloured to the subcategories of skills. Edge lengths show the degree of relatedness between skill pairs.

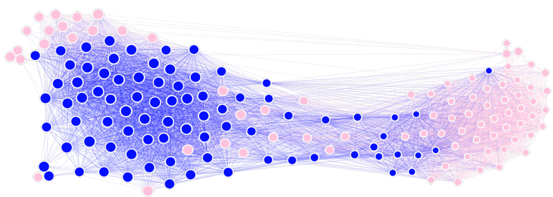
[10]The network is visualised in Gephi software by using Multi-Scale Force Atlas, which is a force-based algorithm. The two highly polarised skill clusters are robust to different layout algorithm choices.



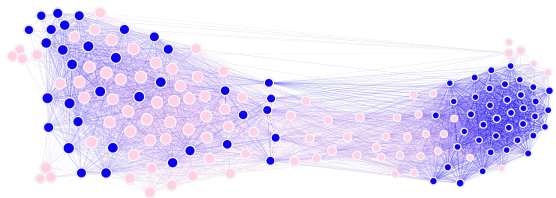
(a) Artistic creation



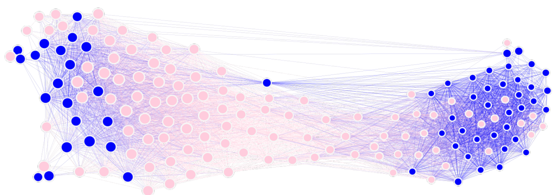
(b) Manufacture of electronic components



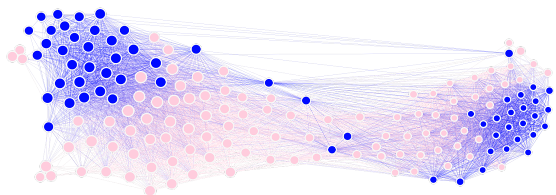
(c) Satellite telecommunications activities



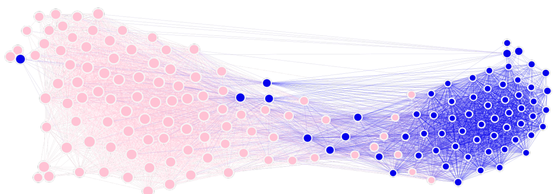
(d) Restaurants and mobile food service activities



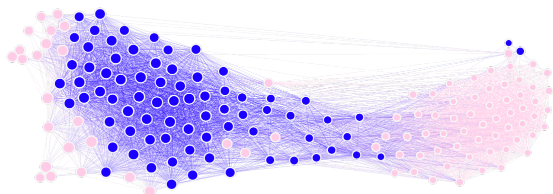
(e) Taxi operation



(f) Retail sale of textiles in specialised stores



(g) Construction of roads and motorways

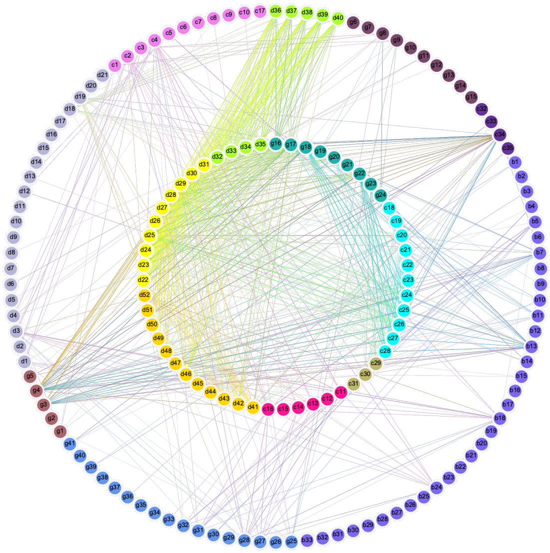


(h) Activities of holding companies

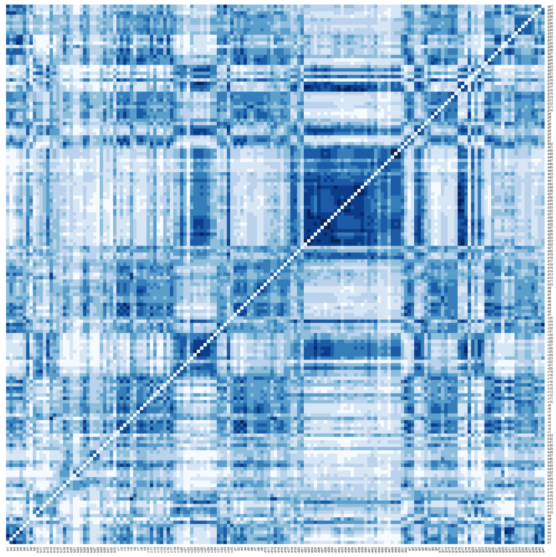
Figure 2. Example Industries and Effective Use of Skills. Skills in which the industries have $RSA > 1$ are projected onto the skill space using navy nodes.

A striking feature at first glance is that the skill space explicitly forms two highly polarised skill clusters. Physical, psychomotor, sensory, systems, and technical skills are located on the right. Basic, social, management, interacting, and higher concentration of knowledge and cognitive skills are located on the left. More related skills tend to cluster together; therefore, the skill space gives a first impression of which skills are mostly used together by industries.

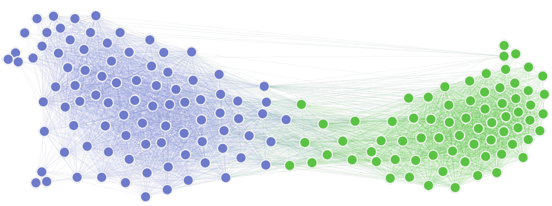
We further analyse the skill polarisation by employing a data-driven community detection method. Figure 3(c) presents the detected communities with the multi-level modularity optimisation



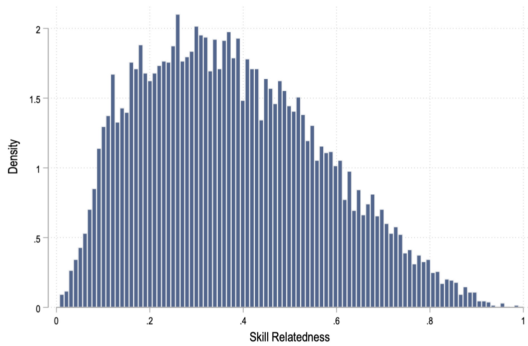
(a) Most related skills (> 0.75)



(b) Skill relatedness matrix



(c) Detected communities



(d) Density of skill relatedness

Figure 3. The Properties of the Skill Space

(Louvain) algorithm (Blondel et al., 2008) and affirms the polarisation of skills. The resulting modularity is 0.42, showing that partition is solid and community structure is significant. The two communities of skills are unfolded in Table A2.

The polarisation of workplace skills is in line with the findings of Alabdulkareem et al. (2018). They analyse the skill relatedness of occupations in the USA labour market and find two polarised skill clusters, which they coin as *social-cognitive* and *sensory-physical*. We embrace a similar approach and term the two skill clusters as *social-cognitive* and *technical-physical* skills.

Figure 3 gives further insights into the structure of the skill space. Figure 3(a) shows the most related skills, *i.e.,* relatedness score above 0.75, using a dual-circular layout. Figure 3(b) displays the skill space as a similarity matrix. Figure 3(d) presents the density of pairwise skill relatedness scores.

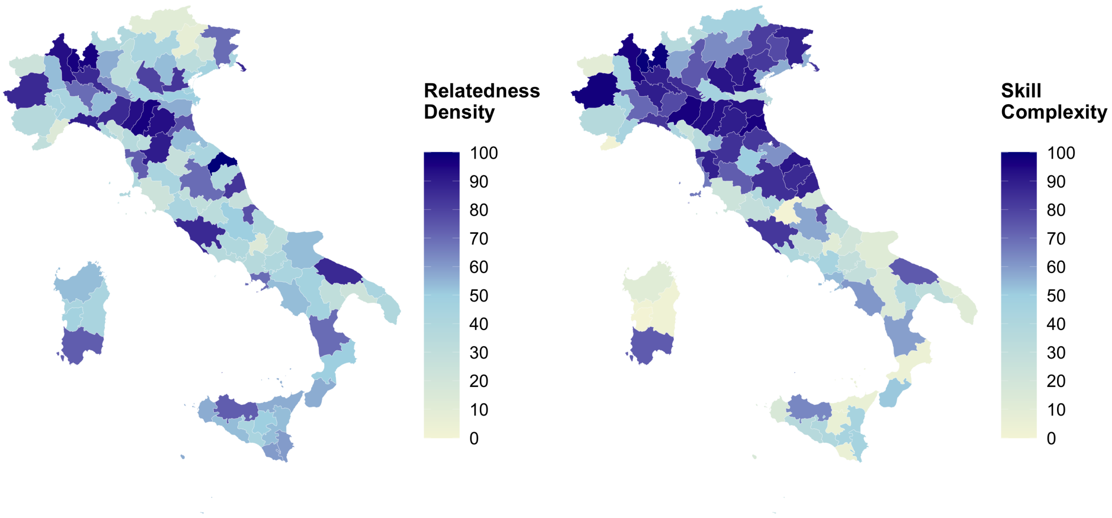
The illustrative and quantitative evidence provided thus far is representative at the national level. It is also possible to analyse the skill space with additional dimensions such as time and space. Recall that the skill relatedness formula in equation 2 yields a skill-to-skill matrix; therefore, the dimensional extension requires an additional indicator of relatedness. To this end, we employ a measure of relatedness density (equation 3) that is developed by Hausmann and Klinger (2007) and reformulated by Balland et al. (2015) as *technological flexibility measure*. We use RSA (equation 1) and the skill relatedness index (equation 2) to construct the density measure.

$$AverageRelatednessDensity_{i,t}^{p} = \frac{\sum_{s\in i}(\frac{\sum_s \phi_{s,j,t} RSA_{s,i,t}}{\sum_s \phi_{s,j,t}} \times 100)_{s,i,t}}{\sum_{s\in i} s} \tag{3}$$

where $\phi_{s,j,t}$ refers to relatedness between skill s and j at time t; $RSA_{s,i,t}$ is a binary variable that takes the value of 1 if industry i effectively uses skill s at time t, and takes the value of 0 otherwise. Accordingly, skill relatedness density around industry i at time t in region p is defined as the sum of relatedness values between all the skill pairs that industry i effectively uses, divided by the sum of relatedness between all the skill pairs available in region p at time t. As a result, the skill relatedness density formula, located inside the parentheses in the numerator of equation 3, gives a matrix (MxN) whose each cell indicates relatedness density between skill s and industry i. We use this skill relatedness density matrix to calculate the average skill relatedness density (ASRD) of industry i in region p as the sum of relatedness densities of skills that industry i effectively uses to the sum of all skill relatedness densities available in region p at time t. ASRD measure allows us to compare the required skill portfolios of different industries across regions and years by combining information on the use of 161 different workplace skills. ASRD will be high for a new industry if its skill space is similar to other industries in which the region has an RSA. In other words, ASRD measures how close a potential new industry is to the region's existing industry mix in terms of human capital.

Figure 4(a) demonstrates ASRD scores averaged for Italian regions for the period 2013-2019. The higher the ASRD of region, the closer its existing set of effectively used skills to the missing skills in the region. This is to say that the region is closer to forming the necessary capabilities to specialise in new industries that require different sets of workplace skills than the existing ones. The figure displays considerable differences in the branching potentials of regions. The northern and upper-central regions have higher potential than the lower-central and southern regions.

Figure 5 displays the dynamic nature of ASRD, exampled by three industries for 2013 and 2019. The subfigures indicate that the same industry might use different skills and be skill-related to different industries in different locations, underlining the importance of regional scale and inadequateness of global relatedness measures.



(a) (b)

Figure 4. Skill Relatedness Density and Skill Complexity Index of Italian Regions

4.2 Skill Complexity of Industries

The complexity measure used in the present study is *the method of reflections* (MOR), introduced in the pioneering work of Hidalgo and Hausmann (2009). MOR sequentially combines two measures: diversity and ubiquity. Diversity ($K_{i,0}$) is the number of skills effectively used ($RSA > 1$) by industry i. Ubiquity ($K_{s,0}$) is the number of industries that effectively use a particular skill s.

$$Diversity = k_{i,0} = \sum_{s} M_{i,s} \tag{4}$$

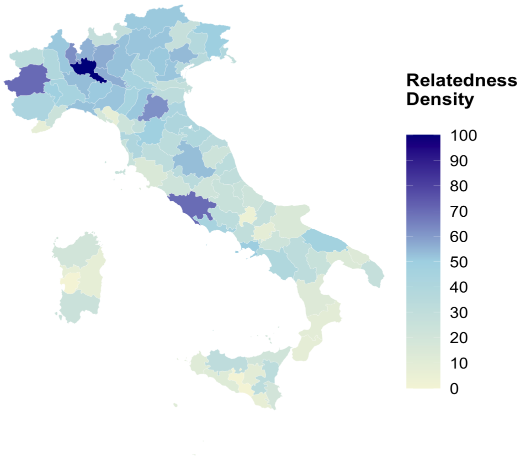
$$Ubiquity = k_{s,0} = \sum_{i} M_{i,s} \tag{5}$$

where $M_{i,s}$ is an adjacency matrix of industries and skills, resulting from equation 1. $M_{i,s} = 1$ if industry i effectively uses ($RSA > 1$) skill s, and $M_{i,s} = 0$ otherwise. These measures indeed summarize the structural properties of a bipartite industry-skill network. Diversity measures the number of edges between industry i and skills, while ubiquity measures the number of edges each particular skill has. After sequentially combining diversity and ubiquity measures for $N \geq 1$ steps, MOR is defined as iterative linear equations that are theoretically infinite. In other words, MOR iteratively calculates the average value of the previous-level properties of a nodeâs neighbours as defined in the following equations.

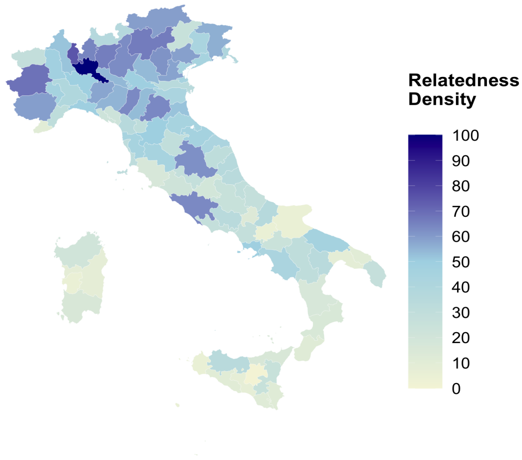
$$K_{i,N} = \frac{1}{k_{i,0}} \sum_{s} M_{i,s} k_{s,N-1} \tag{6}$$

$$K_{s,N} = \frac{1}{k_{s,0}} \sum_{i} M_{i,s} k_{i,N-1} \tag{7}$$

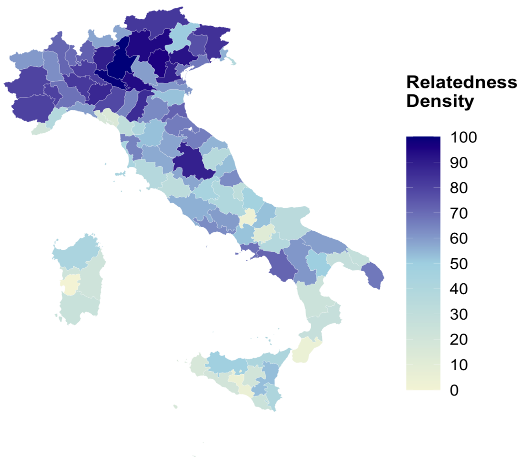
For industries ($K_{i,N}$), even variables ($K_{i,0}, K_{i,2}, K_{i,4}...$) are general measures of skill diversification; odd variables ($K_{i,1}, K_{i,3}, K_{i,5}...$) are measures of ubiquity of their effectively used skills. For skills



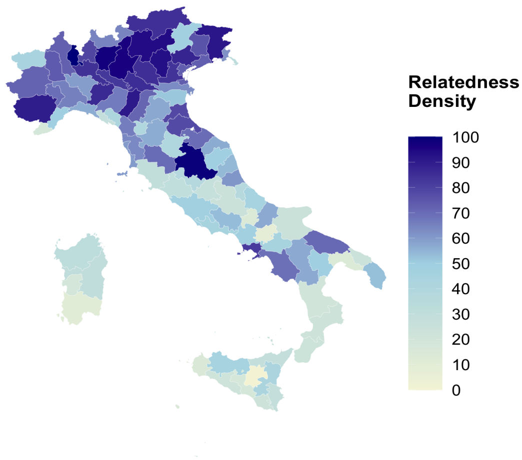
(a) Manufacture of electronic components, 2013



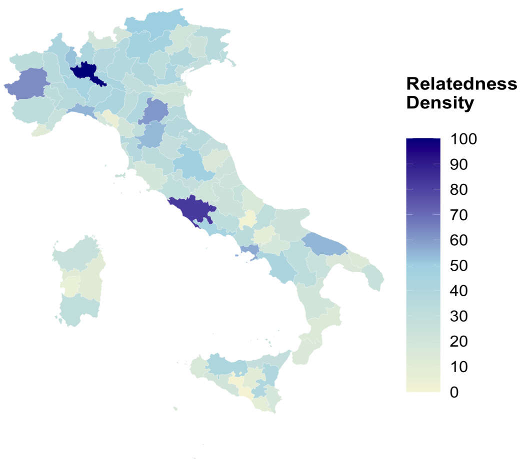
(b) Manufacture of electronic components, 2019



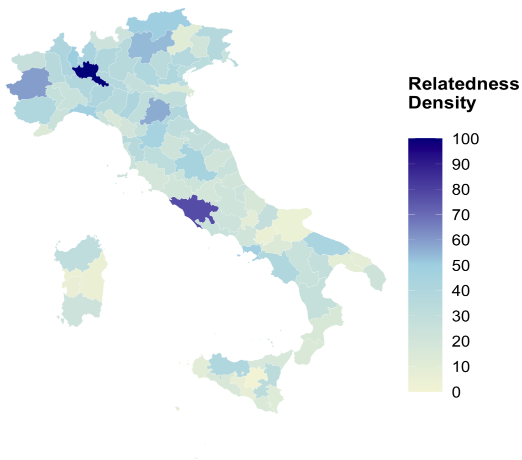
(c) Construction of roads and motorways, 2013



(d) Construction of roads and motorways, 2019



(e) Activities of holding companies, 2013



(f) Activities of holding companies, 2019

Figure 5. Differences in Relatedness Densities Across Industries, Regions and Years

($K_{s,N}$), even variables are related to their ubiquity, while odd variables are related to the diversification of industries that effectively use those skills.

Higher order iterations can be interpreted as a linear combination of the properties of all the nodes in the network that we briefly name complexity score. Complexity score converges to a certain value as the number of iterations increases.

Based on the equations above, the skill complexity of industry i is defined by the composition of the skill set necessary to perform tasks required by that industry and the relative composition of skill sets of all other industries. In other words, a higher skill complexity score for industry i would reflect a relatively higher number of effectively used skills (diversity component) that are relatively effectively used by a small number of industries (ubiquity and scarcity component).

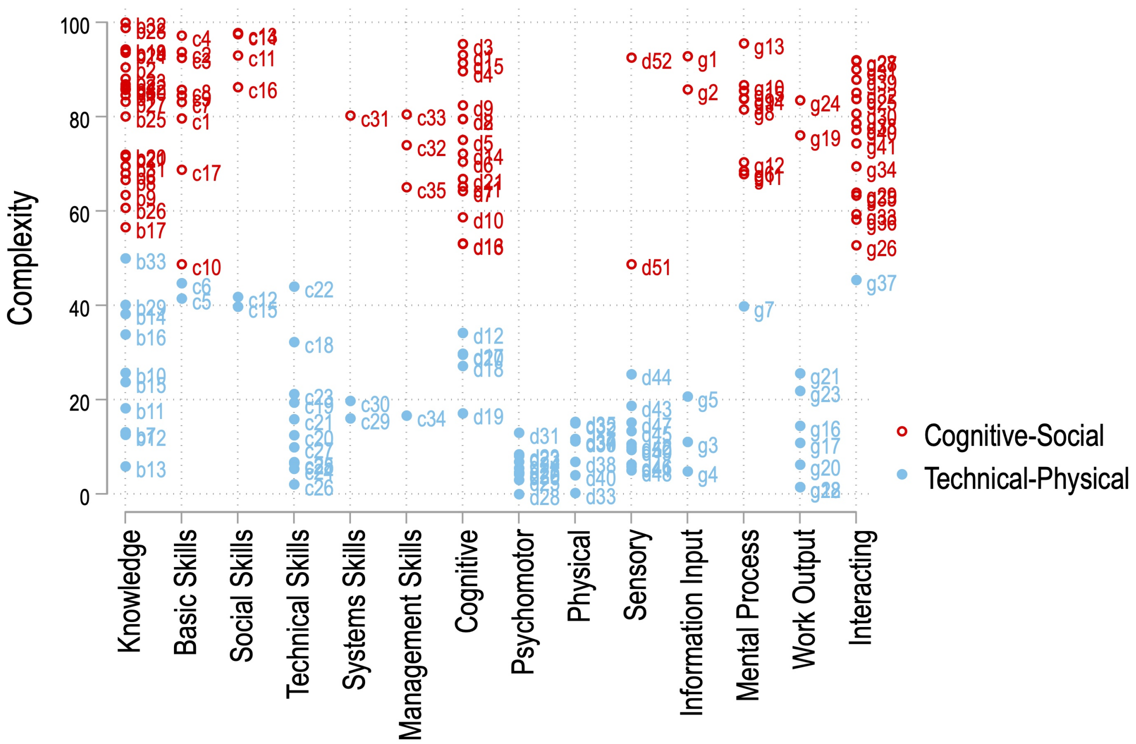


Figure 6. Complexity levels of workplace skills.

Figure 6 displays skill complexity scores of workplace skills (obtained after 17 iterations of MOR) on the y-axis. The sharp divide between social-cognitive and technical-physical skills is striking, posing further evidence for the skill polarisation of industries. Almost all technical-physical skills have below-average complexity scores. Transportation (B33), science (C6), and programming (C22) are the most complex technical-physical skills. Conversely, social-cognitive skills exhibit much higher complexity scores. Communication and media (B32), persuasion (C13), and negotiation (C14) are the most complex social-cognitive skills.

Figure 7 exhibits skill complexity scores of industries[11] at the four-digit NACE level. Construction,

[11] (B-E) Industry, (F) Construction, (G) Wholesale and retail trade; repair of motor vehicles and motorcycles, (I) Accommodation and food service activities, (H) Transportation and storage, (J) Information and communication, (K-N/X) Financial, real estate, scientific and technical, administrative and support service activities, (P/Q) Education; human health and social work activities, (R-U) Arts, entertainment and recreation; other service activities.

a high share of manufacturing, transportation, and storage industries have relatively lower skill complexity. In contrast, education, human health, social work activities, financial and scientific activities, and information and communication industries have the highest skill complexity scores.

Figure 4(b) presents the skill complexity index (based on 17 iterations of MOR) of Italian regions, averaged for the period 2013-2019. Higher skill complexity score for a region indicates diverse and exclusive workplace skills in the population. Apparently, the northern and central regions have higher complexity scores than the southern regions.

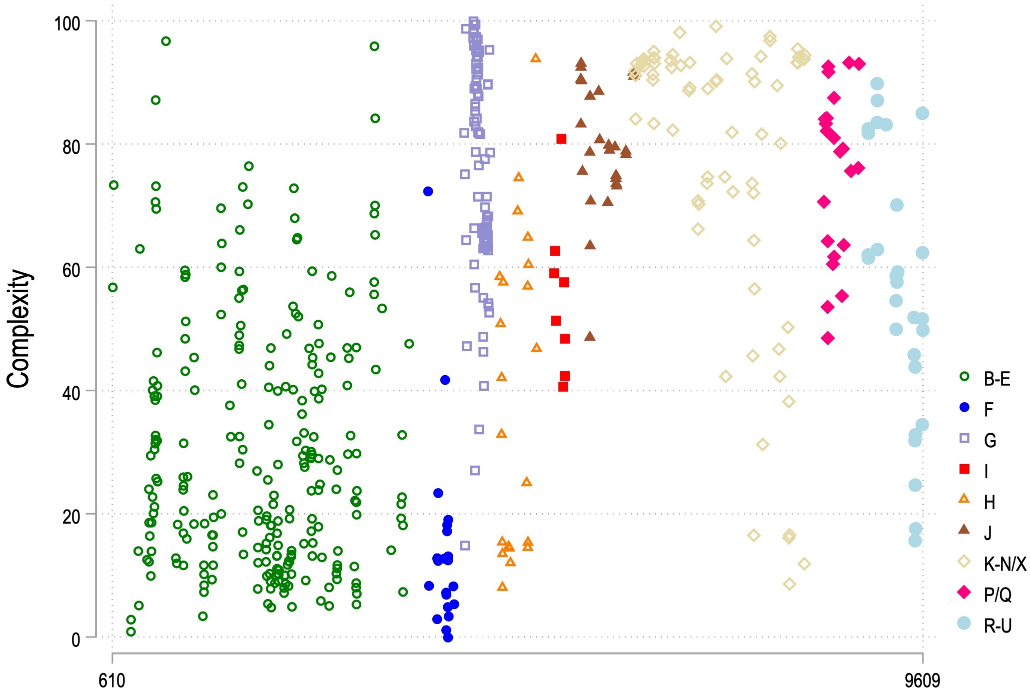


Figure 7. Skill complexity scores of industries.

5 Econometric Analyses: Industrial Diversification

This section conducts econometric analyses to assess the impact of skill relatedness and complexity on the probability that a region specialises in a new industry (entry) or loses it (exit).

We construct two binary dependent variables, *Entry* and *Exit*, to account for the industrial specialisation of regions. *Entry* equals to one if industry i develops a specialisation in region p, *i.e.*, if industry i has revealed comparative advantage (RCA) in region p in period t while it did not have RCA in period $t-3$, and *Entry* equals to zero otherwise. *Exit* equals to one if industry i had RCA in region p in period $t-3$ and loses it in period t. More formally:

$$Entry_{i,p,t} = 1,\ if\ RCA_{i,p,t} > 1\ and\ RCA_{i,p,t-3} \leq 1 \tag{8}$$

$$Exit_{i,p,t} = 1,\ if\ RCA_{i,p,t} \leq 1\ and\ RCA_{i,p,t-3} > 1 \tag{9}$$

$$RCA_{i,p,t} = \frac{E_{i,p,t}/E_{i',p,t}}{E_{i,p',t}/E_{i',p',t}} \tag{10}$$

where $E_{i,p,t}$ is employment in industry i in region p in period t. A higher RCA means that industry i has higher employment in region p in period t, than its average employment in all regions p' in period t. In other words, higher RCA indicates over-presence of an industry in a given region. Tables A3 and A4 provide summary statistics for the dependent variables.

The dependent variables are binary; therefore, logistic, probit, and linear probability (LPM) models can be used. The Hausman test procedure[12] suggests fixed effects with our data set. It is known that logistic and probit models can be inconsistent when estimated with large fixed effects (Greene, 2012). Accordingly, by following the literature (Bahar et al., 2018; Xiao et al., 2018; Farinha et al., 2019; Balland et al., 2019), we use LPM to estimate the following specification at the industry-region level for four overlapping periods of three-years[13].

$$
\begin{aligned}
Y_{i,p,t} &= [Entry_{i,p,t}, Exit_{i,p,t}] \\
Y_{i,p,t} &= \beta_1 ASRD_{i,p,t-1} + \beta_2 SkillComplexity_{i,p,t-1} + \\
&\quad \beta_3 RegionControls_{i,p,t-1} + \beta_4 IndustryControls_{i,p,t-1} + \\
&\quad \beta_5 GeoDiversification + \rho_p + \iota_i + \gamma_t + \varepsilon_{i,p,t}
\end{aligned}
$$

The principal coefficients of interest are β_1 and β_2. β_1 captures the impact of ASRD of industries on the probability of developing a comparative advantage in a region or losing an existing one. We expect a positive sign in the entry models. If an industry has a more similar skill set to the existing skill portfolios of industries located in a particular region, then that industry would have a relatively higher probability of being specialised in that region. The region already possesses the necessary capabilities that the new industry can combine. Conversely, we expect a negative sign in the exit models, mainly because if an industry's required skill set is not similar to the skill sets of preexisting industries in a particular region, then that industry would have a higher probability of losing specialisation in that region since the region does not have the required capabilities thus the industry must nurture them. β_2 measures the impact of skill complexity on the probability of new industrial specialisations. We anticipate a positive coefficient in the entry models and a negative one in the exit models. If an industry has a complex skill set with respect to other industries in the region, it is more likely to produce more valuable and unique goods and services, thus more (less) likely to develop (lose) comparative advantage.

In the previous section, we showed that workplace skills are divided into two clusters with respect to their effective use by industries. In order to investigate how these skill clusters reflect themselves in the industrial diversification process, ASRD and skill complexity are computed for three different skill aggregation levels: full skills (161 skill types that are presented in Table A1), social-cognitive skills (one of the detected skill clusters, displayed in the first section of Table A2), technical-physical skills (another skill cluster, presented in the second section of Table A2). Figure 8 presents their distributions. The inner-quartile ranges of ASRD variables are pretty narrow compared to skill complexity variables. Regarding the social-cognitive skill cluster, 75% of scores

[12] The Hausman test has been performed with different models, including logistic, probit, and LPM. The null "difference in coefficients not systematic" was rejected in each model specification.

[13] 2013-2016, 2014-2017, 2015-2018, 2016-2019.

are below 50 and below 45 for the technical-physical skill cluster. The skill complexity variables exhibit a more balanced distribution across quartiles. The average social-cognitive skill complexity of industries is around 62, while it is 46 for technical-physical skill complexity. The majority of industries exhibit higher ASRD and complexity scores for the social-cognitive skill cluster.

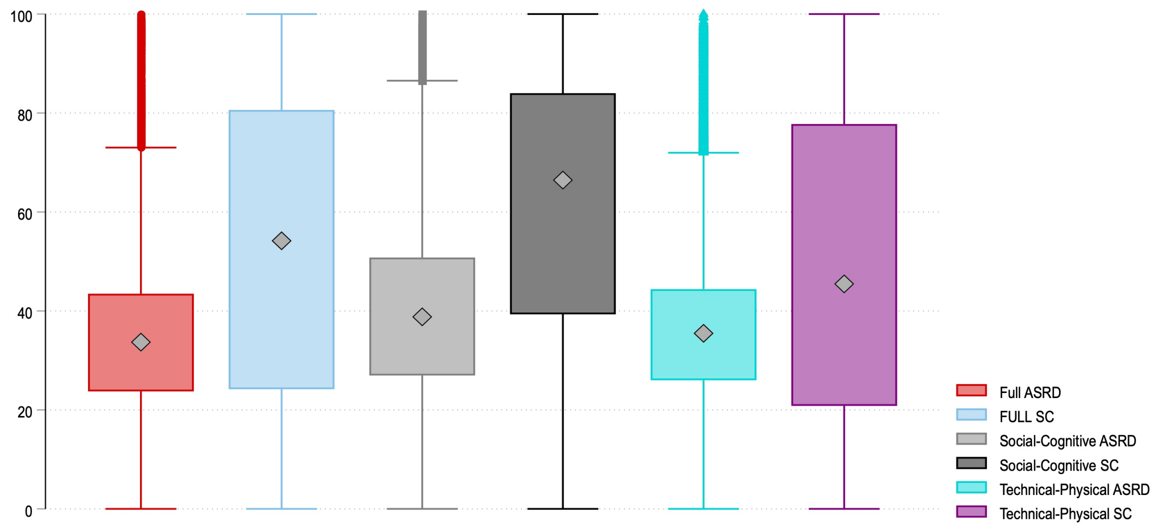


Figure 8. Distributions of Independent Variables

We include three-way fixed effects to all estimates, namely, fixed effects for industries ι_i, regions ρ_p, and time γ_t, assuming that there might be some time-independent features that correlate with the variables. In addition, we include three sets of control variables. The first set stands for regional economic and demographic differences. Population density (log), the number of inhabitants per square kilometre per region, is included to control for urbanisation. GDP per capita (log) controls for the level of economic development. Education is the share of tertiary education in the population and enters the specification as a standard control of human capital. The second set of control variables is at the industrial-regional level. Business growth (percentage) is the net population growth of industries. Churn (percentage) is the sum of growth and death rates of industries. This set controls for the differences in turnovers of industries. Some industries may be inherently more (less) mobile, which might increase (decrease) their entry or exit probability. Lastly, we add two more variables named geographical diversification controls. Industrial ubiquity (number of regions that a particular industry already specialised in, *i.e.,* $RCA > 1$) controls for the rarity of industries since rare industries are expected to form relatively fewer specialisations. Regional diversity (number of specialised industries, *i.e.,* $RCA > 1$, in a given region) accounts for the diversity of industry mixes of regions, given that more diversified regions might be more prone to attract new specialisations. Geographical diversification controls allow us to account for the geographical distribution patterns of industries and regions, helping to isolate the effect of skill-related co-occurrence patterns.

Table 1. Entry and Exit Models: Full Skills

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
Full ASRD	0.00207***		0.00181***	-0.00292***		-0.00237***
	(0.00016)		(0.00016)	(0.00036)		(0.00036)
Full Complexity		0.00035***	0.00030***		-0.00057***	-0.00051***
		(0.00003)	(0.00003)		(0.00005)	(0.00005)
GDP (log)	-0.06612**	-0.04227	-0.06117*	0.07519	0.03401	0.06221
	(0.03261)	(0.03258)	(0.03260)	(0.07535)	(0.07517)	(0.07536)
Pop. Density	-0.07006	-0.01994	-0.06715	0.08037	0.00111	0.05532
	(0.06147)	(0.06132)	(0.06148)	(0.13001)	(0.12967)	(0.13012)
Education	-0.00159*	-0.00113	-0.00166*	-0.00300	-0.00331*	-0.00285
	(0.00088)	(0.00088)	(0.00088)	(0.00197)	(0.00196)	(0.00196)
Churn	-0.00076*	-0.00095**	-0.00074*	0.00137	0.00159*	0.00122
	(0.00040)	(0.00041)	(0.00040)	(0.00095)	(0.00095)	(0.00095)
Bus. Growth	0.00105***	0.00119***	0.00107***	-0.00152**	-0.00181**	-0.00150**
	(0.00029)	(0.00029)	(0.00029)	(0.00074)	(0.00074)	(0.00074)
Reg. Diversity	-0.00118***	-0.00109***	-0.00117***	0.00293***	0.00281***	0.00289***
	(0.00014)	(0.00014)	(0.00014)	(0.00032)	(0.00032)	(0.00032)
Ind. Ubiquity	-0.00561***	-0.00564***	-0.00561***	0.00932***	0.00927***	0.00924***
	(0.00037)	(0.00037)	(0.00037)	(0.00067)	(0.00067)	(0.00067)
Full FE	yes	yes	yes	yes	yes	yes
N	157,602	157,602	157,602	65,814	65,814	65,814
Adj.R^2	0.051	0.051	0.052	0.077	0.078	0.079

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

An examination of the correlation matrix reported in Table A5 reveals low correlations, below 0.49, between independent variables. The only exception is GDP per capita, which is relatively highly correlated, around 0.66, with ASRD variables. However, we included it in the models since it is a standard control variable and is almost always insignificant. Variance inflation factors (VIF) are computed after each baseline specification, as displayed in Table A6. All VIFs are in the acceptable range, *i.e.,* below 5, including GDP per capita.

Since errors are likely to be correlated within regions and industries, all estimates are presented with robust standard errors clustered at the industry-region level. To mitigate potential endogeneity, all independent variables are lagged by one period, denoted as $t-1$ where t is the first year of each three-year period.

We first estimate the baseline specification for the full skill set, meaning that ASRD and skill complexity variables are computed with the full set of 161 workplace skills. Table 1 displays the results for both entry and exit models. In entry model 1, *Full ASRD* has a positive and highly significant effect on the probability that an industry specialises in a new region, as expected. In entry model 2, skill complexity *Full Complexity* also has a positive and significant effect on the probability of new industrial specialisation. In entry model 3, ASRD and skill complexity are present together, and significant effects prevail. The dependent variables are binary; therefore, coefficients must be interpreted as probabilities. Regarding entry, when ASRD increases by ten percentage points, the probability of a new industry specialisation in the region increases by about 1.8%. Skill complexity shows relatively little relevance, a ten per cent increase is associated with a 0.3% increase in the probability of a new industry entry.

Table 2. Entry and Exit Models: Social-Cognitive Skill Cluster

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
SC ASRD	0.00234***		0.00220***	-0.00288***		-0.00251***
	(0.00017)		(0.00017)	(0.00032)		(0.00032)
SC Complexity		0.00043***	0.00041***		-0.00084***	-0.00081***
		(0.00003)	(0.00003)		(0.00005)	(0.00005)
GDP (log)	-0.07196**	-0.04730	-0.07252**	0.08062	0.04377	0.07790
	(0.03262)	(0.03258)	(0.03261)	(0.07542)	(0.07514)	(0.07539)
Pop. Density	-0.08564	-0.01682	-0.08283	0.10038	0.01007	0.08438
	(0.06159)	(0.06136)	(0.06163)	(0.13012)	(0.12928)	(0.12985)
Education	-0.00192**	-0.00107	-0.00198**	-0.00287	-0.00333*	-0.00270
	(0.00089)	(0.00088)	(0.00088)	(0.00197)	(0.00196)	(0.00196)
Churn	-0.00073*	-0.00095**	-0.00070*	0.00150	0.00161*	0.00131
	(0.00040)	(0.00041)	(0.00040)	(0.00095)	(0.00095)	(0.00095)
Bus. Growth	0.00104***	0.00115***	0.00101***	-0.00150**	-0.00161**	-0.00125*
	(0.00029)	(0.00029)	(0.00029)	(0.00074)	(0.00074)	(0.00074)
Reg. Diversity	-0.00121***	-0.00110***	-0.00121***	0.00294***	0.00282***	0.00291***
	(0.00014)	(0.00014)	(0.00014)	(0.00032)	(0.00032)	(0.00032)
Ind. Ubiquity	-0.00560***	-0.00562***	-0.00558***	0.00934***	0.00916***	0.00915***
	(0.00037)	(0.00037)	(0.00037)	(0.00067)	(0.00067)	(0.00067)
Full FE	yes	yes	yes	yes	yes	yes
N	157,602	157,602	157,602	65,814	65,814	65,814
Adj. R^2	0.051	0.052	0.053	0.077	0.081	0.082

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Exit models in Table 1 indicate that ASRD and skill complexity are negatively and significantly affect the probability that an industry loses comparative advantage in a particular region. When ASRD increases by ten percentage points, the probability of an industry losing comparative advantage in the region decreases by about 2.3%. Likewise, when skill complexity increases by ten percentage points, the probability of exit decreases by about 0.5%. In general, the effects of ASRD and skill complexity are more potent for exit models.

In the following step, we turn to the detected skill communities in Section 4 and separately compute the variables ASRD and skill complexity for the two polarised skill clusters: social-cognitive skills and technical-physical skills. We then run the baseline specification with new independent variables. Table 2 shows the results for social-cognitive skills. Regarding entry models, a ten per cent increase in social-cognitive ASRD (*SC ASRD*) positively affects the probability of a new industry specialisation by 2.2%. Negative association endures for exit models; when social-cognitive

ASRD increases by ten percentage points, the probability of exit decreases by 2.5%. Social-cognitive skill complexity *SC Complexity* exhibits a similar look to full skill complexity (*Full Complexity*), the coefficient almost doubles in exit models.

Table 3. Entry and Exit Models: Technical-Physical Skill Cluster

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
TP ASRD	0.00231***		0.00218***	-0.00274***		-0.00252***
	(0.00017)		(0.00017)	(0.00037)		(0.00037)
TP Complexity		0.00038***	0.00035***		-0.00070***	-0.00067***
		(0.00003)	(0.00003)		(0.00006)	(0.00006)
GDP (log)	-0.06515**	-0.04513	-0.06426**	0.06728	0.03891	0.06285
	(0.03260)	(0.03257)	(0.03258)	(0.07526)	(0.07517)	(0.07527)
Pop. Density	-0.05940	-0.01976	-0.06128	0.05413	0.01454	0.05025
	(0.06140)	(0.06135)	(0.06143)	(0.12971)	(0.12953)	(0.12968)
Education	-0.00151*	-0.00101	-0.00154*	-0.00329	-0.00348*	-0.00319
	(0.00088)	(0.00088)	(0.00088)	(0.00197)	(0.00196)	(0.00196)
Churn	-0.00058	-0.00097**	-0.00057	0.00126	0.00173*	0.00118
	(0.00040)	(0.00041)	(0.00040)	(0.00095)	(0.00095)	(0.00095)
Bus. Growth	0.00102***	0.00119***	0.00102***	-0.00157**	-0.00188**	-0.00156**
	(0.00029)	(0.00029)	(0.00029)	(0.00074)	(0.00074)	(0.00074)
Reg. Diversity	-0.00118***	-0.00110***	-0.00118***	0.00294***	0.00281***	0.00290***
	(0.00014)	(0.00014)	(0.00014)	(0.00032)	(0.00032)	(0.00032)
Ind. Ubiquity	-0.00562***	-0.00562***	-0.00559***	0.00934***	0.00924***	0.00923***
	(0.00037)	(0.00037)	(0.00037)	(0.00067)	(0.00067)	(0.00067)
Full FE	yes	yes	yes	yes	yes	yes
N	157,602	157,602	157,602	65,814	65,814	65,814
Adj.R^2	0.051	0.051	0.052	0.077	0.078	0.079

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

The results for technical-physical skills, presented in Table 3, are economically and statistically similar to those of full skills and social-cognitive skills. Overall, similarity to technical-physical skills and complexity in social-cognitive skills pose the most substantial impact on the industrial specialisation process, even though the size effects are pretty similar.

The overall picture changes when considering manufacturing (columns 1-6, Table 4) and service (columns 7-12, Table 4) industries separately. The results show that ASRD and skill complexity affect the specialisation process of manufacturing industries with pretty similar coefficients to the overall results. Regarding service industries, ASRD of social-cognitive skills is insignificant, suggesting that similarity to those skills does not enhance or mitigate the specialisation process. On the other hand, ASRD of technical-physical skills is significant at the 99% level with relatively lower coefficients than the overall results. The skill complexities of both skill types are relevant and significant.

Table 4. Entry and Exit Models: Manufacturing and Service Industries

	Manufacturing						Service					
	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit	(7) Entry	(8) Entry	(9) Entry	(10) Exit	(11) Exit	(12) Exit
Full ASRD	0.00220***			-0.00212***			0.00040*			-0.00117**		
	(0.00019)			(0.00036)			(0.00023)			(0.00050)		
Full Complexity	0.00041***			-0.00060***			0.00020***			-0.00051***		
	(0.00004)			(0.00005)			(0.00005)			(0.00012)		
SC ASRD		0.00196***			-0.00213***			0.00012			-0.00072	
		(0.00018)			(0.00031)			(0.00022)			(0.00051)	
SC Complexity		0.00048***			-0.00077***			0.00023***			-0.00059***	
		(0.00003)			(0.00006)			(0.00005)			(0.00011)	
TP ASRD			0.00211***			-0.00159***			0.00071***			-0.00140***
			(0.00016)			(0.00034)			(0.00024)			(0.00046)
TP Complexity			0.00045***			-0.00063***			0.00021***			-0.00051***
			(0.00004)			(0.00006)			(0.00005)			(0.00013)
GDP (log)	-0.03444	-0.03128	-0.03942	0.03595	0.03529	0.04114	-0.06872	-0.06452	-0.07072	0.08131	0.06692	0.08432
	(0.03725)	(0.03723)	(0.03722)	(0.08652)	(0.08646)	(0.08646)	(0.06407)	(0.06408)	(0.06409)	(0.14723)	(0.14710)	(0.14714)
Pop. Density	0.02474	0.03260	0.03286	0.16489	0.19492	0.13185	-0.18969*	-0.19140*	-0.19104*	-0.28059	-0.27046	-0.26816
	(0.07304)	(0.07307)	(0.07312)	(0.14710)	(0.14690)	(0.14723)	(0.11199)	(0.11203)	(0.11201)	(0.26465)	(0.26480)	(0.26447)
Education	-0.00347***	-0.00338***	-0.00340***	-0.00273	-0.00289	-0.00283	0.00180	0.00188	0.00182	-0.00187	-0.00159	-0.00223
	(0.00103)	(0.00103)	(0.00103)	(0.00225)	(0.00225)	(0.00225)	(0.00169)	(0.00169)	(0.00169)	(0.00388)	(0.00388)	(0.00388)
Churn	-0.00112	-0.00115	-0.00105	0.00213	0.00215	0.00230*	-0.00005	-0.00005	0.00000	0.00282*	0.00287**	0.00272*
	(0.00072)	(0.00072)	(0.00072)	(0.00137)	(0.00136)	(0.00136)	(0.00054)	(0.00054)	(0.00054)	(0.00146)	(0.00146)	(0.00147)
Bus. Growth	0.00156***	0.00140***	0.00167***	-0.00147	-0.00103	-0.00184*	0.00051	0.00050	0.00050	-0.00183*	-0.00175	-0.00182*
	(0.00052)	(0.00052)	(0.00052)	(0.00104)	(0.00104)	(0.00104)	(0.00036)	(0.00036)	(0.00036)	(0.00108)	(0.00108)	(0.00108)
Reg. Diversity	-0.00107***	-0.00108***	-0.00107***	0.00252***	0.00250***	0.00252***	-0.00119***	-0.00119***	-0.00118***	0.00346***	0.00346***	0.00346***
	(0.00016)	(0.00016)	(0.00016)	(0.00036)	(0.00036)	(0.00036)	(0.00027)	(0.00027)	(0.00027)	(0.00062)	(0.00062)	(0.00062)
Ind. Ubiquity	-0.00624***	-0.00618***	-0.00622***	0.00859***	0.00855***	0.00866***	-0.00456***	-0.00455***	-0.00456***	0.01035***	0.01039***	0.01036***
	(0.00049)	(0.00049)	(0.00049)	(0.00080)	(0.00080)	(0.00080)	(0.00055)	(0.00055)	(0.00055)	(0.00122)	(0.00122)	(0.00122)
Full FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	106,662	106,662	106,662	46,562	46,562	46,562	50,940	50,940	50,940	19,252	19,252	19,252
$Adj.R^2$	0.058	0.059	0.058	0.060	0.063	0.060	0.048	0.048	0.048	0.112	0.112	0.112

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

6 Robustness Checks

6.1 Average Marginal Effects: Linear Probability, Logistic, and Probit Models

LPM has been extensively used in the applied microeconometrics and economic geography literature, partly due to its usefulness since the estimated coefficients can be interpreted on the probability scale. In a logistic model, the estimated coefficients are changes in the log-odds scale that is challenging to interpret. In a probit model, the estimated parameters are interpreted as shifts in the cumulative normal distribution, which is also intuitively challenging. Nevertheless, LPM is criticized for having a couple of pitfalls, including inherent heteroscedasticity, inappropriate significance tests due to non-normality of errors, and estimates out of the probability range $(0-1)$ (Greene, 2012). The heteroscedasticity problem can be resolved by using robust options of statistical software. Hallevik (2009) shows that the significance probabilities from linear and logistic regressions are identical; therefore, the argument of inappropriate significance tests does not necessarily hold. On the other hand, estimations below zero and above one are possible. Therefore, we estimate the baseline specifications with logistic and probit models to mitigate potential concerns for LPM. Table A7 displays the results. The first six columns are entry models where columns 1-3 are estimated with logistic regression and columns 4-6 are estimated with probit regression. The second half of Table A7 displays exit models. Columns 7-9 are estimated with logistic, while columns 10-12 are estimated with probit models. The signs of the coefficients and the qualitative results of the estimates are in line with the LPM estimates reported in Tables 1, 2, and 3.

Table 5. Average Marginal Effects for LPM, Logistic, and Probit Estimates

	Entry			Exit		
	(1) LPM	(2) Logistic	(3) Probit	(4) LPM	(5) Logistic	(6) Probit
Full ASRD	0.00180	0.00182	0.00189	-0.00237	-0.00246	-0.00247
	(0.00016)	(0.00016)	(0.00016)	(0.00036)	(0.00035)	(0.00034)
SC ASRD	0.00219	0.00187	0.00195	-0.00251	-0.00272	-0.00272
	(0.00016)	(0.00014)	(0.00014)	(0.00032)	(0.00032)	(0.00031)
TP ASRD	0.00218	0.00216	0.00225	-0.00252	-0.00252	-0.00255
	(0.00017)	(0.00017)	(0.00016)	(0.00036)	(0.00035)	(0.00035)
Full Complexity	0.00030	0.00026	0.00028	-0.00051	-0.00054	-0.00054
	(0.00003)	(0.00002)	(0.00002)	(0.00005)	(0.00005)	(0.00005)
SC Complexity	0.00041	0.00038	0.00039	-0.00081	-0.00083	-0.00084
	(0.00003)	(0.00002)	(0.00002)	(0.00005)	(0.00005)	(0.00005)
TP Complexity	0.00035	0.00032	0.00034	-0.00066	-0.00064	-0.00065
	(0.00003)	(0.00003)	(0.00003)	(0.00006)	(0.00006)	(0.00006)

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

When it comes to size effects, as already mentioned, the estimated parameters of linear probability, logistic, and probit models should not be directly compared since they are at different scales. Hence, we calculate average marginal effects (AME) to be able to compare the estimated effects of these models. Table 5 exhibits AME computed after the baseline specifications reported in the previous tables. Columns 1 and 4 for LPM correspond to Table 1 (column 3 for entry, column

6 for exit) for full range of skills (*Full ASRD*, *Full Complexity*), Table 2 for social-cognitive skill cluster (*SC ASRD*, *SC Complexity*), and Table 3 for technical-physical skill cluster (*TP ASRD*, *TP Complexity*). Columns 2 and 3 of Table 5 correspond to the first six columns of Table A7, while columns 5 and 6 correspond to the last six columns of Table A7. AME estimated with logistic and probit models suggest that a ten per cent increase in *Full ASRD* is associated with, respectively, a 1.82% and 1.89% increase in the entry probability, which is almost the same size effect estimated with LPM. AME of other variables estimated with different models are also pretty similar, suggesting that the estimated probabilities in previous specifications are not severely affected by model selection.

6.2 Different Specifications for Dependent Variables

The dependent variables, entry and exit, account for changes in comparative advantages of industries that are based on the number of employed people, as denoted in equations 8, 9, and 10. Nevertheless, industrial specialisations might also be defined by other types of comparative advantages. We thus aim to assess if the main findings are robust to a different specification of comparative advantage.

Furthermore, Boschma (2017) underline that the effect of relatedness can be overestimated when using the same data set for both dependent and independent variables. Even though we use different data sets (see data section) for dependent and independent variables, both are based on the same survey, ILFS. Motivated by this reasoning, we use data from another survey[14] to construct alternative dependent variables with a different comparative advantage criterion. We change the computation of dependent variables by changing the RCA formula as follows.

$$RCA^1_{i,p,t} = \frac{ENT_{i,p,t}/ENT_{i',p,t}}{ENT_{i,p',t}/ENT_{i',p',t}} \tag{11}$$

where $ENT_{i,p,t}$ is the number of enterprises that are active in industry i, located in region p in period t. $RCA > 1$ indicates that the number of active enterprises in industry i, region p, and period t is greater than the average national level. Afterwards, we define entry and exit variables as in equations 8 and 9 to run the baseline econometric model with new dependent variables. Table 6 displays the results. Technical-physical ASRD has the highest impact, for both entry and exit models, compared to social-cognitive and full skills, even though the differences are negligible. When there is a ten per cent increase in the ASRD of technical-physical skills, the probability of entry (exit) increases (decreases) by 2.7% (3.2%). As in the previous models, social-cognitive skill complexity yields the highest impact. A ten per cent increase in social-cognitive skill complexity is associated with a 0.46% (0.81%) increase (decrease) in the probability of entry (exit).

Overall, the results indicate that the main findings are robust to a different comparative advantage criterion. Moreover, the relatedness effect is not overestimated in the baseline specification; indeed, it is underestimated given the increased coefficients in Table 6 compared to Tables 1, 2, and 3.

As another robustness check, we analyse if the main results are sensitive to a change in the time span of RCA. Recall that we construct variables *Entry* and *Exit* based on the changes in RCA (equation 10) from time $t-3$ to time t, stands for three years. Here, we decrease the time span to

[14] Statistical register of active enterprises (ASIA - Enterprises)

one year, from time $t-1$ to time t, and redefine the dependent variables with the two criteria of comparative advantage, RCA and RCA^1, as follows.

$$Entry_{i,p,t} = 1,\ if\ RCA_{i,p,t} > 1\ and\ RCA_{i,p,t-1} \leq 1 \tag{12}$$
$$Exit_{i,p,t} = 1,\ if\ RCA_{i,p,t} \leq 1\ and\ RCA_{i,p,t-1} > 1 \tag{13}$$

$$Entry_{i,p,t} = 1,\ if\ RCA^1_{i,p,t} > 1\ and\ RCA^1_{i,p,t-1} \leq 1 \tag{14}$$
$$Exit_{i,p,t} = 1,\ if\ RCA^1_{i,p,t} \leq 1\ and\ RCA^1_{i,p,t-1} > 1 \tag{15}$$

Table 6. Entry and Exit Models: RCA in Number of Enterprises

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
Full ASRD	0.00216***			-0.00278***		
	(0.00020)			(0.00031)		
Full Complexity	0.00039***			-0.00058***		
	(0.00003)			(0.00005)		
SC ASRD		0.00246***			-0.00281***	
		(0.00020)			(0.00029)	
SC Complexity		0.00046***			-0.00081***	
		(0.00003)			(0.00005)	
TP ASRD			0.00268***			-0.00322***
			(0.00020)			(0.00032)
TP Complexity			0.00043***			-0.00072***
			(0.00004)			(0.00006)
GDP (log)	-0.00851	-0.01869	-0.01271	-0.00043	0.01619	0.00262
	(0.03949)	(0.03948)	(0.03948)	(0.06750)	(0.06748)	(0.06749)
Pop. Density	-0.06265	-0.07509	-0.05832	0.09983	0.12441	0.09435
	(0.07264)	(0.07266)	(0.07260)	(0.12240)	(0.12256)	(0.12225)
Education	-0.00081	-0.00103	-0.00063	0.00124	0.00154	0.00096
	(0.00104)	(0.00104)	(0.00104)	(0.00188)	(0.00188)	(0.00188)
Churn	-0.00100**	-0.00100**	-0.00083*	0.00281***	0.00289***	0.00267***
	(0.00047)	(0.00047)	(0.00047)	(0.00087)	(0.00086)	(0.00087)
Bus. Growth	0.00126***	0.00119***	0.00120***	-0.00288***	-0.00273***	-0.00289***
	(0.00033)	(0.00033)	(0.00033)	(0.00068)	(0.00068)	(0.00068)
Reg. Diversity	-0.00132***	-0.00138***	-0.00131***	0.00213***	0.00220***	0.00213***
	(0.00016)	(0.00016)	(0.00016)	(0.00027)	(0.00027)	(0.00027)
Ind. Ubiquity	-0.00618***	-0.00615***	-0.00618***	0.00883***	0.00875***	0.00881***
	(0.00038)	(0.00038)	(0.00038)	(0.00063)	(0.00063)	(0.00063)
Full FE	yes	yes	yes	yes	yes	yes
N	144,087	144,087	144,087	79,329	79,329	79,329
Adj.R^2	0.050	0.051	0.050	0.073	0.075	0.074

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 7. Entry and Exit Models: Decrease of Time Span

	Dependent variable specification: number of employees						Dependent variable specification: number of enterprises					
	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit	(7) Entry	(8) Entry	(9) Entry	(10) Exit	(11) Exit	(12) Exit
Full ASRD	0.00104***			-0.00130***			0.00130***			-0.00164***		
	(0.00009)			(0.00020)			(0.00011)			(0.00019)		
Full Complexity	0.00021***			-0.00039***			0.00026***			-0.00040***		
	(0.00002)			(0.00003)			(0.00002)			(0.00003)		
SC ASRD		0.00117***			-0.00137***			0.00148***			-0.00154***	
		(0.00009)			(0.00019)			(0.00011)			(0.00018)	
SC Complexity		0.00024***			-0.00052***			0.00034***			-0.00051***	
		(0.00002)			(0.00003)			(0.00002)			(0.00003)	
TP ASRD			0.00130***			-0.00144***			0.00166***			-0.00191***
			(0.00009)			(0.00021)			(0.00012)			(0.00019)
TP Complexity			0.00024***			-0.00046***			0.00033***			-0.00047***
			(0.00002)			(0.00004)			(0.00002)			(0.00004)
GDP (log)	-0.02936	-0.02996	-0.03006	0.00406	0.00718	0.00674	-0.00838	-0.00789	-0.00893	-0.00282	0.00008	0.00016
	(0.02217)	(0.02217)	(0.02217)	(0.05472)	(0.05473)	(0.05473)	(0.02760)	(0.02762)	(0.02760)	(0.04968)	(0.04973)	(0.04971)
Pop. Density	0.00432	0.00520	0.00475	-0.00672	-0.00750	-0.00603	0.01025**	0.01148**	0.01095**	-0.01167	-0.01190	-0.01127
	(0.00402)	(0.00402)	(0.00401)	(0.00987)	(0.00986)	(0.00987)	(0.00476)	(0.00476)	(0.00476)	(0.01011)	(0.01012)	(0.01012)
Education	-0.00168***	-0.00179***	-0.00166***	-0.00232**	-0.00215*	-0.00238**	0.00091	0.00077	0.00093	0.00093	0.00107	0.00087
	(0.00048)	(0.00048)	(0.00048)	(0.00112)	(0.00112)	(0.00112)	(0.00059)	(0.00059)	(0.00059)	(0.00108)	(0.00108)	(0.00108)
Churn	-0.00048**	-0.00048**	-0.00038*	0.00151***	0.00163***	0.00153***	-0.00076***	-0.00077***	-0.00067**	0.00173***	0.00185***	0.00166***
	(0.00022)	(0.00022)	(0.00022)	(0.00057)	(0.00057)	(0.00057)	(0.00027)	(0.00027)	(0.00027)	(0.00054)	(0.00054)	(0.00054)
Bus. Growth	0.00028	0.00027	0.00024	-0.00051	-0.00049	-0.00054	0.00014	0.00013	0.00010	-0.00067	-0.00071	-0.00068
	(0.00018)	(0.00018)	(0.00018)	(0.00049)	(0.00049)	(0.00049)	(0.00022)	(0.00022)	(0.00022)	(0.00046)	(0.00046)	(0.00046)
Reg. Diversity	-0.00064***	-0.00065***	-0.00064***	0.00199***	0.00199***	0.00198***	-0.00080***	-0.00084***	-0.00080***	0.00172***	0.00175***	0.00173***
	(0.00009)	(0.00009)	(0.00009)	(0.00021)	(0.00021)	(0.00021)	(0.00010)	(0.00010)	(0.00010)	(0.00017)	(0.00017)	(0.00017)
Ind. Ubiquity	-0.00364***	-0.00362***	-0.00363***	0.00638***	0.00636***	0.00639***	-0.00361***	-0.00359***	-0.00360***	0.00720***	0.00717***	0.00720***
	(0.00024)	(0.00024)	(0.00024)	(0.00042)	(0.00042)	(0.00043)	(0.00025)	(0.00025)	(0.00025)	(0.00041)	(0.00041)	(0.00041)
Full FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	234,494	234,494	234,494	98,020	98,020	98,020	214,232	214,232	214,232	118,282	118,282	118,282
Adj.R^2	0.030	0.031	0.031	0.047	0.048	0.046	0.032	0.033	0.032	0.045	0.046	0.045

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 7 shows the results. The first panel presents results for the dependent variables constructed with RCA as denoted in equations 12 and 13 (*i.e.,* comparative advantage in terms of the number of employed people). The second panel exhibits results for dependent variables constructed with RCA^1 as denoted in equations 14 and 15 (*i.e.,* comparative advantage in terms of the number of enterprises). The results confirm the main findings in previous models; ASRD and skill complexity variables are significantly and positively related to entry while significantly and negatively related to exit. The coefficients of interest exhibit some degree of decrease, regardless of the specification. For the dependent variables constructed with RCA, a ten per cent increase in the ASRD for, respectively, full, social-cognitive and technical-physical skills increases the probability of developing a new specialisation by 1%, 1.2%, 1.3%; whilst decreases the probability of losing comparative advantage by 1.3%, 1.4%, 1.4%. For the dependent variables constructed with RCA^1, the coefficients are slightly higher. A ten per cent increase in the ASRD for, respectively, full, social-cognitive and technical-physical skills increases the probability of entry by 1.3%, 1.5%, 1.7%; whilst decreases the probability of losing comparative advantage by 1.6%, 1.5%, 1.9%.

6.3 Alternative Model Specification: Industry−Time and Region−Time Fixed Effects

Here, we present an alternative specification of the fixed effects that are already added to the previous estimates. We intend to eliminate the effects of time-varying heterogeneity across regions or industries. The aim is to test whether the effects of ASRD and skill complexity variables are driven by some unobservables that affect both dependent variables. We thus include industry-time

Table 8. Entry and Exit Models: Industry-Time, Region-Time Fixed Effects

	(1) Entry	(2) Entry	(3) Entry	(4) Exit	(5) Exit	(6) Exit
Full ASRD	0.00264***			-0.00237***		
	(0.00022)			(0.00036)		
Full Complexity	0.00029***			-0.00052***		
	(0.00003)			(0.00005)		
SC ASRD		0.00385***			-0.00410***	
		(0.00025)			(0.00047)	
SC Complexity		0.00041***			-0.00083***	
		(0.00003)			(0.00005)	
TP ASRD			0.00300***			-0.00344***
			(0.00021)			(0.00047)
TP Complexity			0.00035***			-0.00069***
			(0.00003)			(0.00006)
Industry-*Time FE*	yes	yes	yes	yes	yes	yes
Region-*Time FE*	yes	yes	yes	yes	yes	yes
N	157,602	157,602	157,602	65,814	65,812	65,812
Adj.R^2	0.049	0.051	0.050	0.075	0.077	0.073

Notes: Robust standard errors clustered at the region and industry level are in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

and region-time fixed effects instead of their separate inclusion. The control variables are not included in the models, given that most of them become invariant in the present specification. According to the results in Table 8, ASRD and skill complexity indicators for all skill types exhibit positive (negative) and significant effects with increased coefficients on acquiring (losing) new industrial specialisations.

7 Concluding Remarks

Region-specific capabilities are increasingly accepted as a critical source of regional diversification, yet they are seldom included in the measurement process. In this paper, we include regional capabilities into the measurement process of relatedness and complexity by using industry-skill input matrices per region and year, enabling fine-grained analyses of skill relatedness and complexity in the industrial diversification process of regions.

Building on the network-based approach of EEG, we first construct a skill-to-skill relatedness matrix and skill complexity vector, *i.e.*, skill space, of each region in each year. We then combine these measures at the national level and formalise them as a one-mode network, *i.e.*, the skill space of Italy, that consists of 161 workplace skills connected with edges that represent their degree of relatedness. Each node is proportional to its complexity score. After applying a force-based algorithm, the skill space forms two highly polarised skill clusters. The first one consists of basic, social, cognitive, management, and interacting skills, therefore named social-cognitive skills; whilst another one consists of physical, psychomotor, sensory, systems, and technical skills, thus named technical-physical skills.

We then employ linear probability models with three-way fixed effects to quantify the effect of skill relatedness and skill complexity on the probability of (1) forming a new regional comparative advantage in terms of the number of employed people in a given industry (entry), or (2) losing an established comparative advantage (exit). The independent variables are constructed for three different aggregation levels: (1) full skills, (2) social-cognitive skills, and (3) technical-physical skills. The results show a highly significant and positive relationship between higher skill relatedness density and the probability of entry, regardless of the skill type. The more similar the skill set of the new industry to the skill sets of industries that are already established a comparative advantage in that region, the higher the probability of building a new comparative advantage for the new industry in that region. Even though the correlation coefficients are pretty close among the skill types, similarity to technical-physical skills yields a slightly higher probability of entry. When it comes to exit, all estimates show a highly significant and negative association between skill relatedness density and the probability of losing an established comparative advantage, regardless of the skill type. The skill complexity of an industry positively (negatively) affects its entry (exit) probability. We also show that skill relatedness density and skill complexity affect the entry and exit probability of manufacturing industries, regardless of the skill type. On the other hand, only technical-physical skill relatedness density enhances (decreases) the entry (exit) probability of service industries, while skill complexity affects both entry and exit probabilities. By controlling for industrial ubiquity and regional diversity, we show that our results are not just a reflection of the geographical distribution patterns of industries.

The findings mentioned above are robust to a couple of sensitivity analyses. First, we re-estimate the baseline specifications with logistic and probit models to assess the resulting average marginal effects. Second, we re-identified the dependent variables, entry and exit, with a different data set and comparative advantage criterion. Instead of RCA in the number of employed people, we used RCA in the number of active enterprises. The main findings endured with very similar coefficients. Third, we decreased the time span of the comparative advantage to be one year and estimated the baseline specification with both RCA criteria. The results were identical, with a slight decrease in the coefficients. Lastly, we employed region-time and industry-time fixed effects. The results did not change.

Although the present work is far from revealing all aspects of skill relatedness and complexity, the findings clearly underline the role of these measures in the industrial diversification of regions yet also call for further investigation. First, the analysis covers only one country due to data availability. Apparently, further research is needed to understand how skill relatedness and complexity affect the industrial diversification process of regions in a cross-country setting. Second, this study has analysed skill relatedness and complexity in isolation. However, the literature has demonstrated that other relatedness types, such as technological relatedness, are highly relevant in the diversification process. Therefore, further exploration of different relatedness and complexity types would unravel valuable insights regarding underlying mechanisms that cause regions to diversify into different industries.

Regional skill relatedness and complexity analyses might yield rewarding insights for policymakers. They provide a recent panorama of regional human capital obtained with highly granular data and sophisticated methods, unlike previous measures such as completed years of education and work experience. Accordingly, these measures are highly beneficial for human capital formation because skill shortages can be analysed at the industry-region level to define necessary measures that local and national authorities can take. By doing so, regional innovation and smart specialisation policies can be more strategically designed to stimulate new industrial specialisations.

Appendix

Table A1. ICP Categories

1.Knowledge	(B1) Administration and Management, (B2) Office Work, (B3) Economics and Accounting, (B4) Sales and Marketing, (B5) Services to Customers, (B6) Human Resources Management, (B7) Production and Processing, (B8) Food Production, (B9)IT and Electronics, (B10) Engineering and Technology, (B11) Technical Design, (B12) Building and Construction, (B13)Mechanical, (B14) Mathematics, (B15) Physics, (B16) Chemistry, (B17) Biology, (B18) Psychology, (B19) Sociology and Anthropology, (B20) Geography, (B21) Medicine and Dentistry, (B22) Therapy and Counseling, (B23) Education and Training, (B24) Italian Language, (B25) Foreign Language, (B26) Fine Arts, (B27)History and Archaeology, (B28) Philosophy and Theology, (B29) Civil Protection and Public Safety, (B30) Legislation and Institutions, (B31) Telecommunications, (B32) Communication and Media, (B33) Transportation
2.Skills	
2.1 Basic Skills	(C1) Reading Comprehension, (C2) Active Listening, (C3) Writing, (C4) Speaking, (C5) Mathematics, (C6) Science, (C7) Critical Thinking, (C8) Active Learning, (C9) Learning Strategies, (C10) Monitoring
2.2 Social Skills	(C11) Social Perceptiveness, (C12) Coordination, (C13) Persuasion, (C14) Negotiation, (C15) Instructing, (C16) Service Orientation
2.3 Complex Problem	(C17) Complex Problem Solving
2.4 Technical Skills	(C18) Operations Analysis, (C19) Technology Design, (C20) Equipment Selection, (C21) Installation, (C22) Programming, (C23) Quality Control Analysis, (C24) Operation Monitoring, (C25) Operation and Control, (C26) Equipment Maintenance, (C27) Troubleshooting, (C28) Repairing
2.5 Systems Skills	(C29) Systems Analysis, (C30) Systems Evaluation, (C31) Judgment and Decision Making
2.6 Resource Management Skills	(C32) Time Management, (C33) Management of Financial Resources, (C34) Management of Material Resources, (C35) Management of Personnel Resources
3.Attitudes	
3.1 Cognitive	(D1) Oral Comprehension, (D2) Written Comprehension, (D3) Oral Expression, (D4) Written Expression, (D5) Fluency of Ideas, (D6) Originality, (D7) Problem Sensitivity, (D8) Deductive Reasoning, (D9) Inductive Reasoning, (D10) Information Ordering, (D11) Category Flexibility, (D12) Math Reasoning, (D13) Number Facility, (D14) Memorisation, (D15) Speed of Closure, (D16) Flexibility of Closure, (D17) Perceptual Speed, (D18) Spatial Orientation, (D19) Visualisation, (D20)Selective Attention, (D21) Time Sharing
3.2 Psychomotor	(D22) Arms-Hand Steadiness, (D23) Manual Dexterity, (D24) Finger Dexterity, (D25) Control Precision, (D26) Multilimb Coordination, (D27) Response Orientation, (D28) Rate Control, (D29) Reaction Time, (D30) Wrist-Finger Speed, (D31) Speed of Limb Movement
3.3 Psychical	(D32) Static Strength, (D33) Explosive Strength, (D34) Dynamic Strength, (D35) Trunk Strength, (D36) Stamina, (D37) Extent Flexibility, (D38) Dynamic Flexibility, (D39) Gross Body Coordination, (D40) Gross Balance Body Equilibrium
3.4 Sensory	(D41) Near Vision, (D42) Far Vision, (D43) Visual Colour Discrimination, (D44) Night Vision, (D45) Peripheral Vision, (D46) Depth Perception, (D47) Glare Sensitivity, (D48) Hearing Sensitivity, (D49) Auditory Attention, (D50) Sound Localisation, (D51) Speech Recognition, (D52) Speech Clarity
4.Work Activities	
4.1 Information Input	(G1) Getting Information, (G2) Identifying Objects, Actions, and Events, (G3) Monitor Processes, Materials or Surroundings, (G4) Inspecting Equipment, Structures or Material, (G5) Estimate the Quantifiable Characteristics of Products, Events, or Information
4.2 Mental Process	(G6) Judging the Qualities of Things, Services or People, (G7) Evaluating Information to Determine Compliance with Standards, (G8) Processing Information, (G9) Analysing Data or Information, (G10) Making Decisions and Solving Problems, (G11) Thinking Creatively, (G12) Updating and Using Relevant Knowledge, (G13) Developing Objectives and Strategies, (G14) Scheduling Work and Activities, (G15) Organising, Planning, and Prioritising Work
4.3 Work Output	(G16) Performing General Physical Activities, (G17) Handling and Moving Objects, (G18) Controlling Machines and Processes, (G19) Interacting With Computers, (G20) Operating Vehicles, Mechanised Devices, or Equipment, (G21) Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment, (G22) Repairing and Maintaining Mechanical Equipment, (G23) Repairing and Maintaining Electronic Equipment, (G24) Documenting/Recording Information
4.4 Interacting With Others	(G25) Interpreting the Meaning of the Information for Others, (G26) Communicating with Supervisors, Peers, or Subordinates, (G27) Communicating with Persons Outside Organisation, (G28) Establishing and Maintaining Interpersonal Relationships, (G29) Assisting and Caring for Others, (G30) Selling or Influencing Others, (G31) Resolving Conflicts and Negotiating with Others, (G32) Performing for or Working in Directly with the Public, (G33) Coordinating the Work and Activities of Others, (G34) Developing and Building Teams, (G35) Training and Teaching Others, (G36) Guiding, Directing, and Motivating Subordinates, (G37) Train and Nurture Other People, (G38) Provide Consultation and Advice to Others, (G39) Performing Administrative Activities, (G40) Staffing Organisational Units, (G41) Monitoring and Controlling Resources

Author's own elaboration on ICP 2013[15] and O*NET data descriptors[16].

Table A2. Detected Communities

Cluster 1: Social-Cognitive	Critical Thinking, Active Learning, Active Listening, Administration and Management, Analysing Data or Information, Assisting and Caring for Others, Category Flexibility, Communicating with Persons Outside Organisation, Communicating with Supervisors,Peers, or Subordinates, Communication and Media, Complex Problem Solving, Coordinating the Work and Activities of Others, Deductive Reasoning, Developing Objectives and Strategies, Developing and Building Teams, Documenting/Recording Information Economics and Accounting, Education and Training, Establishing and Maintaining Interpersonal Relationships, Fine Arts, Flexibility of Closure, Fluency of Ideas, Food Production, Foreign Language, Geography, Getting information, Guiding, Directing, and Motivating Subordinates, History and Archaeology, Human Resources Management, IT and Electronics, Identifying Objects, Actions, and Events, Inductive Reasoning, Information Ordering, Instructing, Interacting With Computers, Interpreting the Meaning of the Information for Others, Italian Language, Judging the Qualities of Things, Services or People, Judgement and Decision Making, Learning Strategies, Legislation and Institutions, Making Decisions and Solving Problems, Management of Financial Resources, Management of Personnel Resources, Medicine and Dentistry, Memorisation, Monitoring, Monitoring and Controlling Resources, Negotiation, Number Facility, Office Work, Oral Comprehension, Oral Expression, Organising, Planning, and Prioritising Work, Originality, Performing Administrative Activities, Performing for or Working in Directly with the Public, Persuasion, Philosophy and Theology, Problem Sensitivity, Processing Information, Provide Consultation and Advice to Others, Psychology, Reading Comprehension, Resolving Conflicts and Negotiating with Others, Sales and Marketing, Scheduling Work and Activities, Selling or Influencing Others, Service Orientation, Services to Customers, Social Perceptiveness, Sociology and Anthropology, Speaking, Speech Clarity, Speech Recognition, Speed of Closure, Staffing Organisational Units, Telecommunications, Therapy and Counseling, Thinking Creatively, Time Management, Time Sharing, Training and Teaching Others, Updating and Using Relevant Knowledge, Writing, Written Comprehension, Written expression
Cluster 2: Technical-Physical	Mathematics, Science, Biology, Building and Construction, Arms-Hand Steadiness, Auditory Attention, Chemistry, Civil Protection and Public Safety, Control Precision, Controlling Machines and Processes, Coordination, Depth Perception, Drafting, Laying Out, and Specifying Technical Devices Parts and Equipment, Dynamic Flexibility, Dynamic Strength, Engineering and Technology, Equipment Maintenance, Equipment Selection, Estimate the Quantifiable Characteristics of Products, Events or Information, Evaluating Information to Determine Compliance with Standards, Explosive Strength, Extent Flexibility, Far Vision, Finger Dexterity, Glare Sensitivity, Gross Balance Body Equilibrium, Gross Body Coordination, Handling and Moving Objects, Hearing Sensitivity, Inspecting Equipment, Structures or Material, Installation, Management of Material Resources, Manual Dexterity, Math Reasoning, Mathematics, Mechanics, Monitor Processes, Materials or Surroundings, Multilimb Coordination, Near Vision, Night Vision, Operating Vehicles, Mechanised Devices, or Equipment, Operation Monitoring, Operation and Control, Operations Analysis, Perceptual Speed, Performing General Physical Activities, Peripheral Vision, Physics, Production and Processing, Programming, Quality Control Analysis, Rate Control, Reaction Time, Repairing, Repairing and Maintaining Electronic Equipment, Repairing and Maintaining Mechanical Equipment, Response Orientation, Selective Attention, Sound Localisation, Spatial Orientation, Speed of Limb Movement, Stamina, Static Strength, Systems Analysis, Systems Evaluation, Technical Design, Technology Design, Train and Nurture Other People, Transportation, Troubleshooting, Trunk Strength, Visual Colour Discrimination, Visualisation, Wrist-Finger Speed

Table A3. Summary Statistics for Dependent Variables

	2013-2016	2014-2017	2015-2018	2016-2019	Total
Entry (with RCA)					
0	36,854	37,044	37,212	37,255	148,365
1	2,679	2,569	2,365	2,308	9,921
Exit (with RCA)					
0	13,826	13,985	14,137	14,202	56,150
1	2,709	2,470	2,354	2,303	9,836
Entry (with RCA^1)					
0	32,986	33,119	33,118	32,951	132,174
1	3,160	3,044	3,129	3,215	12,548
Exit (with RCA^1)					
0	16,742	16,876	16,852	16,866	67,336
1	3,180	3,029	2,969	3,036	12,214

Table A4. Summary Statistics for Dependent Variables

	obs.	mean	sd	min	max
three years					
Entry (RCA)	158,286	0.062	0.24	0	1
Exit (RCA)	65,986	0.15	0.36	0	1
Entry (RCA^1)	144,722	0.09	0.28	0	1
Exit (RCA^1)	79,550	0.15	0.36	0	1
one year					
Entry (RCA)	236,337	0.038	0.191	0	1
Exit (RCA)	98,787	0.091	0.287	0	1
Entry (RCA^1)	216,001	0.055	0.228	0	1
Exit (RCA^1)	119,123	0.099	0.298	0	1

Notes: Summary statistics for dependent variables are reported.

Table A5. Correlation Matrix

	F. ASRD	*SC ASRD*	*TP ASRD*	*F. Comp.*	*SC Comp.*	*TP Comp.*	*Reg. Diversity*	*Ind. Ubiquity*	*Churn*	*Bus. Growth*	*GDP (log)*	*Pop. Density*	*Educ.*
Full ASRD	1.0000												
SC ASRD	0.9747	1.0000											
TP ASRD	0.9754	0.9469	1.0000										
Full Complexity	0.2633	0.2381	0.2593	1.0000									
SC Complexity	0.2282	0.2191	0.2389	0.6995	1.0000								
TP Complexity	0.2181	0.2092	0.2108	0.6664	0.6958	1.0000							
Reg. Diversity	0.4847	0.4857	0.4710	0.0985	0.0941	0.0899	1.0000						
Ind. Ubiquity	0.0580	0.0884	0.0773	0.2221	0.1530	0.1161	0.0000	1.0000					
Churn	-0.3160	-0.3145	-0.3080	-0.0242	-0.0816	-0.0442	-0.0043	0.1488	1.0000				
Bus. Growth	0.0446	0.0341	0.0259	0.0361	0.0254	0.0678	0.0647	-0.0483	0.2458	1.0000			
GDP (log)	0.6688	0.6766	0.6494	0.1418	0.1703	0.1432	0.1066	0.0000	-0.4751	-0.0166	1.0000		
Pop. Density	0.4310	0.4296	0.4114	0.1048	0.0888	0.0854	0.3897	-0.0000	-0.0283	0.0773	0.2569	1.0000	
Education	0.1181	0.1088	0.1124	0.0235	0.0135	0.0172	0.0435	0.0008	0.0618	0.1105	0.0807	0.1472	1.0000

Table A6. Variance Inflation Factors (VIF)

	Table 1	Table 1	Table 2	Table 2	Table 3	Table 3
	Col. 3	Col. 6	Col. 3	Col. 6	Col. 3	Col. 6
Full ASRD	2.84	3.13				
Full Complexity	1.12	1.16				
SC ASRD			3.02	2.80		
SC Complexity			1.08	1.07		
TP ASRD					2.54	2.72
TP Complexity					1.06	1.07
GDP (log)	2.37	2.29	2.53	2.20	2.24	2.17
Pop. Density	1.28	1.48	1.29	1.48	1.27	1.47
Education	1.04	1.06	1.04	1.06	1.04	1.06
Churn	1.48	1.49	1.48	1.50	1.48	1.49
Bus. Growth	1.12	1.08	1.12	1.08	1.13	1.08
Reg. Diversity	1.52	1.72	1.55	1.69	1.47	1.67
Ind. Ubiquity	1.10	1.21	1.09	1.21	1.08	1.17
Mean VIF	1.54	1.63	1.58	1.57	1.48	1.54

Table A7. Entry and Exit Models: Logistic and Probit Estimates

	Entry						Exit					
	(1) Logistic	(2) Logistic	(3) Logistic	(4) Probit	(5) Probit	(6) Probit	(7) Logistic	(8) Logistic	(9) Logistic	(10) Probit	(11) Probit	(12) Probit
Full ASRD	0.03253*** (0.00289)			0.01679*** (0.00138)			-0.02130*** (0.00305)			-0.01181*** (0.00165)		
Full Complexity	0.00469*** (0.00044)			0.00248*** (0.00022)			-0.00472*** (0.00046)			-0.00258*** (0.00025)		
SC ASRD		0.03351*** (0.00257)			0.01743*** (0.00126)			-0.02370*** (0.00279)			-0.01307*** (0.00149)	
SC Complexity		0.00682*** (0.00044)			0.00353*** (0.00022)			-0.00721*** (0.00046)			-0.00403*** (0.00025)	
TP ASRD			0.03855*** (0.00296)			0.02007*** (0.00141)			-0.02181*** (0.00303)			-0.01218*** (0.00166)
TP Complexity			0.00574*** (0.00049)			0.00301*** (0.00025)			-0.00558*** (0.00057)			-0.00310*** (0.00030)
GDP (log)	-0.26323 (0.41859)	-0.30919 (0.42067)	-0.34681 (0.41935)	-0.13812 (0.20469)	-0.16270 (0.20554)	-0.17445 (0.20522)	-0.17795 (0.43528)	-0.15335 (0.43658)	-0.15867 (0.43479)	-0.11436 (0.24033)	-0.10489 (0.24075)	-0.10378 (0.24013)
Pop. Density	0.07460 (0.74941)	0.08823 (0.75386)	0.18145 (0.75093)	0.04347 (0.36682)	0.05448 (0.36856)	0.08585 (0.36775)	-0.58826 (0.78513)	-0.55737 (0.78710)	-0.63379 (0.78414)	-0.29405 (0.43358)	-0.27422 (0.43410)	-0.31549 (0.43314)
Education	-0.03054* (0.01575)	-0.03381** (0.01583)	-0.02684* (0.01576)	-0.01657** (0.00769)	-0.01859** (0.00773)	-0.01541** (0.00770)	-0.02139 (0.01679)	-0.01947 (0.01689)	-0.02497 (0.01679)	-0.01217 (0.00916)	-0.01148 (0.00919)	-0.01443 (0.00915)
Churn	-0.01209* (0.00693)	-0.01165* (0.00696)	-0.00918 (0.00694)	-0.00684** (0.00337)	-0.00665** (0.00338)	-0.00512 (0.00337)	0.01068 (0.00727)	0.01148 (0.00731)	0.01038 (0.00731)	0.00644 (0.00402)	0.00685* (0.00403)	0.00617 (0.00403)
Bus. Growth	0.01824*** (0.00496)	0.01720*** (0.00498)	0.01773*** (0.00498)	0.00967*** (0.00244)	0.00920*** (0.00245)	0.00919*** (0.00245)	-0.01194** (0.00561)	-0.01047* (0.00562)	-0.01257** (0.00561)	-0.00658** (0.00313)	-0.00563* (0.00314)	-0.00700** (0.00313)
Reg. Diversity	-0.01984*** (0.00255)	-0.02009*** (0.00257)	-0.01984*** (0.00256)	-0.00954*** (0.00125)	-0.00972*** (0.00125)	-0.00952*** (0.00125)	0.02373*** (0.00267)	0.02418*** (0.00267)	0.02396*** (0.00266)	0.01317*** (0.00146)	0.01344*** (0.00146)	0.01335*** (0.00145)
Ind. Ubiquity	-0.06302*** (0.00441)	-0.06256*** (0.00442)	-0.06248*** (0.00442)	-0.03448*** (0.00228)	-0.03421*** (0.00228)	-0.03426*** (0.00228)	0.06533*** (0.00503)	0.06481*** (0.00503)	0.06529*** (0.00503)	0.03784*** (0.00278)	0.03752*** (0.00278)	0.03777*** (0.00278)
Full FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	155,583	155,583	155,583	155,583	155,583	155,583	65,649	65,649	65,649	65,649	65,649	65,649

Notes: Robust standard errors clustered at the region and industry level are in parentheses. All specifications include fixed effects for region, industry, and time. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

References

Alabdulkareem, A., M. R. Frank, L. Sun, B. AlShebli, C. Hidalgo, and I. Rahwan (2018). "Unpacking the polarization of workplace skills". In: *Science Advances* 4, pp. 1–9.

Anderson, K. A. (2017). "Skill networks and measures of complex human capital". In: *Pnas* 114(48, pp. 12720–12724.

Antonietti, R. and R. Boschma (2021). "Social Capital, Resilience, and Regional Diversification in Italy". In: *Industrial and Corporate Change* 30(1), pp. 762–777.

Antonietti, R., L. Cattani, F. Gambarotto, and G. Pedrini (2021). "Education, Routine, and Complexity- Biased Knowledge Enabling Technologies: Evidence from Emilia- Romagna, Italy". In: *Gran Sasso Science Institute* Discussion paper No. 2021-07.

Bahar, D., S. Rosenow, E. Stein, and R: Wagner (2018). "Export Take-Offs and Acceleration: Unpacking Cross-Sector Linkages in the Evolution of Comparative Advantage". In: *Iza Discussion Paper Series* 12061.

Balland, P. A. (2016). "EconGeo: Computing Key Indicators of the Spatial Distribution of Economic Activities". In.

Balland, P. A. and R. Boschma (2019). "Smart Specialisation: beyond patents". In: *Report for European Commission* volume.

Balland, P. A., R. Boschma, J. Crespo, and D. Rigby (2019). "Smart specialization policy in the EU: Relatedness, knowledge complexity and regional diversification". In: *Regional Studies* 53(9), pp. 1252–1268.

Balland, P. A., T. Broekel, D. Diodato, E. Giuliani, R. Hausmann, N. O'Clery, and D. Rigby (2022). "The new paradigm of economic complexity". In: *Research Policy* 51.3, p. 104450. DOI: https://doi.org/10.1016/j.respol.2021.104450.

Balland, P. A. and D. Rigby (2017). "The Geography of Complex Knowledge". In: *Economic Geography* 93(1), pp. 1–23. URL: https://doi.org/10.1080/00130095.2016.1205947.

Balland, P. A., D. Rigby, and R. Boschma (2015). "The technological resilience of US cities". In: *Cambridge Journal of Regions, Economy and Society* 8(2), pp. 167–184. URL: https://doi.org/10.1093/cjres/rsv007.

Blondel, V. D., J. Guillaume, R. Lambiotte, and E. Lefebvre (Oct. 2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. URL: https://doi.org/10.1088/1742-5468/2008/10/p10008.

Boschma, R. (2017). "Relatedness as driver behind regional diversification: A research agenda". In: *Regional Studies* 51(3), pp. 351–364.

Boschma, R., P. A. Balland, and D. F. Kogler (2015). "Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010". In: *Industrial and Corporate Change* 24(1), pp. 223–250.

Boschma, R., R. Eriksson, and U. Lindgren (2014). "Labour market externalities and regional growth in Sweden: the importance of labour mobility between skill-related industries". In: *Regional Studies* 48(10, pp. 1669–1690.

Boschma, R., A. Minondo, and M. Navarro (2013). "The emergence of new industries at the regional level in Spain: A proximity approach based on product relatedness". In: *Economic Geography* 89(I), pp. 29–51.

Breschi, S., F. Lissoni, and F. Malerba (2003). "Knowledge-relatedness in firm technological diversification". In: *Research Policy* 32, pp. 69–87.

Caines, C., F. Hoffmann, and G. Kambourov (2017). "Complex-Task biased technological change and the labor market". In: *nternational Finance Discussion Papers* 1192.

Chang, S. J. (1996). "An evolutionary perspective on diversification and corporate restructuring: entry, exit, and economic performance 1981-89". In: *Strategic Management Journal* 17, pp. 587–611.

Diodato, D. and A. Weterings (2015). "The resilience of regional labour markets to economic shocks: exploring the role of interactions among firms and workers". In: *Journal of Economic Geography* 15(4), pp. 723–742.

Elekes, Z., A. Baranowska-Rataj, and R. Eriksson (2021). "Local access to skill-related high-income jobs facilitates career advancement for low-wage workers". In: *Papers in Evolutionary Economic Geography* 21.36, pp. 365–369.

Farinha, T. F., P. A. Balland, A. Morrison, and R. Boschma (2019). "What drives the geography of jobs in the US? Unpacking relatedness". In: *Industry and Innovation* 26 (9), pp. 988–1022.

Farjoun, M. (1994). "Beyond Industry Boundaries: Human Expertise, Diversification and Resource-Related Industry Groups". In: *Organization science* 5(2), pp. 185–199.

— (1998). "The independent and join effects of the skill and physical bases of relatedness in diversification". In: *Strategic Management Journal* 19(7), pp. 611–630.

Feser, E. J. (2003). "What Regions Do Rather than Make: A Proposed Set of Knowledge-based Occupation Clusters". In: *Urban Studies* 40(10), pp. 1937–1958.

Fitjar, R. D. and B. Timmermans (2017). "Regional skill relatedness: towards a new measure of regional related diversification". In: *European Planning Studies* 25(3), pp. 516–538. URL: https://doi.org/10.1080/09654313.2016.1244515.

Gabe, T. and J. R. Abel (2011). "Agglomeration of Knowledge". In: *Urban Studies* 48(7), pp. 1353–1371.

Greene, W. H. (2012). "Econometric analysis". In: 7th ed.

Guo, Q., S. Zhu, and R. Boschma (2020). "Networks of export markets and export market diversification". In: *Industrial and Corporate Change* volume, pp. 1–17.

Hallevik, O. (2009). *Linear versus logistic regression when the dependent variable is a dichotomy.* URL: https://EconPapers.repec.org/RePEc:spr:qualqt:v:43:y:2009:i:1:p:59-74.

Hausmann, R. and B. Klinger (2007). "The structure of the product space and the evolution of comparative advantage". In: *Working Paper Cambridge, MA: Center for International Development, Harvard University* 146.

Hidalgo, C. A. and R. Hausmann (2009). "The building blocks of economic complexity". In: *Proceedings of the National Academy of Sciences* 106(26), pp. 10570–10575.

Hidalgo, C. A., B. Klinger, A. L. Barabasi, and R. L. Hausmann (2007). "The Product Space Conditions the Development of Nations". In: *Science* 317.317 (5837), pp. 482–487.

Hidalgo, C. A. et al. (2018). "The principle of relatedness". In: *Springer Proceedings in Complexity*, pp. 451–457.

Kogler, D. F. (2017). "Relatedness as driver of regional diversification: a research agenda â a commentary". In: *Regional Studies* 51(3).3, pp. 365–369. DOI: 10.1080/00343404.2016.1276282.

Kogler, D. F., D. L. Rigby, and I. Tucker (2013). "Mapping Knowledge Space and Technological Relatedness in U.S. Cities". In: *European Planning Studies* 21(9), pp. 1374–1391.

Krenz, A. (2014). “Agglomeration of knowledge: A regional economic analysis for the German economy”. In: *Cege Discussion Papers, No. 206* volume.

Lee, G. K. and M. B. Lieberman (2010). “Acquisition vs. internal development as modes of market entry”. In: *Strategic Management Journal* 31, pp. 140–158.

LoTurco, A. and D. Maggioni (2020). “The knowledge and skill content of production complexity”. In: *Research Policy* volume.

Mewes, L. and T. Broekel (2020). “Technological Complexity and Economic Growth of Regions”. In: *Papers in Evolutionary Economic Geography (PEEG)* 2049.

Montesor, S. and F. Quatraaro (2017). “Regional Branching and Key Enabling Technologies: Evidence from European Patent Data”. In: *Economic Geography* 93(4), pp. 367–396.

Muneepeerakul, R., J. Lobo, S. T. Shutters, A. Gomez-Lievano, and M. R. Qubbai (2013). “Urban economies and occupation space: Can they get ‘there’ from ‘here’?” In: *PLoS ONE* 8(9).

Neffke, F., M. Hartog, R. Boschma, and M. Henning (2018). “Agents of Structural Change: The Role of Firms and Entrepreneurs in Regional Diversification”. In: *Economic Geography* 94(1), pp. 23–48. URL: https://doi.org/10.1080/00130095.2017.1391691.

Neffke, F., M. Henning, and R. Boschma (2011). “How do regions diversify over time? Industry relatedness and the development of new growth paths in regions”. In: *Economic Geography* 87(3), pp. 237–265.

Neffke, F. and M. S. Henning (2008). “Revealed Relatedness: Mapping Industry Space”. In: *Papers in Evolutionary Economic Geography* Urban and Regional research centre Utrecht.

— (2013). “Skill-relatedness and firm diversification”. In: *Strategic Management Journal* 34, pp. 297–316.

Neffke, F., A. Otto, and A. Weyh (2017). “Inter-industry labor flows”. In: *Journal of Economic Behavior Organization* 142, pp. 275–292.

Rahmati, R., A. Tafti, J. Westland, and C. Hidalgo (2021). “When All Products Are Digital: Complexity and Intangible Value in the Ecosystem of Digitizing Firms”. In: *MIS Quarterly* 45, pp. 1025–1058. DOI: 10.25300/MISQ/2021/15384.

Rigby, D. L. (2015). “Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes”. In: *Regional Studies* 49(11), pp. 1922–1937. URL: https://doi.org/10.1080/00343404.2013.854878.

Tanner, A. N. (2014). “Regional Branching Reconsidered: Emergence of the Fuel Cell Industry in European Regions”. In: *Economic Geography* 90(4), pp. 403–427. URL: https://onlinelibrary.wiley.com/doi/10.1111/ecge.12055.

— (2016). “The emergence of new technology-based industries: the case of fuel cells and its technological relatedness to regional knowledge bases”. In: *Journal of Economic Geography* 16(3), pp. 611–635.

Timmermans, B. and R. Boschma (2014). “The effect of intra- and inter-regional labour mobility on plant performance in Denmark: the significance of related labour inflows”. In: *Journal of Economic Geography* 14(2), pp. 289–311.

VanEck, N. J. and L. Waltman (2009). “How to normalize cooccurrence data? An analysis of some well-known similarity measures”. In: *Journal of the American Society for Information Science and Technology* 8, 1635â1651.

Whittle, A. (2019). “Local and Non-Local Knowledge Typologies: Technological Complexity in the Irish Knowledge Space”. In: *European Planning Studies* 27(4), pp. 661–677.

Whittle, A. and D. F. Kogler (2019). “Related to what? Reviewing the literature on technological relatedness: Where we are now and where can we go?” In: *Pap Reg Sci.*, pp. 1–17.

Xiao, J., R. Boschma, and M. Anderson (2018). “Industrial diversification in Europe: The differentiated role of relatedness,” in: *Economic Geography* 94(5), pp. 514–49.