Skill capabilities behind the scenes: The role of occupational portfolio in regional industrial evolution

Jing Chen, Xiaojing Li, Xiaoqi Zhou & Rongjun Ao

Papers in Evolutionary Economic Geography

25.08

Utrecht University Human Geography and Planning

Skill capabilities behind the scenes: The role of occupational portfolio in regional industrial evolution

Jing Chen[1], Xiaojing Li[1], Xiaoqi Zhou[2], Rongjun Ao[2]

Affinitions:

1. Key Laboratory for Geographical Process Analysis and Simulation of Hubei Province, Central China Normal University, Wuhan, China
2. College of Urban and Environmental Sciences, Central China Normal University, Wuhan, China

Correspondence author:

Xiaoqi Zhou, College of Urban and Environmental Sciences, Central China Normal University; China; Email: zhouxiaoqi@mails.ccnu.edu.cn

Funding:

The authors would like to acknowledge funding from the Natural Science Foundation of China (grant numbers 42001136 and 42271188).

Note:

This is the working paper version of a published paper. Please cite the published version: Chen, J, Li, X., Zhou, X., & Ao, R. (2025). Skill capabilities behind the scenes: The role of occupational portfolio in regional industrial evolution. *Regional Studies*, *59*(1), 2416535. https://doi.org/10.1080/00343404.2024.2416535

Abstract:

An increasing number of studies confirm that regional diversification is path-dependent, with new industries building on pre-existing ones. However, these studies typically overlook the role of skill capabilities and the interactions between skills and industries. In this research, based on the concept of industry-occupation cross-relatedness, the influence of skill capabilities on industrial diversification across Chinese regions was investigated. Findings indicate that regions can diversify into skill-related activities following a skill path-dependent process. Furthermore, a theoretical framework integrating industry-occupation cross-relatedness and economic complexity was introduced, which enabled the adoption of various diversification strategies based on regional skill capabilities.

Keywords:

Industrial evolution, cross-relatedness, path dependence, skill capabilities, Chinese regions

JEL codes:

R23, J24, O18, R11

1 Introduction

Over the past decade, the notions of relatedness and complexity have been widely applied to explain the economic diversification of firms, cities, regions, and countries (Whittle and Kogler, 2020; Hidalgo, 2023). Relatedness assumes that regions do not diversify into new economic activities randomly, but draw on existing capabilities (Boschma, 2017). Specifically, the industrial evolution of a region generally features path dependency and its future development path is determined by current economic activities. Meanwhile, complexity assumes that economic complexity plays a decisive role in the rewards of economic diversification (Balland *et al.*, 2019). It is further argued that less ubiquitous products ought to be more complex and require higher levels of regional capabilities. Drawing on these insights, empirical research in evolutionary economic geography (EEG) applies relatedness and complexity to interpret the evolving mechanism of industries, products, skills, and occupations (Xu *et al.*, 2021; Buyukyazici *et al.*, 2024).

More recently, an emerging strand of research in skill relatedness has attracted scholarly interest. Based on advances in inter-industry labor flows and occupational skills in labor economics, EEG scholars further defined and measured skill relatedness in forms of industry space, occupation space, and skill space (Muneepeerakul *et al.*, 2013; Neffke and Henning, 2013; Alabdulkareem *et al.*, 2018). It is argued that skill relatedness provides at least four additional insights compared with technological or industrial relatedness (Wixe and Andersson, 2017; Galetti *et al.*, 2021). First, skill relatedness starts from a micro perspective on labor concerning their skills, tasks, and knowledge, which is different from industrial relatedness taking firms as the starting point to analyze economic diversification (Whittle and

Kogler, 2020). Second, rather than focusing on the final outputs of economic activities, skill relatedness emphasizes understanding what regions do—highlighting the processes and capabilities involved rather than just the products made (Farinha *et al.*, 2019). Third, industry relatedness is usually defined broadly and covers various factors like labor skills and input-output linkages (Boschma, 2017), whereas skill relatedness pays attention to similarities and complementarities between skills. Fourth, several authors underscored that skill relatedness can cover the impacts of both manufacturing and service sectors, whereas industrial relatedness focuses solely on the manufacturing sector (Hane-Weijman *et al.*, 2022). Along these lines, the corresponding empirical findings have not only validated the path-dependent feature of skill diversification (Muneepeerakul *et al.*, 2013), but also explored the role of skill capabilities in regional diversification (Buyukyazici *et al.*, 2024) despite disparities in economic settings, relatedness measures, and study periods.

While considerable progress has been made in understanding the skill aspects of regional capabilities, a significant research gap still exists. That is, consideration has not been given to the endogenous interaction between skill portfolios and industry activities when the relationship between skill capabilities and regional industry evolution is studied. Given that this interaction is rooted in relevant studies (Thompson and Thompson, 1987; Sunley *et al.*, 2020), the absence of such linkages may underestimate the importance of skill capabilities. For instance, university graduates usually choose to work in industries closely related to their knowledge and skills; in the meantime, firms often examine the knowledge and skills of job seekers before hiring them. More broadly, labor can impact the entry of industrial activities and industrial activities can also impact the evolution of skill capabilities because of labor pooling and matching effects (Holm *et al.*, 2017; Chen *et al.*, 2024). To fill the research gap in skill-industry interactions, this study aimed to establish the concept of industry-occupation cross-relatedness based on existing studies on industry relatedness, occupation relatedness, cross-relatedness on the one hand, and to examine the role of skill capabilities in regional diversification through the lens of industry-occupation cross-relatedness on the other hand.

With these aims, the contributions of this study can be summarized into the following aspects. First, it contributes to studies on the relationship between skill capabilities and industrial diversification by considering the interplay between skill and industry portfolios. To do that, this study systematically overviewed studies on the role of skill capabilities in economic diversification (Galetti *et al.*, 2021; Buyukyazici *et al.*, 2024). Then, it proposed the concept of industry-occupation cross-relatedness that considers the interplay between skill capabilities and industry activities. To our knowledge, unlike existing studies that focus on either industry relatedness or occupation relatedness, this paper is one of the early studies explicitly considering the interaction between skill and industry structures from the perspective of cross-relatedness when analyzing the relationship between skill capabilities and regional diversification.

Second, this study also complements the literature on diversification trajectories based on relatedness and complexity, especially the hidden role of skill capabilities in economic diversification. To be specific, it is found that economic activities can diversify into not only industry-related but also skill-related activities, which yields a critical implication for literature. Previously, industry relatedness was thought to capture a set of regional capabilities, including knowledge, technologies, skills, etc. (Boschma, 2017; Whittle and

Kogler, 2020) and it is challenging to isolate the impact of skill capabilities. Industry relatedness is thus deemed as the only metric to differentiate between related and unrelated diversification. Nevertheless, current research notes that this argument might not hold for skill-based industry diversification and these *behind-the-scene* skill capabilities can be a potential source of regional development. Furthermore, a cross-relatedness and complexity space was also established to differentiate four types of diversification trajectories, including (1) high cross-relatedness and high complexity, (2) high cross-relatedness and low complexity, (3) low cross-relatedness and high complexity, and (4) low cross-relatedness and low complexity. The framework can be integrated with the existing relatedness/complexity framework (Balland *et al.*, 2019; Hidalgo, 2023) to further identify new industries that align with or diverge from the current skill and/or industry portfolios, and also to assess whether their economic complexity will increase or decrease. These insights allow policymakers to craft customized skill policies that leverage regional branching opportunities for regional development.

Third, this study enriches the literature on cross-relatedness among multiple types of economic activities, especially the applicability of cross-relatedness in studying the nexus between occupations and industries. Theoretically, the concept of cross-relatedness enables researchers to explore relatedness not restrained to only industries, technologies, or products and demonstrates that relatedness is not unidimensional but multidimensional across different forms of economic activities (Catalán *et al.*, 2022; Hidalgo, 2023). Unlike existing studies that have examined the endogenous interactions between scientific, industrial, and technological portfolios (Balland and Boschma, 2022), this study integrated occupational relatedness, industrial evolution, and cross-relatedness into the concept of industry-occupation cross-relatedness and probed the theoretical constructions in the nexus between skill capabilities and regional diversification.

The remainder of this study is structured as follows. Section 2 introduces the theoretical background, followed by the presentation of data and methods in Section 3. The results are illustrated in Section 4 and the final section concludes.

2 Theoretical background

2.1 Relatedness, cross-relatedness, and complexity

Within the past two decades, relatedness has seen extensive application in explaining economic diversification and covers products (Hidalgo *et al.*, 2007), industries (Neffke *et al.*, 2011), skills (Muneepeerakul *et al.*, 2013), and technologies (Kogler *et al.*, 2013). These studies have reached a consensus that local capabilities influence the likelihood of new activities developing in regions. It is assumed that a positive effect of relatedness on the entry probability of a new activity indicates related diversification, whereas a negative one suggests unrelated diversification (Boschma, 2017). Although related diversification is more frequent by and large, unrelated diversification plays a crucial role in creating new development paths (Pinheiro *et al.*, 2022). For instance, existing studies like Zhu et al. (2017), Hu and Hassink (2017) and He et al. (2018) revealed that extra-linkages and institutions play a nonnegligible role in path creation. Particularly, based on firm-level industrial data from 1998 to 2008, He et al. (2018) confirmed the roles of global linkages, economic liberalization, and state involvement in creating new industrial development paths among Chinese prefectures.

Despite this, it is noteworthy that relatedness is not the sole factor shaping regional diversification. As suggested by EEG scholars, complexity also can be used to explain the evolution of products, industries, occupations, and technologies (Hidalgo, 2023). To take one example, Balland et al. (2019) proposed a framework integrating both relatedness and complexity to illustrate different diversification paths. Unlike relatedness, complexity grants a different way of evaluating the potential economic benefits of activities. By definition, economic complexity refers to activities that are difficult to copy and thus have huge economic value (Hidalgo and Hausmann, 2009). In other words, simple economic activities can be easily replicated and relocated, while complex ones are challenging to imitate and less mobile. Another feature of complex economic activities is tacit knowledge rooted in individual firms and local networks, which provides a source of competitive advantage (Maskell and Malmberg, 1999). Considering that complex economic activities can generate high rents and embody tacit knowledge, it is assumed that the higher the complexity of an activity, the higher its potential economic benefits will be. Correspondingly, empirical evidence unveiled that complex economic activities are hard to accomplish despite their attractiveness (Balland *et al.*, 2019; Hane-Weijman *et al.*, 2022).

Apart from focusing on relatedness and complexity, recent EEG scholars have paid attention to the crossing phenomena in economic diversification by introducing the concept of cross-relatedness. To illustrate, Pugliese et al. (2019) first established a three-layer network visualizing the interactions among patents, industries, and scientific publications. In a follow-up study, Catalán et al. (2022) probed into the connection between scientific portfolios and technological diversification among 182 countries by formulating the concept of scientific and technological cross-relatedness. It is discovered that technological diversification is associated with the scientific and technological portfolios of a region. Their concept of cross-relatedness is vital since it involves the effects of scientific and technological structures as a whole and extends the principle of relatedness to more knowledge dimensions. At the regional level, Balland and Boschma (2022) examined the role of scientific domains in technological diversification in Europe. Similarly, Castaldi and Drivas (2023) explored how relatedness and cross-relatedness among technologies, designs, and market activities condition the evolution of regional innovation specializations in the United States and Europe. Notwithstanding fruitful empirical studies on cross-relatedness, they are mostly technology-centered, which further motivates this study to analyze the relationship between skill capabilities and regional diversification from the perspective of cross-relatedness. In view of this, the next subsection explores the corresponding literature.

2.2 Skill capabilities and regional industrial diversification

The accumulation of human capital and skills has long been recognized as the key driving force of economic growth and regional development (Romer, 1990). For instance, Schultz (1961) identified the economic value of a worker’s education, skills, and health as capital input into the economic process. In the past decades, more attention has been paid to the role of human capital and skills in the course of knowledge spillover and structural transformation of regional economies. This shift can possibly be attributed to changes in the spatial division of labor and the era of globalization and digitalization, given that labor employed by the same industry may vary in skills across regions, and capabilities cannot be fully reflected in specific sectors or products (Barbour and Markusen, 2007; Whittle and Kogler, 2020).

Within the field of EEG, scholars have developed the concept of skill relatedness and measured the skill aspect of regional capabilities with three general space categories, industry space, occupation space, and skill space. First, inspired by Hidalgo et al. (2007), Neffke and Henning (2008) were the first to measure the degree of skill relatedness between industries based on the intensity of interindustry labor flows in Sweden and visualized the relatedness in the form of industry space. It is posited that different industries are interconnected by skill similarities or complementarities. From a micro perspective, workers are more likely to make use of their skill sets to avoid skill destruction when switching jobs from one industry to another (Holm *et al.*, 2017; Farinha *et al.*, 2019). Therefore, it is argued that industries requiring similar skills simultaneously are considered skill-related.

Second, the occupation-based approach to understanding regional dynamics is well-established in regional science and economic geography (Thompson and Thompson, 1987; Sunley *et al.*, 2020). Muneepeerakul et al. (2013) were pioneers in measuring skill relatedness between occupations based on co-location patterns, which led to the creation of an occupation space. Following this, empirical studies not only verified the path dependency in the evolution of occupations (Farinha *et al.*, 2019) but also examined the influence of skill relatedness on regional development (Cappelli *et al.*, 2019). Typically, these studies operate on the assumption that occupations are utilized as proxies for skills because it is challenging to measure skill combinations accurately.

Third, because occupational data and labor mobility indirectly capture skill relatedness, scholars further set up skill spaces (Alabdulkareem *et al.*, 2018; Xu *et al.*, 2021), where nodes represent skills like stamina and time management and edges between skills stand for their relatedness degree. When required by occupations or industries simultaneously, skill pairs are defined as skill-related.

Based on these understandings, empirical studies further examined the relationship between skill capabilities and industrial diversification. Inspired by Rey (1998), these studies can be classified into two general types concerning their approaches to integrating skill and industry portfolios. As illustrated in Figure 1a, the first approach is *embedding*. Different from the broad concept of industry relatedness, skill capabilities are embedded in the industry portfolio of regions, namely the skill relatedness of industries. This means that industry relatedness captured by inter-industry labor mobility manifests the skill perspective of relatedness among industries. For example, Neffke and Henning (2013) developed the revealed skill relatedness method. Subsequent studies like Cappelli et al. (2019) and Jara-Figueroa et al. (2018) followed this direction to investigate how skill relatedness can exert an impact on the dynamics of industries among firms and regions. Nonetheless, this approach has been questioned given the fact that labor flows might be impacted by seasonal variations (Wixe and Andersson, 2017; Whittle and Kogler, 2020).

[INSERT FIGURE 1]

The second approach is *linking*, where industry-skill matrices are established by linking skills with industry datasets (Figure 1b). To illustrate, Galetti et al. (2021) parsed the co-occurrence patterns among industry-skill pairs and established an industry space to illustrate the skill relatedness of industries. In specific terms, their study multiplied a matrix with skill requirements in rows and occupations in columns by a matrix with occupations in rows and industries in columns. The corresponding result is industry-skill pairs that mirror the skill requirements of industries. Likewise, Buyukyazici et al. (2024) calculated the minimum conditional probability of industry co-occurrences in skill classes (i.e., the industry relatedness of skills). They established a skill space and measured the average skill-relatedness density to quantify how close a potential industry is to the existing portfolio of the region regarding human capital. However, the linking approach requires specific data on the skill requirements of industries, which are often hard to acquire. In this way, only a few studies have successfully merged skills with industries.

Although the embedding or linking approach acknowledges the significance of skill capabilities in industrial diversification, these approaches have faced criticism for their failure to account for the endogenous interactions between labor and firms. As such, this study further formulated a third *crossing* approach by establishing the concept of occupation-industry cross-relatedness when evaluating the role of skill capabilities in regional industrial diversification (Figure 1c). This approach differs essentially from the former two, not only because it uses occupation information as a proxy for labor skills given the difficulty of obtaining skill-level datasets for regions (Wixe and Andersson, 2017; Sunley *et al.*, 2020), but also because the endogenous interactions between occupations and industries are considered.

Theoretically, industry-occupation interactions are grounded in the relevant literature and at least two micro mechanisms between firms and labor can be identified despite the difficulty in differentiating them empirically. From a resource-based perspective (Penrose, 1959; Neffke and Henning, 2013), firms are motivated to diversify into new economic activities based on their resources, including human capital and the skill sets of their workforce. As a pivotal component of economic production, human capital is primarily accumulated at the firm level. Individuals, however, are the actual carriers of essential information, experiences, knowledge, and skills (Wixe and Andersson, 2017). Consequently, firms may periodically enhance the skills of their employees through approaches like external labor recruitment or internal on-the-job training when diversifying into new economic activities. Meanwhile, workers normally apply their knowledge and skills when fulfilling their tasks and obligations. They may seek new opportunities in skill-related industries if failing to meet the skill requirements of a firm (Neffke and Henning, 2013; Holm *et al.*, 2017). From a collective learning perspective, the recombination of industry-specific knowledge from firms and occupation-specific knowledge from individual workers can catalyze a learning process (Frenken *et al.*, 2007; Jara-Figueroa *et al.*, 2018). Nonetheless, it becomes challenging to achieve effective learning and knowledge spillovers when the cognitive distance between firms and individuals is either too great or too small.

Based on these insights, the concept of industry-occupation cross-relatedness in this study is fundamentally distinct from existing studies on cross-relatedness, industry relatedness, and occupational relatedness mainly in two aspects. First, industry relatedness is generally assessed in an *ex post* manner and encompasses a range of factors, including skill capabilities (Boschma, 2017; Whittle and Kogler, 2020), which makes it challenging to isolate the impact of skill portfolios. In contrast, occupational relatedness primarily examines the skill similarities among individual workers. From the perspective of industry chains, skills and knowledge serve as crucial inputs for industrial activities, with the concept of occupation-industry cross-relatedness acting as a vital link. It links industry relatedness with occupational relatedness, which thereby quantifies the interconnections between labor skills and industrial activities. Consequently, unlike technology-based cross-relatedness (Catalán *et al.*, 2022; Castaldi and Drivas, 2023), occupation-industry cross-relatedness can be used for delineating the influence of skill capabilities on industrial diversification.

Second, drawing on the concept of product space (Hidalgo *et al.*, 2007), the notion of industry space explores relatedness between industries, and occupation space examines relatedness between occupations, while occupation-industry cross space evaluates the connections between industries and occupations. This means that nodes in cross relatedness networks are heterogeneous. Hence, regions can leverage skill capabilities to explore potential branching opportunities. For instance, industrial upgrading can be achieved by fostering more robust links to relevant occupations, whereas industry diversification is unlikely to occur through fewer and weaker connections. In other words, scholars and policymakers can determine whether regional diversification trajectories are path-dependent on existing skills or represent a departure from established paths by examining occupation-industry cross-relatedness. Moreover, based on the existing framework of relatedness and complexity (Balland *et al.*, 2019; Hidalgo, 2023), this cross-relatedness approach can also be combined with economic complexity. Accordingly, the subsequent subsection details a framework of cross-relatedness and complexity.

2.3 Towards a framework of cross-relatedness and complexity

For many years, EEG scholars have recognized that industry diversification paths are either path-dependent or path-breaking and influenced by their relatedness to existing industries (Boschma, 2017; Hidalgo, 2023). More recently, developments in economic complexity have led researchers to distinguish four types of diversification trajectories based on relatedness and complexity (Balland *et al.*, 2019; Hidalgo, 2023). These include (1) high-complexity and high-relatedness, (2) high-relatedness and low-complexity, (3) low-complexity and high-relatedness, and (4) low-complexity and low-relatedness. On this basis, this research further introduced a framework that includes both cross-relatedness and complexity at the intersection of industries and occupations. This framework measures the relatedness between labor skills and industrial activities, and also positions these relationships within the broader economic complexity landscape, which unveils regional branching opportunities stemming from skill capabilities rather than industrial structures as a whole. As illustrated in Figure 2, the horizontal axis represents the cross-relatedness between industries and occupations, while the vertical axis denotes the knowledge complexity of industries. The quadrants defined by cross-relatedness and complexity space illustrate how industrial activities interact with skill capabilities. This classification helps scholars and policymakers identify potential diversification opportunities specific to the occupational portfolio of a region and devise targeted skill policies to foster industrial development.

[INSERT FIGURE 2]

The northeast quadrant depicts a scenario where regions diversify into new activities based on their skill structure, and economic complexity increases. This attractive scenario can be found in advanced economies, where firms can reach a varied and sophisticated portfolio of skill capabilities like university graduates and migrated workers. Then, the northwest quadrant manifests a situation where regions can diversify into skill-unrelated but complex economic activities. For example, regions without plenty of skill capabilities—such as cognitive skills, scientific knowledge, and programming—have to recruit non-local workers or provide vocational education for complex industrial development, which can be referred to as a "casino" or "leapfrog" policy (Balland *et al.*, 2019; Qiao *et al.*, 2024). Subsequently, the southeast quadrant suggests that regions can capitalize on existing skill capabilities, but the overall economic complexity decreases. In this regard, current skill portfolios might not contribute to industry upgrading but limit its diversification opportunities. To illustrate, labor-intensive industrial activities like packaging and repairing typically require physical skills. This low-skill trap makes it difficult for the corresponding economy to climb the ladder of economic complexity (Buyukyazici, 2023; Corradini *et al.*, 2023). Finally, the southwest quadrant denotes a scenario where regions branch into activities with low levels of complexity and skill-relatedness. This may occur when newly recruited employees are unable to significantly contribute to the diversification of industries into more complex activities.

The core idea of this cross-relatedness and complexity framework is that industrial capabilities are broadly defined and skill capabilities might be behind the scenes of industrial portfolios. By embracing this idea, the significance of human capital and skills in industrial diversification can be recognized, which constitutes the basis for place-based skill policy making. Previously, regional development policies were too often focused on industry structure as a whole, could not evaluate the role of individual factors like workforce and skills in regional diversification, and thus paid insufficient attention to hidden and localized skill capabilities. To this end, the framework of cross-relatedness and complexity in this study examines the role of skill capabilities in driving related or unrelated diversification into more or less complex industrial activities, which reshapes skill policies from a regional perspective for industrial development. This differs from existing studies on relatedness and complexity that concentrate on regional diversification within technologies, occupations, and industries as a whole (Li and Rigby, 2023; Buyukyazici *et al.*, 2024).

From a practical standpoint, while localized skill capabilities are crucial, occupation portfolios only capture the skill aspect of regional capabilities. By comparison, other factors, such as technologies, input-output relationships, and markets, can also influence regional diversification. Hence, it is argued that the cross-relatedness/complexity framework complements the existing relatedness/complexity framework (Balland *et al.*, 2019; Hidalgo, 2023), rather than replacing it, by working in conjunction. More specifically, industries can be plotted to identify (1) which industries are more sophisticated with high benefits and (2) which industries are more inclined to diversify in the dual perspectives of industry-occupation cross-relatedness and industry relatedness. For example, policies should prioritize workforce and skill development through measures like on-the-job training, formal education, and external recruitment if skill capabilities do not meet the skill requirements (e.g., digital and soft skill) for potential industrial diversification. Conversely, policies should focus on how to make full use of existing skill capabilities if skills are sufficient for industrial goals. To this end, skill policy design becomes more place-based and industry-oriented when integrating the frameworks of relatedness/complexity and cross-relatedness/complexity, given that skill intervention has long followed a place-neutral approach for industrial development (Corradini *et al.*, 2023).

3 Data and methods

3.1 Data

In this study, two categories of data sets were mainly employed. City-level industry data were retrieved from the Chinese Customs Trade Statistics (CCTS), which has been applied frequently for the industrial analysis of Chinese regions (Zhu *et al.*, 2017). Compared with the Annual Survey of Industrial Firms (ASIF), the CCTS covers more time periods and works well with the occupation data for the investigation of cross-relatedness. Nevertheless, since not all industries export products (Fritz and Manduca, 2021), one thing worthy of attention is the potential discrepancies between the export and actual industrial structures of regions. At this point, the CCTS includes each export record that contains detailed information on import and export time, commodity code, export volume, export amount, etc. The study period is 2000–2015, and the data pre-processing procedure is detailed as follows. To begin with, the Harmonized System (HS) codes from different years were adjusted to the 2007 edition of codes. Then, these six-digit HS codes were further transformed into three-digit National Economic Industry Classification (GB/T 4754-2017) codes. To minimize the impact of trading firms, this study identified and eliminated trading firms not involved in manufacturing activities using the method of Ahn, Khandelwal, and Wei (2011). Finally, the research adjusted prefecture-level city administrative division codes between 2000 and 2015, and aggregated enterprise-level data into city-level data. As a result, a matrix containing 176 industries and 277 prefecture-level cities was established.

Occupation data were retrieved from the 2000 Chinese Census and the 2015 Population

Sampling Survey (also known as the 2015 Mini-census), both of which have been widely used in the research on skill structure evolution in China (Shen and Liu, 2016). Since measuring human capital and skills is challenging, most studies have to use the proxies of skills to some extent. For example, education attainment is utilized to figure out the dynamics of human capital (Liu *et al.*, 2017). More recently, scholars have indicated that education attainment might not be an effective indicator of labor skill and thus advocated using occupation to represent skill structure (Sunley *et al.*, 2020). Based on the 2008 edition of the International Standard Classification of Occupations (ISCO-2008) and the 2015 edition of the People's Republic of China Classification of Occupations, an occupation is depicted as a series of jobs sharing numerous key activities and responsibilities. Concurrently, skills are defined as the capabilities required to perform the tasks and obligations associated with a job. The national standards used for occupation classification are GB/T 6565-1999 (used in the 2000 Census) and GB/T 6565-2015 (used in the 2015 Mini-census). Occupation codes across difference years typically matched in four patterns: one-to-one, many-to-one, one-to-many, and many-to-many. When a 1999 occupation code corresponded to multiple 2015 codes, all the latter codes were converted to match the former's code. Conversely, when several 1999 codes corresponded to a single 2015 code, all the former codes were updated to the latter's code. For many-to-many matches, both sets of occupation classification standards were reviewed; the detailed descriptions associated with the codes in each standard were compared; flexible adjustments were made based on these descriptions. Predominantly, many-to-one and one-to-many relationships constitute the bulk of cases, with fewer many-to-many matches having a negligible effect on the overall process. To this end, following Zhou et al. (2023), 250 jobs and 13 occupational categories[1] were formed from occupational subcategories. Finally, occupational data were further aggregated at the city level to generate the preliminary processed occupational skill data and integrate them with industry data. Resultingly, a matrix containing 250 occupations and 277 prefecture-level cities was established. In addition to the above-mentioned occupation and industry data, some other socio-economic variables like gross domestic product (GDP) and population were obtained from the China City Statistical Yearbooks (CCSY) in corresponding years.

3.2 Measuring relatedness, cross-relatedness, and complexity

According to Balland et al. (2019), Hane-Weijman et al. (2022), and Li and Rigby (2023), relatedness and complexity can be calculated in four steps. First, the revealed comparative advantage (RCA) is calculated for each industry with the formula:

$$RCA_{i,c,t} = \frac{E_{i,c,t}/\sum_i E_{i,c,t}}{\sum_c E_{i,c,t}/\sum_{c,i} E_{i,c,t}} \tag{1}$$

By definition, RCA refers to another form of location quotient and examines whether a region has a higher share of employment in an industry than the national average. As suggested by Hidalgo et al. (2007), region *c* is specialized in industry *i* if its RCA is above one at time *t*.

$$x_{i,c,t} = \begin{cases} 1 & if\ RCA_{i,c,t} \geq 1 \\ 0 & otherwise \end{cases} \tag{2}$$

Second, the co-occurrence probability approach is adopted to measure the relatedness between industries *i* and *j* is measured. To be more specific, the number of instances where a region specializes in industry *i* (RCA>1) is counted, given that the region also specializes in industry *j* (*i*≠*j*). Then, the probability $P(x_{i,t}|x_{j,t})$ is obtained by dividing this number by the instances where industry *j* has an RCA. In a similar way, $P(x_{i,t}|x_{j,t})$ is calculated where a region specializes in industry *j* (RCA>1), concerning that the region also specializes in industry i. To further reconcile the asymmetric distance between industries, the minimum of each pair of conditional probabilities is calculated as below:

$$\phi_{i,j,t} = min\{P(x_{i,t}|x_{j,t}), P(x_{j,t}|x_{i,t})\} \tag{3}$$

Third, the density of industries around industry i in region c is derived from the sum of relatedness $\phi_{i,j,t}$ of industry i to all the other industries j where the region has an RCA by the sum of relatedness of industry i to all the other industries j in the reference region.

$$Density_{c,i,t} = \frac{\sum_j \phi_{i,j,t} x_{c,i,t}}{\sum_j \phi_{i,j,t}} \tag{4}$$

Finally, several approaches can be employed to measure economic complexity like the method of reflection (Hidalgo and Hausmann, 2009) and the fitness method (Tacchella *et al.*, 2012). Nevertheless, studies that have compared these metrics generally report high consistency in their results in regional contexts (Gao and Zhou, 2018; Fritz and Manduca, 2021). To this end, this study followed the method developed by Hidalgo and Hausmann (2009) and extended by Balland and Rigby (2017) to measure complexity. Specifically, the RCA matrix M is row standardized along with its transpose M^T to construct the index of economic complexity. The product matrix $B=M \times M^T$ is a square matrix with a dimension equal to the number of industries. The complexity index for each of these industries is given by the elements of the second eigenvector Q of matrix B, and elements are standardized as:

$$Complexity_{i,t} = \frac{Q - \bar{Q}}{\text{stdev}(Q)} \tag{5}$$

In addition to relatedness and complexity, occupation-industry cross-relatedness can be measured in two more steps. First, cross-relatedness between occupations and industries is described as the minimum conditional probability that a region develops industry i because it has some degree of comparative advantage in occupation *k*, and *vice versa*. This form was originally developed by Hidalgo et al. (2007) and further extended by Pugliese et al. (2019), Catalán et al. (2022), and Castaldi and Drivas (2023) with the form:

$$\phi^X_{k,i,t} = min\{P(x_{k,t}|x_{i,t}), P(x_{i,t}|x_{k,t})\} \tag{6}$$

By definition, $\phi^X_{k,i,t}$ ranges from zero to one. A value close to zero implies no co-occurrence between occupation *k* and industry *i*, whereas a value close to one indicates frequent co-occurrence between occupation *k* and industry *i*.

Second, the occupation-industry cross-relatedness density $CrossDensity_{c,i,t}$ corresponds to the average occupation-industry cross-relatedness of a potential new industry *i* in relation to the occupational skill structure of region *c* at time *t* with the form:

$$CrossDensity_{c,i,t} = \frac{\sum_k x_{c,k,t}\phi^X_{k,i,t}}{\sum_k \phi^X_{k,i,t}} \tag{7}$$

In a likewise manner, $CrossDensity_{c,i,t}$ ranges from zero and one. A value close to zero suggests that region *c* has a high degree of occupational development in areas close to industry *i* at time *t*. In other words, the current state of the labor market is characterized by a high degree of proximity to the industry to be developed.

3.3 Identifying the diversification trajectories of new industries

Based on the application of the relatedness and complexity framework (Balland *et al.*, 2019; Zhou and Hu, 2021; Qiao *et al.*, 2024), the cross-relatedness and complexity framework can be empirically established. This enabled researchers and policymakers to identify new industries that are either related or unrelated to the existing skill portfolio, and to assess whether the economic complexity of these industries increases or decreases over time.

To define whether a new industry is more or less related to the skill structure of a city, the city-level average cross-relatedness density of industries without an RCA is calculated and compared with the cross-relatedness density of an industry. The entry of an industry is classified as skill path-dependent if its cross-relatedness density exceeds the average cross-relatedness density of the city, but path-breaking if its cross-relatedness density falls below the average cross-relatedness density of the city. More formally, the average cross-relatedness density of a city is defined as:

$$CityCrossRelatedness_{c,t} = \frac{\sum_i (1 - RCA_{c,i,t})^{*} CrossDensity_{c,i,t}}{\sum_i (1 - RCA_{c,i,t})} \tag{8}$$

where $CityCrossRelatedness_{c,t}$ represents the cross-relatedness density of city *c* and $CrossDensity_{c,i,t}$ stands for the cross-relatedness density of industry *i* in city *c* at time *t*. Note that similar steps can be used to calculate the average relatedness density of industries without an RCA ($CityRelatedness_{c,t}$) and compare it with the relatedness density of an industry ($Density_{c,i,t}$) to define whether the industry is more or less related to the industrial structure of a city.

To define whether an industry entry has higher complexity, the average knowledge complexity of industries with an RCA in a city is calculated and compared with the knowledge complexity of an industry. Similarly, the entry of an industry is defined as high complexity if its knowledge complexity surpasses that of the city, but low complexity if its complexity is lower than that of the city. The knowledge complexity of a city is defined as follows:

$$CityComplexity_{c,t} = \frac{\sum_i RCA_{c,i,t} * Complexity_{i,t}}{\sum_i RCA_{c,i,t}} \tag{9}$$

where $CityComplexity_{c,t}$ represents the complexity of city *c* and $Complexity_{i,t}$ stands for the knowledge complexity of industry *i* at time *t*.

Based on these, four industry diversification trajectories depicted in Figure 2 can be classified. All the new industries can be categorized into one of these diversification trajectories. Given that occupation-industry cross-relatedness measures the proximity between occupation structure and industrial activities, this classification framework uncovers the role of skill capabilities in regional diversification. For this analysis, the diversification trajectories of new industries that lacked comparative advantages at time t−1 but gained them by time *t* and the spatial distribution of these new industries were specifically examined.

3.4 Econometric model

To examine the relationship between skill capabilities and regional industry evolution, the following model is considered:

$$
\begin{aligned}
x_{j,c,t} = \alpha_0 + \alpha_1 \times x_{j,c,t-1} + \alpha_2 \times CrossDensity_{j,c,t-1} \\
+ \alpha_3 \times Complexity_{j,t-1} + \alpha_4 \times Industry_{j,t-1} \\
+ \alpha_5 \times Region_{c,t-1} + \gamma_j + \varphi_c + \varepsilon_{j,c,t-1}
\end{aligned}
\tag{10}
$$

where $x_{j,c,t}$ is 1 if industry *j* in region *c* has an RCA at time *t*; $CrossDensity_{c,j,t-1}$ indicates the occupation-industry cross-relatedness density (hereinafter referred to as cross-density) of region *c* at time *t*-1; $Complexity_{j,t-1}$ stands for the economic complexity of industry *j* at time *t*-1. $Region_{c,t-1}$ denotes three control variables for regions, including the size of the region (population), the level of economic development (per capita GDP), regional agglomeration economies (logged population density), and industrial stock (the number of specialized industries in a region). $Industry_{j,t-1}$ indicates one control variable for industries, i.e., industrial size (the number of cities specializing in a given industry). Furthermore, the fixed effects of regions (γ_j) and industries (φ_c) are considered. To compare the effects of occupation-industry cross-relatedness and industry relatedness on industry evolution, the cross-density variable is replaced with the industry relatedness density variable (hereinafter referred to as density). Tables 1 and A1 illustrate the description of all the variables and the descriptive statistics, respectively.

[INSERT TABLE 1]

As a well-known Chinese proverb goes, "It takes ten years to grow trees, but a hundred to rear people", human capital accumulation is a "patient" process that unfolds over time (Yang, 2023). Thus, this study set its timeframe from 2000 to 2015 to adequately reflect this gradual development. Nonetheless, empirical evidence generally set a five-year period in technological and industrial diversification studies (Zhu *et al.*, 2017; Balland *et al.*, 2019). This research also examined the effect of skill capabilities on regional diversification in shorter periods, including the five-year from 2000 to 2005 and the ten-year period from 2000 to 2010, when checking the robustness of the estimation results[2].

As for the modeling method, Equation 10 is estimated via a linear probability model due to a large number of fixed effects. By comparison, nonlinear models like logit and probit models may give rise to biased estimates in the presence of many zeros in dependent variables (Gomila, 2021; Castaldi and Drivas, 2023). Nevertheless, the results were also re-estimated using non-linear models as a robustness check. Following Castaldi and Drivas (2023), the dummy variables were dropped. This is because the estimation results of non-linear models might be biased and inconsistent when the regression includes a number of fixed effects. Additionally, the variance inflation factor (VIF) values are all below five, and as indicated by the correlation matrix in Table A2, collinearity is not a concern.

4 Empirical results

4.1 Descriptive analysis

The evolution of cross-space and some network statistics are demonstrated in Figure 3. Note that the average degree refers to the ratio of twice the number of edges to the number of nodes. For visualization, edges with a weight greater than 0.40 were included. The number of nodes, edges, and average degree increased drastically, and more industry nodes were included, which indicated that more strong links between occupations and industries were established in 2015. Meanwhile, two concrete examples were further included to display the concept of occupation-industry cross-relatedness. In Figure 3a, the node representing personnel in glass and glass product manufacturing is closely related to the nodes of glass fiber, glass product, and glass manufacturing industries, which indicates that the skills and knowledge of glass and glass product manufacturing personnel can be applied in these three glass-related manufacturing industries. Similarly, Figure 3b suggests that personnel in steel refining can work across several sectors, such as iron smelting, steel refining, and ferroalloy smelting.

[INSERT FIGURE 3]

The maps of average relatedness and cross-relatedness densities of cities are presented in Figures 4a and 4b. High consistency can be found in the spatial distribution of these two density variables. Large cities in the Yangtze River Delta, the Pearl River Delta, and the

Beijing-Tianjin-Hebei urban agglomerations have high levels of relatedness and cross-relatedness. This distinction supports the argument that industrial and occupational structures in large cities tend to be better aligned (Mion and Naticchioni, 2009). This enhanced match is possibly due to the greater diversity and specialization of industries in larger urban areas, which can house a wider range of occupations and more specific skill sets. The relationship between relatedness and cross-relatedness across different population sizes of regions is further illustrated in Figure B1. These two variables move in the same direction. Medium and small regions have low levels of relatedness and cross-relatedness, whereas large regions seem to have high levels of relatedness and cross-relatedness.

[INSERT FIGURE 4]

The spatial distributions of new industries are denoted in Figure 5a. The southwest region demonstrated high frequencies of industry entry, followed by the southeast and northeast regions. In contrast, the spatial distributions of new industries in terms of relatedness and cross-relatedness are depicted in Figures 5b and 5c, respectively. Consistent with existing literature (Zhu *et al.*, 2017; He *et al.*, 2018), path-breaking industries emerged mainly in underdeveloped western regions. By comparison, regions in the northwest and southwest like Kunming and Lanzhou showed high frequencies of skill-unrelated diversification and jumped from their current skill portfolios. This can possibly be explained by the re-skilling of workforce development via the external recruitment of labor and skills, formal education, job training, etc. In addition, the number of new industries for different types of diversification trajectories is listed in Table A3 from the perspectives of relatedness, cross-relatedness and complexity. Industry-based related diversification accounts for 90% of total industry entries, while skill-based related diversification accounts for 74%. Both confirm the path-dependent nature of industrial diversification (Boschma, 2017; Pinheiro *et al.*, 2022).

[INSERT FIGURE 5]

The practical integration of cross-relatedness/complexity and relatedness/complexity frameworks for designing place-based skill policies is demonstrated in Figure 6. Specifically, Figures 6a-6c depict the cross-relatedness/complexity space for three cities—Beijing, Dalian and Xuzhou in 2000, while Figures 6d-6f illustrate the corresponding relatedness/complexity space. In these figures, the biopharmaceutical manufacturing sector is highlighted in red for illustration. The results from Figures 6a and 6d suggest that developing or maintaining the revealed comparative advantage in this high-complexity sector is viable and strategically advantageous for Beijing. By comparison, Dalian, a coastal city in eastern China, possesses the requisite skill capabilities to support the biopharmaceutical sector, but its industrial portfolios are unrelated, as shown in Figures 6b and 6e. Therefore, policies should take advantage of the local labor structure to nurture this sector. In contrast, Xuzhou, a resource-based city in eastern China, lacks the necessary skills for the biopharmaceutical sector, as depicted in Figures 6c and 6f. Consequently, industrial development should prioritize adjusting occupational composition possibly through skill-enhancing strategies like in-house training and external recruitment.

[INSERT FIGURE 6]

4.2 Regression analysis

The results of baseline regressions are displayed in Table 2. Note that Models 1-4 focus on industry relatedness, whereas Models 5-8 focus on occupation-industry cross-relatedness.

Consistent with Boschma et al. (2013) and Zhu et al. (2017), the parameter of $x_{j,c,t-1}$ is significant and positive in all models, which means that having a comparative advantage at the beginning of the study period increases the probability of having a comparative advantage at the end of the period. Moreover, it is evident that both *Density* and *Cross-density* are significant in Models 1-8. On the one hand, *Density* has a positive effect on having a comparative advantage in an industry, which indicates that regional diversification is associated with the existing industry structure of a region. This echoes the current understanding of the path dependency of regional diversification in China (He *et al.*, 2018; Qiao *et al.*, 2024). On the other hand, the interpretation of the cross-density variable is that regional diversification is linked to its skill structure as well. To this end, diversifying into skill-related industries can be a potential way of developing new growth paths in the form of related diversification. In terms of magnitude, the effect of *Density* in Models 1-4 is greater than that of *Cross-density* in Models 5-8. Every 10% increase in *Density* is related to an increase of 2.0% ([0.871×10%]/0.043) in the probability of forming industrial comparative advantages in Model 4. By comparison, every 10% increase in *Cross-density* is correlated with an increase of 0.8% ([0.159×10%]/0.021) in the probability of forming industrial comparative advantages in Model 8. In addition, *Complexity* in all models except Model 4 is positive and significant, which suggests that complexity at large contributes to a comparative advantage in an industry.

[INSERT TABLE 2]

For the control variables, *Population*, *GDP per capita*, and *Size* consistently show positive and significant effects, indicating a positive correlation between maintaining or achieving

RCA and the region's population size, economic development, and industry scale. However, unlike *Density* and *Cross-density*, the effect of agglomeration, as indicated by *Population Density*, is consistently negative, suggesting that regions with greater agglomeration are less likely to have specialized industries. By comparison, *Stock* is negative in Models 3-4 but positive in Models 7-8, possibly due to the impacts from the relatedness and cross-relatedness variables. As reflected by the adjusted R^2 in Models 1-8, adding controls and fixed effects can increase the explanatory power of variables, but replacing the density variable with the cross-density variable results in a slight decrease in the adjusted R^2.

The original sample was further divided into four subsamples based on medians, and the heterogeneity of the empirical results is checked in Table 3. For brevity, only full models with control variables and fixed effects were included. The cross-density variable is always significant and positive in all models with the exception of Model 2, which unveils the significance of skill capabilities in driving industrial evolution. Conversely, the cross-density variable is negative and significant in Model 2. The interpretation is that skill structure can hardly contribute to industrial diversification when occupation-industry cross-relatedness is low. Under this circumstance, regional diversification has to follow a path-breaking process in terms of skill capabilities given that labor skills are essential inputs for industrial activities. At the same time, the complexity variable is only significant and positive in Models 1 and 4. These results are not in line with the full sample, which reveals that complexity contributes to regional industrial evolution when the relatedness between a potential industry and an existing skill structure is relatively high and industry complexity is low.

[INSERT TABLE 3]

To test the robustness of the results, the following estimations were made. First, the original study period was changed to five and ten years, covering 2000–2005 and 2000–2010, respectively. Although human capital accumulation requires time (Yang, 2023), the effect of skill capabilities on industrial evolution concerning shorter time periods was examined. Second, alternative values of threshold (0.8 and 1.2) were utilized to determine a regional comparative advantage. Third, the results of the baseline regressions were re-estimated by using non-linear models, including logit and probit models. However, nonlinear models can encounter incidental parameter problems when equations include a number of fixed effects, which can result in biased and inconsistent coefficients, especially when standard errors are clustered at the region-class level (Boschma *et al.*, 2013; Gomila, 2021). Because of this, following Castaldi and Drivas (2023), dummy variables in logit and probit models were dropped. Since the re-estimated results were not significantly different from the baseline regressions, these results were relocated to the appendix (Table A4).

5 Concluding remarks

This paper introduced the concept of occupation-industry cross-relatedness, which diverged from existing studies that primarily utilize the linking or embedding approach to integrating skill and industry portfolios (Neffke and Henning, 2013; Buyukyazici *et al.*, 2024). This new concept incorporates the interactions between occupations and industries and draws on established research on industry relatedness, occupation relatedness and technology-science cross-relatedness (Boschma, 2017). The current study utilized this concept, and further isolated and evaluated the impact of skill capabilities on the industrial evolution of Chinese regions from 2000 to 2015. Meanwhile, based on the relatedness-complexity space

(Balland *et al.*, 2019; Hidalgo, 2023), a cross-relatedness and complexity space in the nexus of skills and industries was established. This allowed researchers and policymakers to not only discern which new industries are related or unrelated to the existing skill portfolio but also evaluate whether their economic complexity increases or decreases. The results of this analysis suggest that industry evolution is associated with the skill portfolio of a region, which extends the current understanding of regional capabilities and cross-relatedness. In addition, the impacts of skills and human capital on regional industry evolution are insignificant when industries have low levels of relatedness to skills, which emphasizes skill-industry matching in economic activities (Mion and Naticchioni, 2009; Holm *et al.*, 2017).

This study underscored several crucial policy implications. First, the findings of this study suggest that the skill capabilities hidden behind industry portfolios may serve as pivotal sources for regional diversification paths. If a potential industry is aligned with the local skill structure, this new industry is likely to emerge. Conversely, skills and human capital are less likely to play a significant role in the evolution of a potential industry if a disconnect exists between the industry and the existing skill portfolio. Second, skill-related diversification is common, but this study acknowledges that skill path-breaking diversification can also occur, especially in underdeveloped regions with limited access to diverse and complex skills. Under these circumstances, policy interventions becomes essential (Buyukyazici, 2023). Third, the framework of cross-relatedness and complexity can be effectively integrated with the existing relatedness/complexity framework (Balland *et al.*, 2019; Hidalgo, 2023), serving as a comprehensive analytical tool to identify place-specific deficiencies and imbalances in skills for regional development. By utilizing these frameworks, both endogenous and exogenous measures can be implemented to enhance regional capabilities. For endogenous measures, such as job training, the focus is on building new skills primarily using existing resources. In contrast, exogenous measures involve introducing non-local actors, such as through the external recruitment of star scientists and top managers, to foster industrial evolution.

While this paper offers valuable insights into regional diversification, future research should explore a few important avenues. First of all, future studies could examine the distinct impacts of high-, medium-, and low-skilled occupations on regional industrial diversification (Shen and Liu, 2016; Sunley *et al.*, 2020). This differentiation is critical because, unlike industries, the occupational landscape is highly polarized, characterized by a growing proportion of high- and low-skill occupations alongside a relative decline in medium-skill jobs (Alabdulkareem *et al.*, 2018; Xu *et al.*, 2021). Second, most of the current literature focuses on one or two of complexity, relatedness, and cross-relatedness (Balland *et al.*, 2019; Hane-Weijman *et al.*, 2022), which is far away from a comprehensive framework. Thus, it is necessary to consider the interplay of these three dimensions in a unified manner. One possible way is to integrate the frameworks of self-relatedness/cross-relatedness (Castaldi and Drivas, 2023) and relatedness/complexity (Balland *et al.*, 2019; Hidalgo, 2023) with the cross-relatedness and complexity space of this research. Third, the concept of cross-relatedness has been used for studying the interface between scientific fields and technological patents (Balland and Boschma, 2022; Catalán *et al.*, 2022) and between occupations and industries in this analysis. Future research should pay more attention to other forms of crossing phenomena and explore the applicability of cross-relatedness in explaining regional diversification across varying contexts.

6 Endnotes

1. Military personnel were excluded. To illustrate, 13 occupational categories included managers, technicians, professionals, office personnel, commercial sales workers, mechanical and electrical assemblers, production and operation workers, mining and construction workers, transportation operators, safety and security workers, delivery service workers, residential service workers, and agricultural production workers.
2. This choice can be attributed to the issue of data availability. Industries were continuously recorded, but the occupation data and classification standards might be different across different years. The 2005 occupation data only have 78 occupation groups across 341 cities, and the 2010 occupation data have 407 occupation categories across 102 cities. If a consistent occupation classification system was established across 2000, 2005, 2010 and 2015, occupation information might be lost greatly when minor occupation groups were aggregated.

7 References

Ahn, J., Khandelwal, A.K. and Wei, S.-J. 2011. The role of intermediaries in facilitating trade. *Journal of International Economics* 84: 73–85.

Alabdulkareem, A., Frank, M.R., Sun, L., AlShebli, B., Hidalgo, C. and Rahwan, I. 2018. Unpacking the polarization of workplace skills. *Science Advances* 4: eaao6030.

Balland, P.-A. and Rigby, D. 2017. The Geography of Complex Knowledge. *Economic Geography* 93: 1–23.

Balland, P.-A. and Boschma, R. 2022. Do scientific capabilities in specific domains matter for technological diversification in European regions? *Research Policy* 51: 104594.

Balland, P.-A., Boschma, R., Crespo, J. and Rigby, D.L. 2019. Smart Specialization Policy in the European Union: Relatedness, Knowledge Complexity and Regional Diversification. *Regional Studies* 53: 1252–1268.

Barbour, E. and Markusen, A. 2007. Regional occupational and industrial structure: does one imply the other? *International Regional Science Review* 30: 72–90.

Boschma, Minondo, A. and Navarro, M. 2013. The Emergence of New Industries at the Regional Level in Spain: A Proximity Approach Based on Product Relatedness. *Economic Geography* 89: 29–51.

Boschma, R. 2017. Relatedness as Driver of Regional Diversification: A Research Agenda. *Regional Studies* 51: 351–364.

Buyukyazici, D. 2023. Skills for smart specialisation: Relatedness, complexity and evaluation of priorities. *Papers in Regional Science* 102: 1007–1030.

Buyukyazici, D., Mazzoni, L., Riccaboni, M. and Serti, F. 2024. Workplace skills as regional capabilities: relatedness, complexity and industrial diversification of regions. *Regional Studies* 58: 469–489.

Cappelli, R., Boschma, R. and Weterings, A. 2019. Labour mobility, skill-relatedness and new plant survival across different development stages of an industry. *Environment and Planning A: Economy and Space* 51: 869–890.

Castaldi, C. and Drivas, K. 2023. Relatedness, Cross-relatedness and Regional Innovation Specializations: An Analysis of Technology, Design, and Market Activities in Europe and the US. *Economic Geography* 99: 253–284.

Catalán, P., Navarrete, C. and Figueroa, F. 2022. The scientific and technological cross-space: Is technological diversification driven by scientific endogenous capacity? *Research Policy* 51: 104016.

Chen, J., Li, X. and Zhu, Y. 2024. Shock absorber and shock diffuser: the multiple roles of industrial diversity in shaping regional economic resilience after the Great Recession. *The Annals of Regional Science* 72: 1015–1045.

Corradini, C., Morris, D. and Vanino, E. 2023. Towards a regional approach for skills policy. *Regional Studies* 57: 1043–1054.

Farinha, T., Balland, P.-A., Morrison, A. and Boschma, R. 2019. What drives the geography of jobs in the US? Unpacking relatedness. *Industry and Innovation* 26: 988–1022.

Frenken, K., Van Oort, F. and Verburg, T. 2007. Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies* 5: 685–697.

Fritz, B.S.L. and Manduca, R.A. 2021. The economic complexity of US metropolitan areas. *Regional Studies* 55: 1299–1310.

Galetti, J.R.B., Tessarin, M.S. and Morceiro, P.C. 2021. Skill relatedness, structural change and heterogeneous regions: evidence from a developing country. *Papers in Regional Science* 100: 1355–1376.

Gao, J. and Zhou, T. 2018. Quantifying China's regional economic complexity. *Physica A: Statistical Mechanics and Its Applications* 492: 1591–1603.

Gomila, R. 2021. Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General* 150: 700–709.

Hane-Weijman, E., Eriksson, R.H. and Rigby, D. 2022. How do occupational relatedness and complexity condition employment dynamics in periods of growth and recession? *Regional Studies* 7: 1176–1189.

He, C., Yan, Y. and Rigby, D. 2018. Regional Industrial Evolution in China. *Papers in Regional Science* 97: 173–198.

Hidalgo, C.A. 2023. The policy implications of economic complexity. *Research Policy* 52: 104863.

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

Hidalgo, C.A., Klinger, B., Barabási, A. and Hausmann, R. 2007. The product space conditions the development of nations. *Science* 317: 482–487.

Holm, J.R., Østergaard, C.R. and Olesen, T.R. 2017. Destruction and reallocation of skills following large company closures. *Journal of Regional Science* 57: 245–265.

Hu, X. and Hassink, R. 2017. Exploring adaptation and adaptability in uneven economic resilience: A tale of two Chinese mining regions. *Cambridge Journal of Regions, Economy and Society* 10: 527–541.

Jara-Figueroa, C., Jun, B., Glaeser, E.L. and Hidalgo, C.A. 2018. The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences* 115: 12646–12653.

Kogler, D., Rigby, D. and Tucker, I. 2013. Mapping knowledge space and technological relatedness

in US cities. *European Planning Studies* 21: 1374–1391.
Li, Y. and Rigby, D. 2023. Relatedness, Complexity, and Economic Growth in Chinese Cities. *International Regional Science Review* 46: 3–37.
Liu, Y., Shen, J., Xu, W. and Wang, G. 2017. From school to university to work: migration of highly educated youths in China. *The Annals of Regional Science* 59: 651–676.
Maskell, P. and Malmberg, A. 1999. Localised learning and industrial competitiveness. *Cambridge Journal of Economics* 23: 167–185.
Mion, G. and Naticchioni, P. 2009. The spatial sorting and matching of skills and firms. *Canadian Journal of Economics/Revue Canadienne d'économique* 42: 28–55.
Muneepeerakul, R., Lobo, J., Shutters, S.T., Goméz-Liévano, A. and Qubbaj, M.R. 2013. Urban Economies and Occupation Space: Can They Get "There" from "Here"? *PLoS ONE* 8: e73676-8.
Neffke, F. and Henning, M. 2008. Revealed relatedness: Mapping industry space. .
Neffke, F. and Henning, M. 2013. Skill Relatedness and Firm Diversification. *Strategic Management Journal* 34: 297–316.
Neffke, F., Henning, M. and Boschma, R.A. 2011. How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography* 87: 237–265.
Penrose, E. 1959. *The Theory of the Growth of the Firm*. Oxford University Press, .
Pinheiro, F.L., Hartmann, D., Boschma, R. and Hidalgo, C.A. 2022. The time and frequency of unrelated diversification. *Research Policy* 51: 104323.
Pugliese, E., Cimini, G., Patelli, A., Zaccaria, A., Pietronero, L. and Gabrielli, A. 2019. Unfolding the innovation system for the development of countries: coevolution of Science, Technology and Production. *Scientific Reports* 9: 16440.
Qiao, Y., Ascani, A. and Morrison, A. 2024. External linkages and regional diversification in China: The role of foreign multinational enterprises. *Environment and Planning A: Economy and Space* 0308518X231223621.
Rey, S.J. 1998. The Performance of Alternative Integration Strategies for Combining Regional Econometric and Input-Output Models. *International Regional Science Review* 21: 1–35.
Romer, P. 1990. Endogenous technological change. *Journal of Political Economy* 98: 71–102.
Schultz, T.W. 1961. Investment in Human Capital. *The American Economic Review* 51: 1–17.
Shen, J. and Liu, Y. 2016. Skilled and less-skilled interregional migration in China: A comparative analysis of spatial patterns and the decision to migrate in 2000–2005. *Habitat International* 57: 1–10.
Sunley, P., Martin, R., Gardiner, B. and Pike, A. 2020. In search of the skilled city: Skills and the occupational evolution of British cities. *Urban Studies* 57: 109–133.
Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. and Pietronero, L. 2012. A New Metrics for Countries' Fitness and Products' Complexity. *Scientific Reports* 2: 723.
Thompson, W.R. and Thompson, P.R. 1987. National Industries and Local Occupational Strengths: the Cross-Hairs of Targeting. *Urban Studies* 24: 547–560.
Whittle, A. and Kogler, D.F. 2020. Related to What? Reviewing the Literature on Technological Relatedness: Where We Are Now and Where Can We Go? *Papers in Regional Science* 99: 97–113.
Wixe, S. and Andersson, M. 2017. Which types of relatedness matter in regional growth? Industry,

occupation and education. *Regional Studies* 51: 523–536.

Xu, W., Qin, X., Li, X., Chen, H., Frank, M., Rutherford, A., Reeson, A. and Rahwan, I. 2021. Developing China's workforce skill taxonomy reveals extent of labor market polarization. *Humanities and Social Sciences Communications* 8: 1–10.

Yang, Z. 2023. Human capital space: a spatial perspective of the dynamics of people and economic relationships. *Humanities and Social Sciences Communications* 10: 1–14.

Zhou, X., Ao, R., Aihemaitijiang, Y., Chen, J. and Tang, H. 2023. Influence of skill relatedness on the location choice of heterogeneous labor force in Chinese prefecture-level cities. *PLOS ONE* 18: e0289803.

Zhou, Y. and Hu, C. 2021. Industrial policy and differentiated regional diversifications: Evidence from Chinese cities. *Cities* 119: 103348.

Zhu, S., He, C. and Zhou, Y. 2017. How to jump further and catch up? Path-breaking in an uneven industry space. *Journal of Economic Geography* 17: 521–545.

8 Appendices

[INSERT TABLES A1-A4]

[INSERT FIGURE B1]

9 Figures and tables

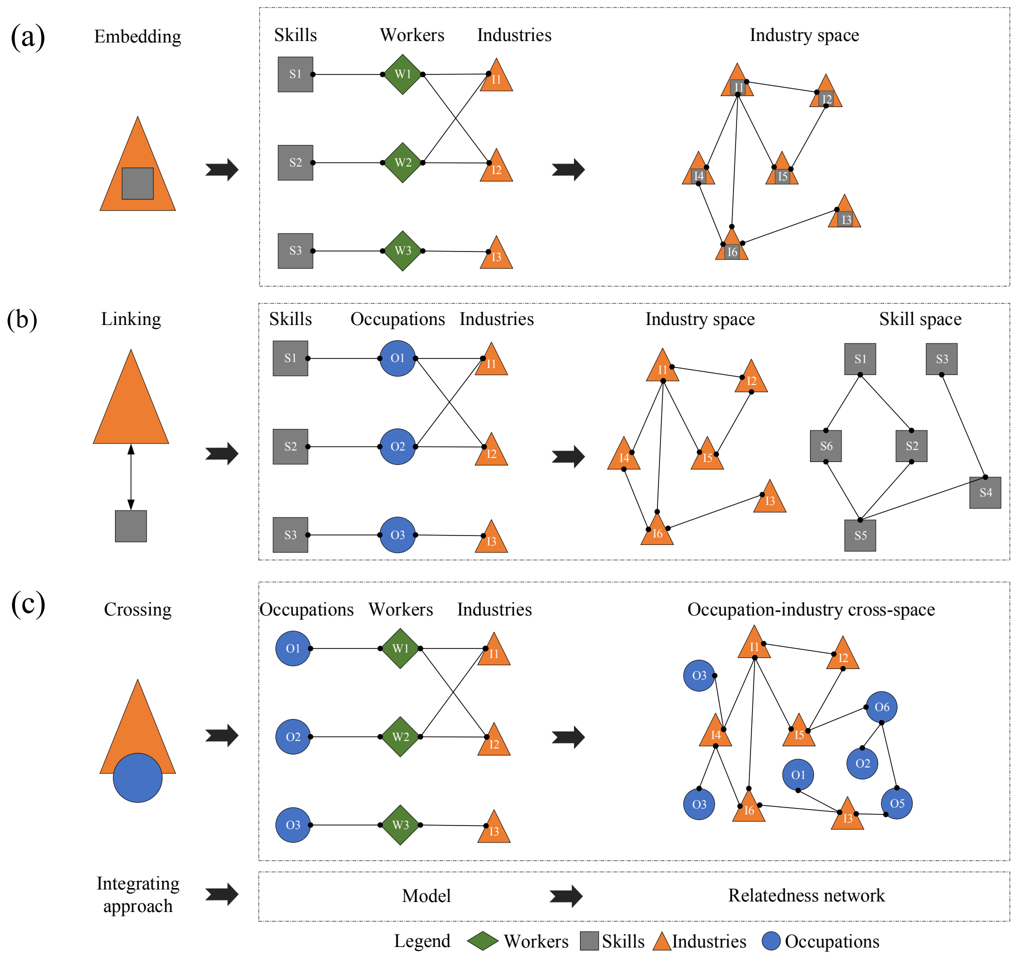

Figure 1. Three approaches to integrating industry and skill portfolios

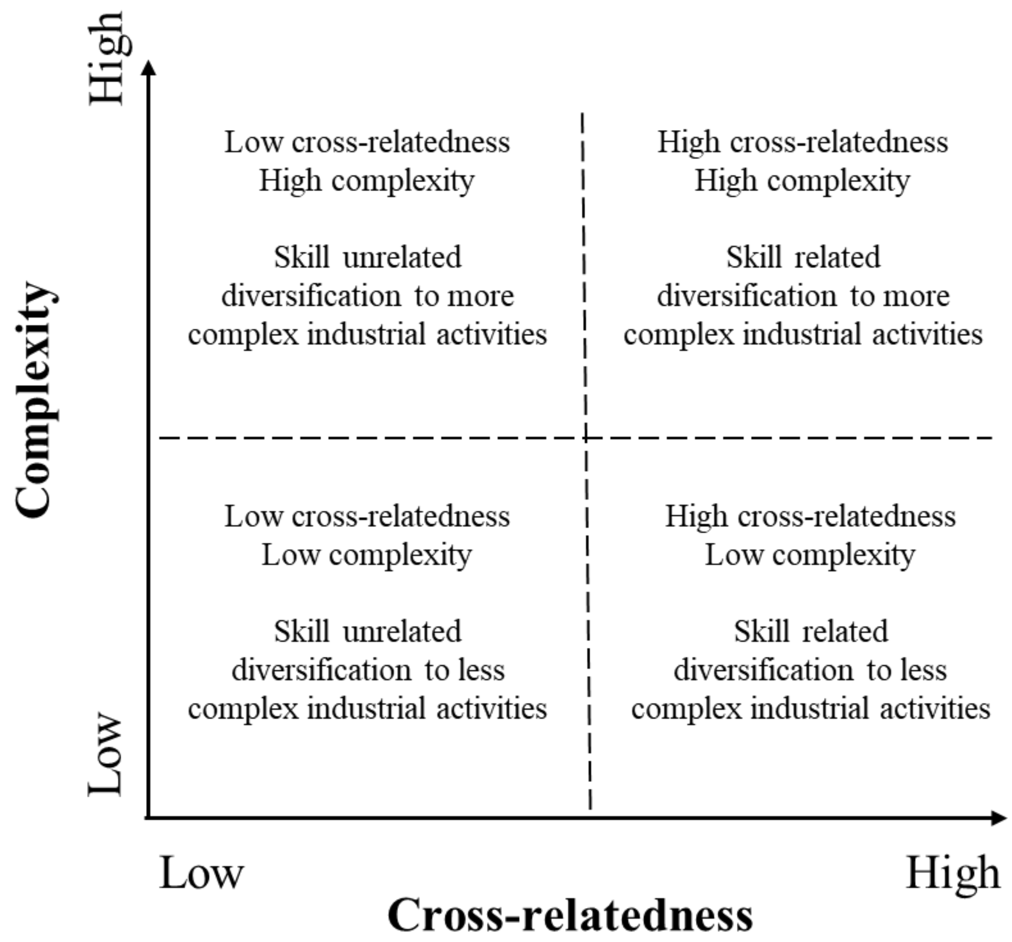

Figure 2. A theoretical framework of cross-relatedness and complexity between industries and occupations

Figure 3. Occupation-industry cross-space, 2000 (a) and 2015 (b)

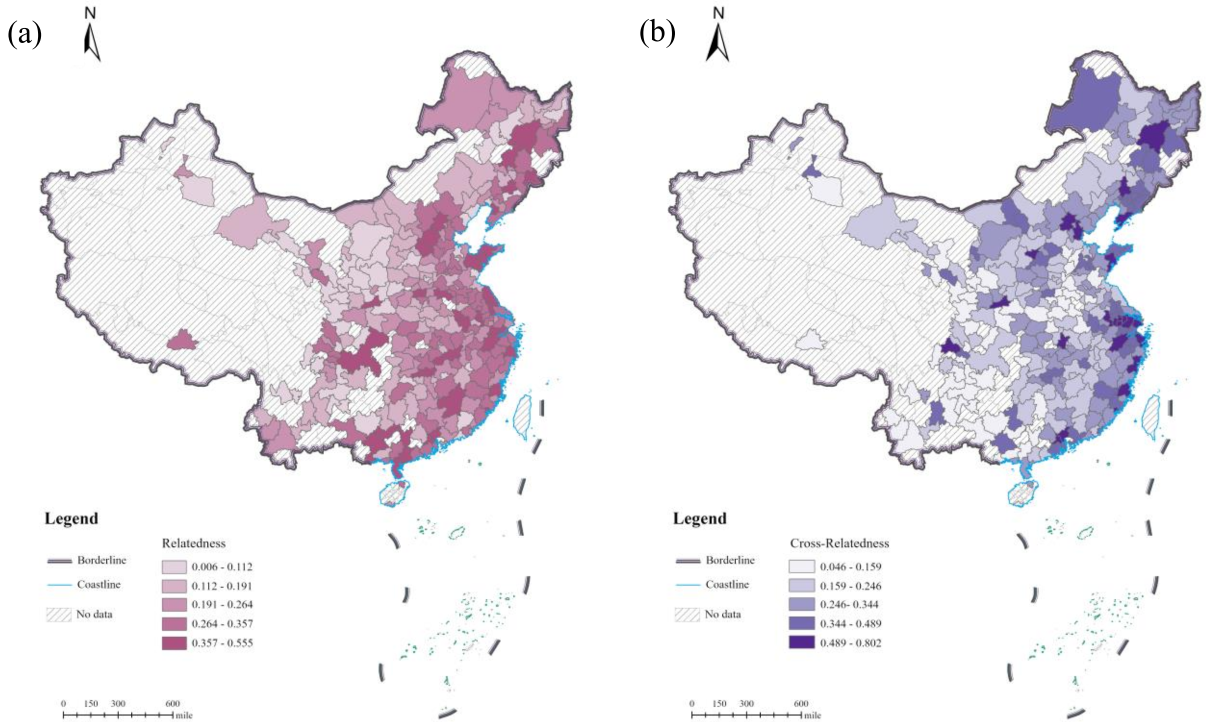

Figure 4. The spatial distributions of relatedness (a) and cross-relatedness (b) densities

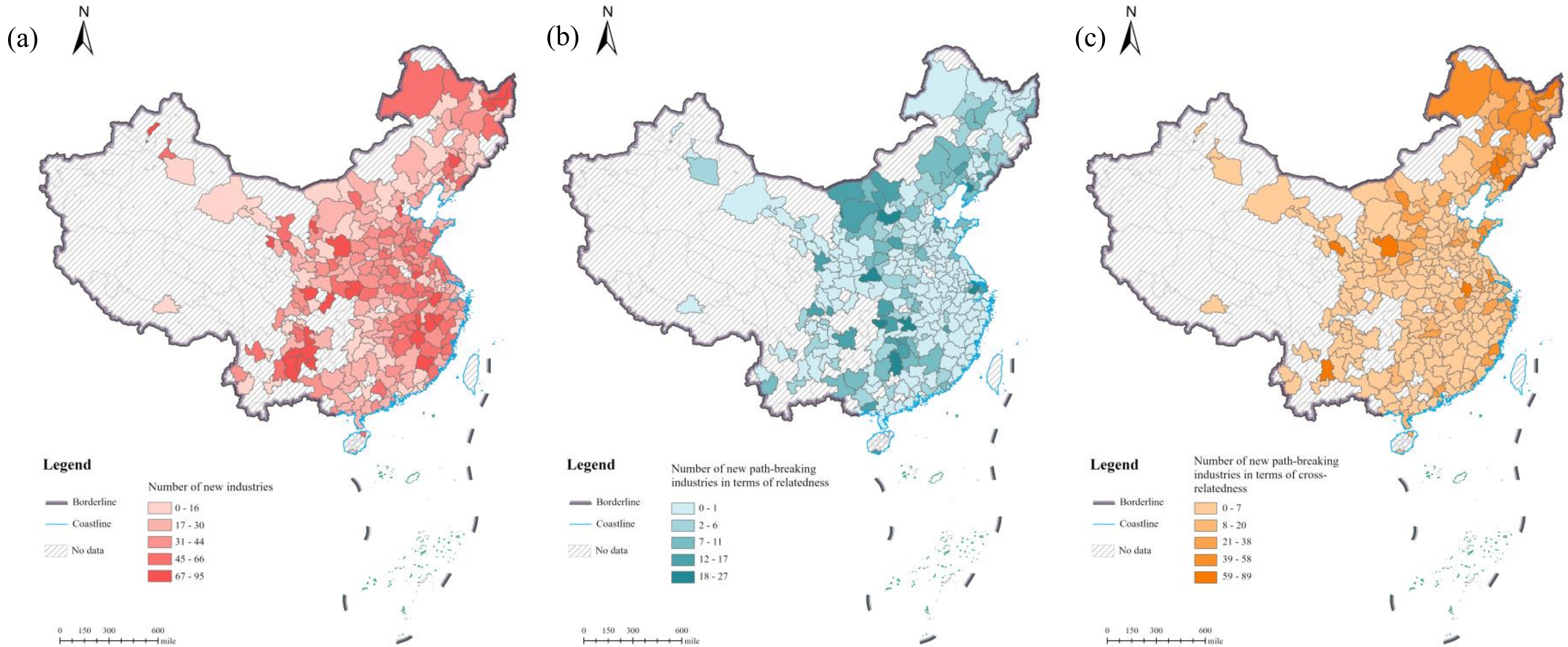

Figure 5. The spatial distributions of industry entry (a), path-breaking industry entry (b), and skill path-breaking industry entry (c)

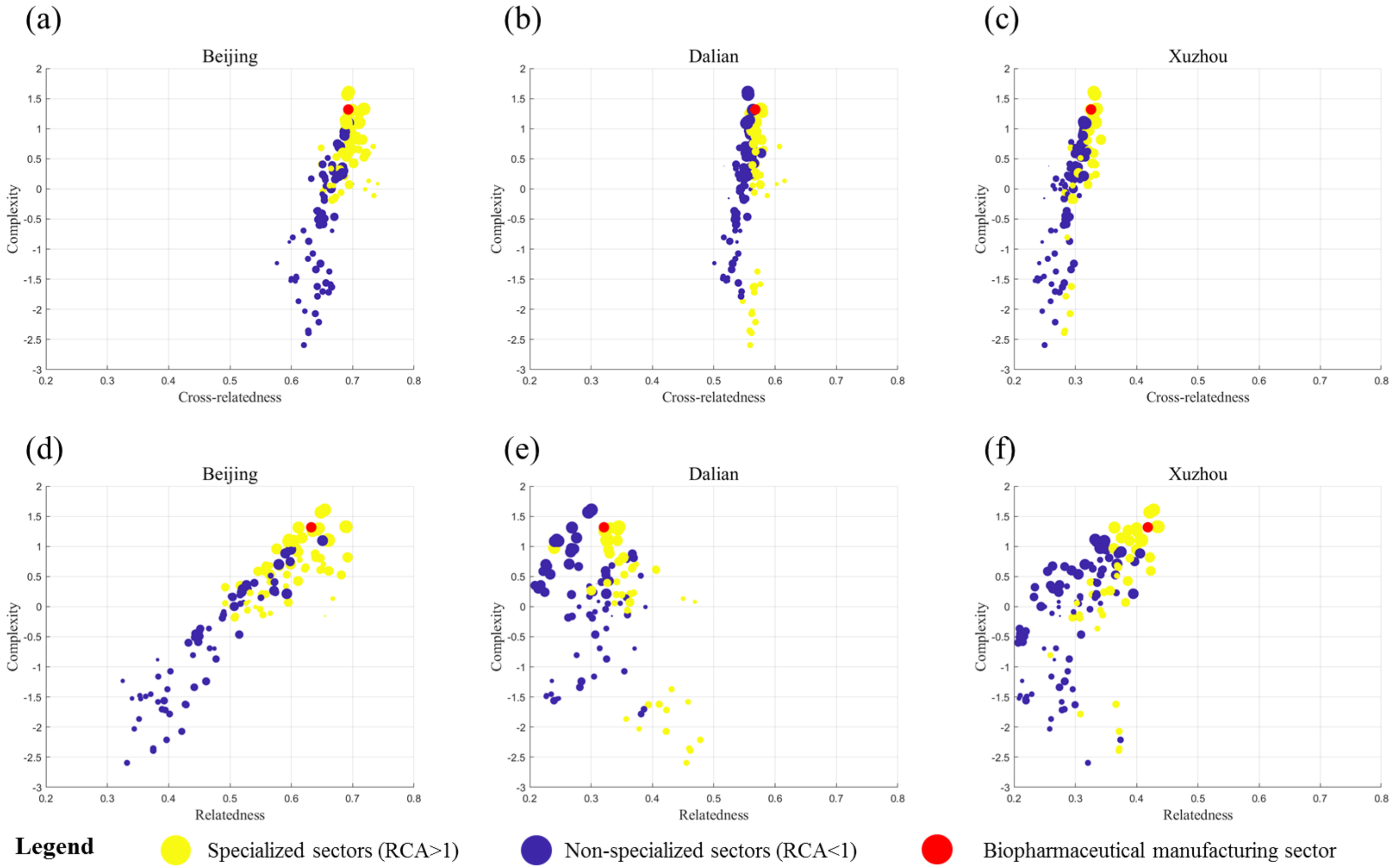

Figure 6. Integrating the cross-relatedness/complexity framework and the relatedness/complexity framework for place-based skills policies

Table 1. Description of variables

Variables	Definition	Source
$x_{j,c,t}$	Dummy variable identifying the existence of specialized industry *j* at time *t* in region *c*	Chinese Customs Trade Statistics
Density	Density of industry relatedness that industry *j* reveals with respect to all the industries observed at time *t*-1 in region *c*	Chinese Customs Trade Statistics
CrossDensity	Density of occupation-industry cross-relatedness that industry *j* reveals with respect to all the occupations observed at time *t*-1 in region *c*	Chinese Customs Trade Statistics, the 2000 Chinese Census, and the 2015 Chinese 1% Population Sampling Survey
Complexity	Economic complexity of industry *j* observed at time *t*-1	Chinese Customs Trade Statistics
Population	Population of region *c* at time *t*-1	China City Statistical Yearbooks
GDP per capita	GDP per capita of region *c* at time *t*-1	China City Statistical Yearbooks
Population density	Logged population density of region *c* at time *t*-1	China City Statistical Yearbooks
Stock	Number of specialized industries in region *c* at time *t*-1	Chinese Customs Trade Statistics
Size	Number of cities with specialized industry *j* at time *t*-1	Chinese Customs Trade Statistics

Table 2. Baseline regressions

	Relatedness				Cross-relatedness			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
x00	0.277^{***}	0.253^{***}	0.181^{***}	0.180^{***}	0.343^{***}	0.311^{***}	0.261^{***}	0.261^{***}
	(0.006)	(0.006)	(0.007)	(0.007)	(0.005)	(0.005)	(0.006)	(0.006)
Density	0.413^{***}	0.367^{***}	0.861^{***}	0.871^{***}				
	(0.016)	(0.016)	(0.043)	(0.043)				
Cross-density					0.193^{***}	0.166^{***}	0.142^{***}	0.159^{***}
					(0.015)	(0.015)	(0.020)	(0.021)
Complexity		0.068^{***}	0.014^{***}	0.004		0.071^{***}	0.023^{***}	0.013^{***}
		(0.002)	(0.002)	(0.003)		(0.002)	(0.002)	(0.003)
Population			0.167^{***}	0.169^{***}			0.102^{***}	0.102^{***}
			(0.023)	(0.023)			(0.024)	(0.024)
Population density			-0.192^{***}	-0.193^{***}			-0.232^{***}	-0.236^{***}
			(0.015)	(0.015)			(0.016)	(0.016)
GDP per capita			0.019^{***}	0.020^{***}			0.008^{**}	0.008^{***}
			(0.003)	(0.003)			(0.003)	(0.003)
Stock			-3.452^{***}	-3.488^{***}			1.342^{***}	1.359^{***}
			(0.261)	(0.262)			(0.114)	(0.114)
Size			1.959^{***}	2.023^{***}			2.226^{***}	2.282^{***}
			(0.073)	(0.074)			(0.073)	(0.074)
Industry FE	No	No	No	Yes	No	No	No	Yes
Region FE	No	No	No	Yes	No	No	No	Yes
Constant	0.162^{***}	0.180^{***}	0.114^{***}	0.134^{***}	0.199^{***}	0.214^{***}	0.057^{***}	0.065^{***}
	(0.004)	(0.004)	(0.008)	(0.009)	(0.004)	(0.004)	(0.009)	(0.011)
N	48,752	48,752	48,752	48,752	48,752	48,752	48,752	48,752
Adjusted R^2	0.116	0.136	0.157	0.158	0.107	0.129	0.151	0.152

Notes: Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. Heterogeneity analysis for different levels of cross-relatedness and complexity

	(1)	(2)	(3)	(4)
	High cross-relatedness	**Low cross-relatedness**	**High complexity**	**Low complexity**
x00	0.274***	0.237***	0.272***	0.236***
	(0.007)	(0.009)	(0.007)	(0.009)
Cross-density	0.251***	-0.238***	0.109***	0.191***
	(0.040)	(0.058)	(0.030)	(0.030)
Complexity	0.025***	0.004	-0.042***	0.037***
	(0.004)	(0.004)	(0.012)	(0.004)
Population	0.056	0.341***	0.165***	0.040
	(0.035)	(0.039)	(0.034)	(0.033)
Population density	-0.300***	0.108***	-0.334***	-0.129***
	(0.019)	(0.037)	(0.022)	(0.023)
GDP per capita	0.003	0.027***	0.005	0.014***
	(0.005)	(0.005)	(0.005)	(0.004)
Stock	1.242***	0.726***	1.152***	1.605***
	(0.167)	(0.163)	(0.172)	(0.149)
Size	1.628***	3.003***	2.520***	3.794***
	(0.109)	(0.101)	(0.125)	(0.164)
Industry FE	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Constant	0.052***	0.149***	0.145***	-0.032**
	(0.019)	(0.018)	(0.017)	(0.014)
N	24,376	24,376	24,376	2,4376
Adjusted R^2	0.145	0.150	0.143	0.113

Notes: Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A1. Descriptive statistics

Variables	No. of observations	Mean	Std. Dev.	Min	Max
x_{00}	48,752	0.238	0.426	0	1
x_{05}	48,752	0.256	0.436	0	1
x_{10}	48,752	0.293	0.455	0	1
x_{15}	48,752	0.334	0.472	0	1
Density	48,752	0.229	0.138	0	1
Cross-density	48,752	0.241	0.130	0.024	1
Complexity	48,752	0.000	0.997	-3.078	1.916
Population	48,752	0.103	0.129	0.014	1.137
Population density	48,752	0.120	0.177	0.002	1.793
GDP per capita	48,752	-1.237	0.723	-3.081	0.672
Size	48,752	0.066	0.036	0.001	0.176
Stock	48,752	0.042	0.022	0.001	0.101

Table A2. Correlation matrix

Variables	Index	1	2	3	4	5	6	7	8	9	10	11	12
x_{00}	1	1.000											
x_{05}	2	0.511	1.000										
x_{10}	3	0.344	0.474	1.000									
x_{15}	4	0.322	0.416	0.504	1.000								
Density	5	0.546	0.377	0.270	0.269	1.000							
Cross-density	6	0.224	0.207	0.150	0.127	0.461	1.000						
Complexity	7	0.202	0.188	0.206	0.212	0.182	0.096	1.000					
Population	8	0.100	0.080	0.014	0.034	0.286	0.520	0.000	1.000				
Population density	9	0.076	0.050	0.021	-0.013	0.214	0.507	0.000	0.682	1.000			
GDP per capita	10	0.107	0.110	0.072	0.076	0.319	0.321	0.000	-0.089	-0.067	1.000		
Size	11	0.305	0.278	0.296	0.279	0.217	0.132	0.661	0.000	-0.000	0.000	1.000	
Stock	12	0.291	0.221	0.139	0.142	0.890	0.493	-0.000	0.346	0.263	0.323	-0.000	1.0000

Table A3. The distribution of new industries in terms of relatedness, cross-relatedness, and complexity

Dimensions	Categories	Number of new industries	Percentage
Relatedness	High relatedness	8,353	90%
	Low relatedness	923	10%
Cross-relatedness	High cross-relatedness	2,416	74%
	Low cross-relatedness	6,860	26%
Complexity	High complexity	3,806	41%
	Low complexity	5,470	59%
Relatedness and complexity	High relatedness and high complexity	3,435	37%
	High relatedness and low complexity	4,918	53%
	Low relatedness and high complexity	371	4%
	Low relatedness and low complexity	552	6%
Cross-relatedness and complexity	High cross-relatedness relatedness and high complexity	2,932	29%
	High cross-relatedness relatedness and low complexity	4,128	45%
	Low cross-relatedness relatedness and high complexity	1,074	12%
	Low cross-relatedness relatedness and low complexity	1,342	14%
Total		9,276	100%

Table A4. Robustness analysis

	2000-2005		2000-2010		LQ=0.8		LQ=1.2		Logit	Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
x00	0.449***	0.449***	0.274***	0.275***	0.264***	0.263***	0.261***	0.261***	1.142***	0.700***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.025)	(0.015)
Cross-density	0.233***	0.219***	0.262***	0.207***	0.140***	0.156***	0.129***	0.136***	0.773***	0.456***
	(0.018)	(0.018)	(0.019)	(0.020)	(0.020)	(0.022)	(0.021)	(0.022)	(0.103)	(0.062)
Complexity	0.003*	0.003	0.007***	0.003	0.010***	0.016***	0.014***	0.018***	0.121***	0.070***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.014)	(0.008)
Population	0.017	0.018	-0.239***	-0.239***	0.086***	0.086***	0.112***	0.114***	0.492***	0.287***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.024)	(0.024)	(0.024)	(0.024)	(0.124)	(0.073)
Population density	-0.098***	-0.095***	-0.018	-0.004	-0.237***	-0.240***	-0.231***	-0.233***	-1.222***	-0.706***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)	(0.016)	(0.090)	(0.052)
GDP per capita	0.009***	0.008***	-0.005	-0.007**	0.005	0.005	0.009***	0.010***	0.043**	0.024**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.017)	(0.010)
Stock	1.200***	1.186***	1.099***	1.045***	1.138***	1.147***	1.397***	1.422***	6.966***	4.197***
	(0.097)	(0.097)	(0.109)	(0.109)	(0.103)	(0.103)	(0.113)	(0.113)	(0.602)	(0.357)
Size	1.566***	1.571***	2.481***	2.538***	2.432***	2.360***	2.856***	2.757***	10.781***	6.572***
	(0.063)	(0.064)	(0.070)	(0.071)	(0.060)	(0.061)	(0.067)	(0.067)	(0.391)	(0.233)
Industry FE	No	Yes	No	Yes	No	Yes	No	Yes	No	No
Region FE	No	Yes	No	Yes	No	Yes	No	Yes	No	No
Constant	-0.048***	-0.037***	-0.034***	0.024**	0.013	0.051***	0.023***	0.059***	-2.114***	-1.287***
	(0.008)	(0.009)	(0.008)	(0.010)	(0.009)	(0.011)	(0.009)	(0.011)	(0.050)	(0.029)
N	48,752	48,752	48,752	48,752	48,752	48,752	48,752	48,752	48,752	48,752
Adjusted R^2/Log likelihood	0.290	0.290	0.168	0.170	0.149	0.152	0.150	0.152	-27,386	-27,379

Notes: Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

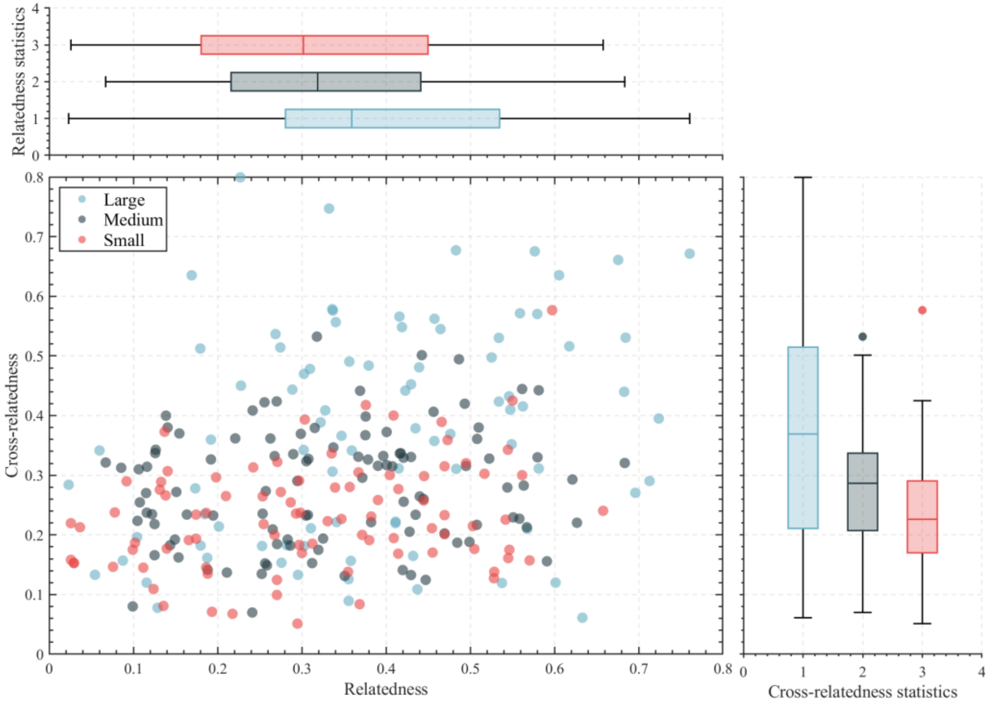

Figure B1. Marginal plot analysis of relatedness versus cross-relatedness