Understanding the Role of Smart Specialization Strategies (S3) within a Regional Innovation System: Evidence from Digital Industries in the Yangtze River Delta, China

Zhen Yue 1,2,†, Meisha Zhang 3,†, Shuran Yang 4 and Kai Zhao 2,4,*

1 School of Foreign Studies, Xi’an Jiaotong University, Xi’an 710049, China; yuezhen@mail.xjtu.edu.cn
2 Soft Science Research Base of Shaanxi High-Quality Economic Development, Xi’an Jiaotong University, Xi’an 710061, China
3 School of Marxism, Chang’an University, Chang’an 710064, China; zms5233@chd.edu.cn
4 School of Economics and Finance, Xi’an Jiaotong University, Xi’an 710061, China; yangshr22@stu.xjtu.edu.cn
* Correspondence: kaizhao@mail.xjtu.edu.cn
† These authors contributed equally to this work.

Abstract: In response to Boschma’s concern that the implications of relatedness- and unrelatedness-based diversification strategies lack empirical evidence at disaggregated levels and in the context of the Global South, this study generates a unique dataset at the city level and explores how these smart specialization strategies (S3) may explain digital industry innovations within a specific regional innovation system, i.e., the Yangtze River Delta, China. The findings reveal that both relatedness density and knowledge complexity play a positive role in explaining digital industry innovations. However, the relationship between relatedness and knowledge complexity and its interactive effects on innovation performance are less straightforward. In our study, we found that efficient cooperation between relatedness and complexity can only be achieved if the level of government intervention is moderate. Therefore, the discussion of S3 focuses on more than the dichotomous argument between relatedness and unrelatedness. Many socio-economic factors also impact the effectiveness of these theoretical components within different innovation systems, which are largely overlooked by present studies.

Keywords: smart specialization; relatedness; knowledge complexity; innovation system; Yangtze River Delta; digital industry

1. Introduction

According to Austrian economic theorists such as Hayek [1], the issue with the concept of a competitive market proposed by neo-classical economists is that it describes an equilibrium but not the process to achieve such a point. Thus, it remains difficult to explain how producers are naturally connected with their competition and how they increase or decrease prices, promote products, or change cost structures during the process of competition. Although he is not a representative figure among Austrian economic theorists, Schumpeter [2] further notes that the focus on economic development should be shifted from resource allocation to creating or destroying resources. Schumpeter believes that innovation and entrepreneurship play a central role during the process of production, which, in turn, results in enormous internal economic well-being, which is far more important than the traditionally defined concept of allocation efficiency. In comparison, the modern Austrian economist Kirzner [3] enriches the core concept of individual behaviors and argues that the purposes and means of economic activities will not remain unchanged. Instead, these factors are determined by individuals with creativity because the
assumptions proposed by neo-economists, such as a constant return to scale, information symmetry, etc., are impossible to realize. To a certain extent, these concepts, which focus more on a free market and individual freedom than on government intervention and central planning, introduce new perspectives to traditional location theories and further clarify the important position of new economic geographies (NEGs), which underscore the discussion around the spatial agglomeration of specific economic activities and associated driving forces [4,5]. The core elements of NEGs, i.e., an increased return to scale, information asymmetry, and path dependency, are intertwined with the concept of Austrian economic theories, as they represent a “good” economic theory that reflects the real situation of markets.

Extensive studies from various perspectives have attempted to understand the realistic interactions between knowledge, innovation, and location under the dominant milieu of the knowledge economy’s development. In particular, smart specialization strategies (S3) that aim to stimulate relatedness diversification within innovation systems have received increasing attention in the domain of economic geography. Such strategies follow the seminal work of Boschma [6], who proposes the concept of smart specialization beyond the geographical proximity perspectives of traditional regional economics. However, when a system is transitioning from a labor-intensive to knowledge-intensive model, institutional features, human capital reserve, and the traditional industrial structure may impede such a system from entering new domains. This process can lead to the phenomenon of “high relatedness and low complexity” or a “lock-in-state”, in which only increasing related diversification in a few directions removes opportunities for further diversification. This contradiction is particularly evident in some old industrial systems with an aim to develop new technologies [7]. Therefore, unlike European regional innovation systems in which the importance of relatedness diversification is repeatedly emphasized, determining how to effectively use the concept of relatedness and implement an S3 strategy is less straightforward in the regional innovation systems of developing countries such as China. These systems face the burden of overcoming the “middle income trap”. Thus, involving the role of complexity appears to complicate our understanding of the importance of relatedness diversification policy making and introduces new challenges to innovation systems with fewer capabilities to engage in more complex activities.

To solve the above conundrum, we used the development of the digital industry in the Yangtze River Delta region, China, as an example. This area is a representative regional innovation system that is more likely than other regions to serve as a center of high-tech industries. This region can also be used as an example to validate the theoretical claims of the S3 theory. In this study, we sought answers to the following questions: (1) Does relatedness outperform unrelatedness in explaining the inter-regional innovation process of a newly emerged high-tech industry? (2) What would the interactive effect be between relatedness and complexity? Here, we aim to provide region-specific solutions to overcome the potential low complexity trap. To the best of our knowledge, this is one of the first studies that empirically and directly investigates the role of relatedness and unrelatedness in explaining industrial innovation outputs at a sub-national level in the context of the Global South. In doing so, and in response to the issue of missing data and inconsistency, we provide a novel perspective to measure the concept of relatedness [8,9] and knowledge complexity [10] using both primary and secondary data. We aim to not only shed more light on the debate regarding “related or unrelated diversification” with non-Northern empirical evidence but also provide more meaningful materials and references for enriching and revisiting the S3 theory based on socio-economic reality. This study aims to achieve the following objectives: (1) enrich empirical analysis at the sub-national level; (2) shift the perspective from the Global North to the Global South; (3) further integrate a region-specific situation with the development of digital industries; and (4) describe how the effects of S3 policy on innovation are moderated by other, broader socio-economic factors.
This paper is organized as follows. The following section presents the relevant literature and provides an overview of the debate on the relatedness and complexity of smart specialization policy, as well as the development of the digital economy. The third section discusses the empirical data, variables, and methods required for this study. This section is followed by a section that calculates the characteristics of the relatedness of digital industry in the Yangtze River Delta region. The penultimate section constructs a cross-section regression model to estimate the impact of relatedness density and knowledge complexity on digital industry innovation.

2. Literature Review

2.1. Digital Industry Innovation

With the rapid development of information technologies such as big data, cloud computing, and artificial intelligence, the development of the modern economy and industrial enterprises has entered the digital age. There are several definitions of the digital economy [11]. Yoo et al. specifically define the digital economy as the process by which the digital and physical components of a product or service use new combinations to produce new products or provide new services. Digital Croatia defines the digital economy as a new form of economy based on digital technologies [13] and argue that this area represents one of the most attractive opportunities for growth. Bukht and Heeks divide the digital economy into three levels [14]. The first level is the “IT field”, which includes hardware manufacturing, software and IT consulting, and information services; the second level is the narrow sense of the digital economy, which includes electronic business, digital services, and the platform economy; the third level is the digital economy in the broad sense, which includes E-commerce, industry 4.0, precise agriculture, and the algorithm economy.

The present studies mainly focus on the establishment and evolution of digital infrastructure from a technical perspective [15]. The most important driving force for the development of the digital industry appears to be innovation capacity. However, scholars have reached mixed conclusions regarding the definition of digital innovation. Such definitions can be generally summarized using the following categories: (1) digital product innovation [16], such as smart home products; (2) digital process innovation [17] (for example, digital technology greatly reduces R & D costs); (3) digital organization innovation [18], such as the establishment of a chief digital office for organizational innovation; (4) digital business model innovation [19], which includes supply connection, marketing, and other business models with the aid of updates to digital technology; and (5) results and technological innovation, which emphasize innovative results and improved IT technologies. In this study, we selected patent data as a proxy of definition (4) to measure the level of technological innovation in the digital technology industry of the Yangtze River Delta region, as datasets related to innovation processes and intermediate services are very limited in China.

Digital innovation is influenced by many factors. Kohli and Melville maintain that the emergence of digital innovation is an outcome of the increasing pressure on organizations to apply digital technology in their products and change existing business models [20]. As one of the most significant drivers of innovation, digitization has a disruptive impact on almost every industry [21]. From a corporate management perspective, Chen et al. believe that the CIO’s issue-selling effectiveness, rather than its structural location [22], directly affects the level of corporate digital innovation. Some research focuses on disruptive innovation in digital technology from the perspective of technological change [23]. For instance, Sandberg et al. further analyzed the differences in various digital technology application levels (the base layer, platform layer, and application layer) [24]. Some scholars instead argue that innovative applications at all levels are based on standardized technologies that increasingly contain standard essential patents and that owning standard essential patents is essential to gaining and maintaining a significant market share [25].
Moreover, the formation of technical standards can effectively solve the problem of technological discontinuity and subsequently promote industrial innovation and development [26]. However, this branch of the literature only focuses on key technologies, management strategies, and property rights protection at the micro level. The concepts of digital development and innovation have not yet been used to understand the disparity in socio-economic development caused by the lack of adaptability towards radical structural change, policy implementation, etc.

2.2. Smart Specialization, Relatedness, and Complexity

S3 is an economic development approach that explores the potential advantages of an innovation system within, e.g., a city or region [27–29]. The core concept of S3 is smart specialization, which determines a system’s growth probability based on certain place-based strengths [30,31]. Smart specialization aims to target economic domains related to its potential strengths. Unlike its name suggests, the goal of smart specialization is not to make an innovation system more specialized but to leverage existing strengths to promote diversification into new economic platforms and domains. A smart specialization strategy follows a core principle: When innovation systems develop a new domain, they should target and start from where those systems have relevant capabilities or variety to promote knowledge and technology exchanges across domains [32]. Related diversification reflects a path-dependent process that borrows from and combines relevant local technologies and capabilities [33]. For instance, the development of the car industry is likely to be based on several relevant capabilities such as engineering, training, etc. [34] because the implementation is easier and less costly [35]. Many other approaches also incorporate such patterns, including branching, path dependence, bounded rationality, etc. [32,36,37].

However, critics argue that this type of S3 overemphasizes the power of related diversification and is conservative and less altered [38], which could remove diversification opportunities and make regions less resilient over time [9,39,40]. Therefore, some scholars seek a type of S3 that focuses on unrelated diversification [38,40,41]. This type of S3 aims to realize a potential brand-new path [41]. The development of the economy in the long term may face the risk of lock-in [42,43], and unrelated diversification may lead to radical change and new support in regions [44]. Nevertheless, there also remain questions and weaknesses related to the policy focus on unrelated diversification. The potential for systems to start from zero with little experience creates a high risk of policy failure. It is much easier for regions to show higher economic growth when diversifying into related and complex activities, rather than unrelated activities [35,45].

The theory of S3 has been frequently mentioned by local governments in Europe only since 2014. Even though extensive studies have investigated how the implementation of S3 affects regional/city growth and innovation, this theory remains a relatively new research domain. Innovation systems at different levels require time to adjust and unfold structural changes. Thus, associated results, both positive and negative, may be difficult to accurately measure over the short term. However, there is consensus that the development and improvement of smart specialization depend on the situations that innovation systems generate and upgrades to sustainable advantages. Therefore, the proposed concept of complexity appears to represent a compromise between relatedness- and unrelatedness-oriented policies, which at least provide a “benchmark” or “guarantee” that the implementation of S3 will avoid the low-complexity trap documented by a plethora of studies [9,46,47].

To date, mixed conclusions have been reached regarding the definition and usefulness of the S3 theoretical framework. However, these mixed conclusions are unlikely to be a consequence of the debates regarding the validity of theoretical frameworks or methodological improvements. Instead, this uncertainty remains because the role of S3 within the process of socio-economic development has been less empirically explored. As Boschma notes [48], “it is not a matter of S3’s focus either on related or unrelated diversification, but that this choice depends on the region-specific context”. Therefore, this study
aims to shed some light on the current research domain as follows: (1) enriching empirical analysis at the sub-national level, particularly at a new economic–geographical level, i.e., metropolitan groups that are quite different from the conventional typology such as major urban regions, old industrial regions, and peripheral regions [49]; (2) shifting the perspective from the Global North to the Global South, as almost all present studies focus on European or North American contexts; (3) further integrating a region-specific situation within the development of digital industries; and (4) describing how the effect of S3 policy on innovation is moderated by other broader socio-economic factors. We expect that our research settings, which are different and more complicated than those in previous studies, will provide meaningful empirical evidence for developing an appropriate tool to identify regional lock-ins and traps. Instead of discussing a polarized S3 policy, i.e., related or unrelated diversification, the effect of specification strategies should be dynamic and based on different inter-system linkages, institutional settings, industrial features, etc. Thus, there would be no one-size-fits-all panaceas but instead individual remedies that suit specific cases.

3. Data Collection, Variable Generation, and Model Specification

3.1. Data Collection and the Generation of Core Variables

The Yangtze River Delta metropolitan area, including the Shanghai municipality and Zhejiang, Jiangsu, and Anhui provinces, is the most developed region in China, with an integrated industry system and chain. In the Yangtze River Delta Regional Integration Development Plan proposed by the Chinese central government in 2019, the role of this area is clearly defined as a “growth pole”, “science and technology center”, and “demonstration area” similar to metropolitan areas in the U.S., Europe, and Japan. The corresponding positioning and development purposes are different from those of other comparatively developed regions such as the Guangdong–Hong Kong–Macao Greater Bay Area, as the development of high-tech industry is a priority. This featured background provides us a good opportunity to investigate the features of digital industries.

Based on the Innojoy database, a total of 70 four-digit industries in China’s national economy industry category were selected, including computer, communication, and other electronic equipment manufacturing industries (C39) and information transmission, software, and information technology service industries (I63-65) for all 27 cities in the Yangtze river delta region in 2021 (Table 1). The number of invention patents and utility model patents was categorized using each four-digit industry in each city according to “the GB/T 4754-2007 classification standard” and “international patent classification and national economic classification”. The central explanatory variables were obtained from the basic information and financial indicator data of listed companies belonging to the four-digit industries in the IFIND financial database. We selected the listed companies that remained active until 31 December 2021, with the latest data disclosed in 2021. Control variables were obtained from the IFIND financial database and the China Statistical Yearbook 2020. Due to space limitations, only the selected double-digit industries are listed here, as shown in the table below.

<table>
<thead>
<tr>
<th>Industry Name</th>
<th>Industry Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer, communication, and other electronic equipment manufacturing industries</td>
<td>C39</td>
</tr>
<tr>
<td>Telecommunications, radio and television, and satellite transmission services</td>
<td>I63</td>
</tr>
<tr>
<td>The internet and related services</td>
<td>I64</td>
</tr>
<tr>
<td>Software and information technology services</td>
<td>I65</td>
</tr>
</tbody>
</table>

Due to limitations in data availability, there is no unified definition for industrial innovation. Many studies use patent applications or the output value of new products as a proxy. In comparison, Balland et al. proposed a novel approach using the number of new...
patents belonging to a specific industrial category in a city or region as the proxy of innovation capacity [35]. However, related data are not accessible in Chinese databases. Thus, we still used the number of four-digit industry invention patents and utility model patents to express the innovation capacity of each four-digit industry using the variable $Patent_{r,j}$.

Here, we mainly focus on relatedness density and knowledge complexity, with the variables $Relatedness\_Density_{r,j}$ and $TCI_{r,j}$, respectively. At present, there are roughly three approaches for measuring relatedness density. The first approach follows the simple logic that two industries are related to each other if they are located within the same two- or three-digit category [6]. However, this method ignores the cross-industry input–output and technology spillover links between the upstream and downstream areas of the same product industry chain [50]. Based on the input–output table, the second approach assumes that industries with more similar resource inputs, such as the means of production, labor, and R&D activities, are more likely to have a higher degree of technological similarities and relatedness. For example, Lemelin and Guo et al. calculated an industrial resource structure through input–output tables to compare the degree of similarity among inputs between different industries [51,52]. However, while different industries invest similar resources, the proportion and importance of such resources remain different [53]. For example, human input is more suitable for evaluating labor-intensive industries, and patents are more suitable for evaluating technology-intensive industries. The third method is coexistence analysis, which measures the degree of relatedness between industries with the conditional probability of two industries’ production activities in the same region. If two industries are frequently identified in the same region, these two industries are likely more closely related to certain resources or technologies. This method directly uses a larger range of industrial data based on the final output of enterprises without concern for the endogeneity and heterogeneity of more complex intermediate data such as factor inputs. Therefore, we also employed this approach to investigate the degree of digital industry relatedness in the Yangtze River Delta region. The specific calculation method is shown below. $Relatedness_{ij}$ indicates the relatedness between two industries; the larger the value is, the more related the two industries are:

\[
LQ = \frac{X_{ri}/\Sigma X_{ri}}{\Sigma X_{ri}/\Sigma_{r}X_{ri}}, \quad (1)
\]

\[
Relatedness_{ij} = \min \left[ P(RCA_j/RCA_i), P(RCA_i/RCA_j) \right], \quad (2)
\]

\[
RCA = \begin{cases} 
1, & LQ \geq 1 \\
0, & LQ < 1 
\end{cases}, \quad (3)
\]

where $LQ$ indicates the total asset location entropy of listed enterprises belonging to a four-digit industry in a certain city, $r$ is the city, $i$ and $j$ are the four-digit industries, and $X_{ri}$ is the total assets of a listed enterprise in $r$ city and $i$ industry. RCA is the dominant comparative advantage. If the location entropy of $r$ city and $j$ industry is larger than 1, this city and the industry are considered to have a dominant comparative advantage, which means that $RCA = 1$. $P(RCA_j/RCA_i)$ is a conditional probability expression which refers to the probability that industry $j$ has a dominant comparative advantage when industry $i$ has a dominant comparative advantage. For example, in the 27 cities studied in this article, $n$ cities have a dominant comparative advantage in industry $i$ and $m$ cities have a dominant comparative advantage in industry $j$, while $c$ cities have comparative advantages in both industry $i$ and industry $j$. We can thus conclude that $Relatedness_{ij} = \min [(c/m) + (c/n)]$ using the above principle. After the above calculation, the relatedness matrix of 70 four-digit industries in the Yangtze River Delta region can calculated and recorded as $B$, which represents a real symmetric matrix of $70 \times 70$, and all diagonal elements are 1. We next apply the work of Balland et al. (2019) [35] to represent the dominant comparative
advantage of city and industry (when city and industry have the dominant comparative advantage, \( \sigma_{rj} = 1 \)). The relatedness density index between each industry at the aggregate level or within a particular city can be calculated via the following method:

\[
Relatedness\_Density_{rj} = \frac{\sum_{j \neq i} Relatedness_{ij}}{\sum_{j \neq i} Relatedness_{ij}}
\]

(4)

We also follow Balland et al.'s approach to estimate knowledge complexity [35]. We multiply the relatedness matrix \( B \) composed of all paired industries with its transposed matrices \( BT \) to obtain a new matrix \( M \), where \( M = B \times BT \). \( M \) is a square matrix with the same dimension as the four-digit industries (70 \( \times \) 70). Finally, the knowledge complexity index of each industry is given by the value of each element's second eigenvector \( \hat{Q} \) within the matrix \( M \), and then all elements in vector \( \hat{Q} \) are standardized for the subsequent regression analysis, which is \( TC\_I_j = \frac{\hat{Q} - \langle \hat{Q} \rangle}{\text{std}(\hat{Q})} \).

3.2. Model Specification

This section presents a cross-sectional regression analysis to explore the impact of relatedness and knowledge complexity on digital industry innovation in the Yangtze River Delta region. Besides the main variable, relatedness density and knowledge complexity, we also include control variables: (1) FDI. The technology spillover effect of FDI indirectly affects regional innovation capacity. (2) Human capital (Human) is an important factor affecting regional or industrial innovation activities. Thus, the number of regional college students per capita is used here to indicate the level of human capital reserves. (3) Government intervention (GOV) has an impact on all industrial production activities in a region. We assume a local government with a higher level of fiscal expenditures is more likely to have resources and time to intervene in local industries’ activities. Thus, the share of fiscal expenditures over the total GDP of a city is used as a proxy. (4) The population density (Pop_density) indirectly affects the regional innovation ability by strengthening the human capital level. (5) The per capita GDP (GDP_per_capita) directly measures the level of regional economic development. Regions with a higher level of economic development generally have greater innovation vitality. (6) The level of technological expenditures (Tech) reflects the general quality of technology and science of a local environment. We assume that innovative activities among enterprises are more likely to be motivated if a local government focuses on supporting technological and scientific developments directly or indirectly.

To exclude cases where control variables affect the significance of core variables, we first construct the following basic regression to verify whether relatedness density and knowledge complexity affect the digital industry innovation capacity of cities in the Yangtze River Delta region:

\[
Patent_{rj} = \beta_0 + \beta_1 Relatedness\_Density_{rj} + \beta_2 TC\_I_j + \epsilon_{rj}
\]

(5)

The dependent variable \( Patent_{rj} \) is the industry’s innovation capabilities and uses the number of applications for invention patents and utility model patents in industry \( j \) and city \( r \) in 2021. \( Relatedness\_Density_{rj} \) is the relatedness density of city \( r \), industry \( j \), and other industries in the city. \( TC\_I_j \) is the knowledge complexity of each industry, and \( \epsilon_{rj} \) is the perturbation term.

Ensuring the coefficients of main explanatory variables are unbiased, we next introduce the control variables for the regression analysis:

\[
Patent_{rj} = \beta_0 + \beta_1 Relatedness\_Density_{rj} + \beta_2 TC\_I_j + \beta_3 FDI_r + \beta_4 Human_r + \beta_5 GOV_r + \beta_6 Tech_r + \beta_7 ln(Pop\_density_r) + \beta_8 ln(GDP\_pc_r) + \gamma_{rj}
\]

(6)

Because an increase in innovation capacity also reversely affects the level of technological relatedness and complexity, the three-stage least squares method (3SLS) was used to resolve the endogeneity issue caused by simultaneity.
The following is the system of equations used for 3SLS estimation:

\[
\begin{align*}
\text{Patent}_{ij} &= \theta_0 + \theta_1 \text{Relatedness}_i \text{Density}_{ij} + \theta_2 TCI_j + \theta_3 FDI_i + \theta_4 \text{GOV}_i + \theta_5 \text{Tech}_i + \mu_{ij} \\
\text{Relatedness}_i \text{Density}_{ij} &= \alpha_0 + \alpha_1 \text{Patent}_{ij} + \alpha_2 TCI_i + \alpha_3 \ln(\text{GDP}_i) + \alpha_4 \ln(\text{Pop}_i) + \alpha_5 \text{Human}_i + \phi_{ij}
\end{align*}
\] (7)

We also constructed a spatial measurement model of relatedness and complexity based on industrial innovation capacity to handle the potential endogeneity issue caused by omitted variables:

\[
\begin{align*}
Y &= \rho WY + X\beta + \theta WX + \mu \\
\mu &= \lambda W\mu + \varepsilon \sim N[0, \sigma^2I]
\end{align*}
\] (8) (9)

where \( Y \) indicates the explained variable, \( X \) represents the explanatory variables \( \text{Relatedness}_i \text{Density}_{ij} \) and \( TCI_j \), \( W \) represents the spatial weight matrix of \( n \times n \) dimensions, \( \beta \) represents the correlation coefficient of \( X \), and \( \rho \) and \( \theta \) represent the spatial correlation coefficients. \( \Lambda \) represents the spatial error coefficient, \( \mu \) and \( \varepsilon \) represent the random error, and \( \varepsilon \) follows a normal distribution. When \( \rho \neq 0 \), \( \theta = 0 \), and \( \lambda = 0 \), the above equation conforms to the spatial autoregressive model (SAR). When \( \rho = 0 \), \( \theta = 0 \), and \( \lambda \neq 0 \), the above formula conforms to the spatial error model (SEM). When \( \rho \neq 0 \), \( \theta \neq 0 \), and \( \lambda = 0 \), the above formula conforms to the spatial Dubin model (SDM).

Together with other important factors, the effects of relatedness and complexity on a city’s innovation capacity are likely to be determined by disparities in socio-economic development. Therefore, we lastly apply a threshold model to capture potential heterogeneity as follows:

\[
\begin{align*}
\text{Patent}_{ij} &= \mu_0 + \mu_1 \delta_{ij} + \gamma_1 X_j \cdot I(q < \rho_1) + \gamma_2 X_j \cdot I(\rho_1 < q < \rho_2) + \gamma_3 X_j \cdot I(\rho_2 < q < \rho_3) + \gamma_4 X_j \cdot I(q > \rho_3) + \varepsilon_{ij}
\end{align*}
\] (10)

where \( q \) is the threshold variable, \( \rho_i \) is the threshold estimate for this regression, and \( X_i \) is an explanatory variable. We specifically chose GDP per capita and the level of government intervention (i.e., the share of government expenditures) as the possible threshold proxies that could affect industrial innovation at different stages. The interactive term of these proxies is also included for further analysis.

4. Analysis

4.1. Characteristics of Relatedness Density in the Yangtze River Delta Region

The total asset information of 1890 listed four-digit industry companies in 2021 was sorted through the data screening and matching of 70 four-digit industries within the category of digital industry among 27 cities in the Yangtze River Delta region. Then, we used python programming to calculate the \( \text{Relatedness}_i \text{Density}_{ij} \) between two industries and \( \text{Relatedness}_i \text{Density}_{ij} \) of each four-digit industry at the city level, with 2415 industrial pairs and 1890 industrial–city pairs. Among them, relatedness industry pairs accounted for 43.23% of the total, which means that 56.77% of industries did not have a comparative advantage in the 27 cities. Regarding the overall situation of the Yangtze River Delta region, the average relatedness density among 2415 industrial pairs in 2021 was approximately 0.36, and the variance was approximately 0.07. Thus, the dispersion degree was relatively small. The minimum value of relatedness density of related industries was approximately 0.11, and the maximum value was 1, indicating the presence of some fully related digital industry pairs in the Yangtze River Delta region in 2021.

In addition to the different characteristics of relatedness in the region, the relatedness among the four-digit digital industries also varied from the perspective of the Yangtze River Delta’s integration. Observing the probability of relatedness between each four-digit
industry and other industries, the relatedness of each industry was 37.17 on average, accounting for more than 50% of the total number of industries. This result shows that digital industries in the Yangtze River Delta region are well-coordinated. In addition, the highest probability of relating with other industries was identified in the information system integration service (6531) and Internet of Things technology service (6532) industries, which were both at 62/70. The smallest value was yielded by the internet security services (6440) industry, as no other industries were found to be related. Figure 1 shows at a high level how each four-digit digital industry relates to the others in terms of quantity.

![Figure 1. The breadth of relatedness in four-digit industries. Source: Authors’ own calculation.](image)

As shown in Figure 1, from the perspective of the strength of association, industrial pairs with an association strength of 1 such as internet platform services (including I6432 and I6434), operation and maintenance services (I6540), and radio and television satellite and animation and game digital content services (I6572) were effectively participating in the same service chain. Because the service process is interlinked, the flow of knowledge between industries is generally more frequent.

We further calculated and constructed the technical complexity index (TCI) [28]. According to the technical complexity of 70 four-digit digital industries, we found that the
most complex industries were information security equipment manufacturing (C3915) and computer parts manufacturing (C3912). Overall, the technical complexity of internet information platform and professional equipment manufacturing industries was high, while the technical complexity of information storage and satellite broadcasting industries was relatively low, which is in line with the innovative characteristics of current digital technology. Figure 2 shows the technical complexity indexes of 70 digital industries and four-digit industries in the Yangtze River Delta region (to facilitate the subsequent regression analysis, the TCI index is the result of a standardized treatment; thus, a larger value indicates higher complexity).

![Figure 2. Technical complexity of the four-digit industries. Source: Authors' own calculation.](image)

To explore the spatial pattern of relatedness and determine the disparity in relatedness among the 27 cities in the Yangtze River Delta region, we calculated the average relatedness density value for the 70 four-digit industries in each city \( \bar{Relatedness}_{e} = \frac{1}{10} \sum_j Relatedness_{Density}_{r_j} \). The larger the value of \( \bar{Relatedness}_{e} \), the higher the average level of coordination among all four-digit industries in a city. The average number of patent applications in 70 four-digit industries was also calculated in each city \( \bar{Patent}_{r} = \frac{1}{10} \sum_j Patent_{r_j} \) to determine each city’s innovation capabilities. We then used a scatter diagram to visualize the potential association between relatedness and innovation. As shown in Figure 3, only Shanghai, Hangzhou, Suzhou, and Nanjing, as the major developed cities and creative hubs, performed well in both the innovation and relatedness dimensions. However, many cities, despite an evident level of relatedness, did not equally produce solid innovation outputs. This finding suggests a less straightforward relationship between technological relatedness and regional innovation capabilities in the scenario of this study, where differences in socio-economic status may play a vital role in the process of digital industry development.
4.2. Relatedness, Knowledge Complexity, and Digital Industry Innovation in the Yangtze River Delta Region

$GDP_{\text{per capita}}$ and $Pop_{\text{density}}$ are natural logs. Table 2 shows the descriptive statistical results. Considering that the model’s validity is affected by the issue of multicollinearity, a correlation analysis between various explanatory variables and control variables was first conducted before performing the cross-sectional model regression estimation. The results are shown in Table 3. Here, the correlation coefficients of all explanatory variables are generally low, with a maximum value of 0.612. Thus, multicollinearity is unlikely to be a serious issue for our model specification.

Table 2. Descriptive statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Max</th>
<th>Min</th>
<th>Sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Patent_{rj}$</td>
<td>365.38</td>
<td>14,804</td>
<td>0</td>
<td>1236.02</td>
</tr>
<tr>
<td>$Relatedness_{Density}_{rj}$</td>
<td>0.13</td>
<td>1</td>
<td>0</td>
<td>0.21</td>
</tr>
</tbody>
</table>
Next, we performed a regression analysis of the cross-sectional model. The regression results are presented in Table 4. The basic regression results in the first column show positive coefficients for relatedness and complexity with significance at 1% and 5% levels, respectively. This result indicates that the level of relatedness among industries has significantly improved the innovation capacity of digital industries in the Yangtze River Delta region. In industries with a higher level of relatedness, the technological innovation level was significantly higher than that in industries with a lower level of relatedness. To more specifically explore the effects of the core explanatory variables, a cross-sectional regression was performed again by introducing the control variables in the second column of Table 4. The results show that the coefficients of relatedness and complexity remained significantly positive. Economic variables are usually interdependent and mutually causal. Thus, to solve the problem of endogeneity caused by bidirectionality between relatedness and the level of industrial innovation, and considering the limitations of handling endogeneity when using cross-sectional data for empirical analysis, the third column presents a robustness test of the model using the three-stage least-squares method (3SLS). According to the test results, the coefficients of core explanatory variables remained positively significant. This outcome demonstrates that the results of the basic regression are robust. However, through the second group of regressions, we found that the influence of complexity on relatedness was significantly negative, indicating that the degree of relatedness in a region is likely to reduce in association with an increase in the complexity level.
Table 4. Basic regression and 3SLS robustness test of relatedness and industrial innovation level.

<table>
<thead>
<tr>
<th></th>
<th>Patent_{rj} (OLS)</th>
<th>Patent_{rj} (OLS)</th>
<th>Patent_{rj} (3SLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relatedness_Density_{rj}</td>
<td>3.057 *** (0.202)</td>
<td>0.742 *** (0.184)</td>
<td>23.262 *** (3.244)</td>
</tr>
<tr>
<td>TCI_{rj}</td>
<td>0.280 ** (0.104)</td>
<td>0.282 ** (0.083)</td>
<td>0.411 * (0.191)</td>
</tr>
<tr>
<td>FDI_{r}</td>
<td>0.025 *** (0.001)</td>
<td>-0.027</td>
<td></td>
</tr>
<tr>
<td>Human_{r}</td>
<td>0.135 * (0.071)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOV_{r}</td>
<td>28.740 *** (2.275)</td>
<td>-2.065 ** (0.781)</td>
<td></td>
</tr>
<tr>
<td>Tech_{r}</td>
<td>0.063</td>
<td>-18.868 *** (4.414)</td>
<td></td>
</tr>
<tr>
<td>GDP_per_capita_{r}(log)</td>
<td>0.405 * (0.194)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop_density_{r}(log)</td>
<td>-0.765 *** (0.075)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.736 *** (0.053)</td>
<td>2.366</td>
<td>2.936 *** (0.196)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Relatedness_Density_{rj} (3SLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent_{rj}</td>
<td>0.089 *** (0.005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCI_{r}</td>
<td></td>
<td>-0.030 * (0.013)</td>
<td></td>
</tr>
<tr>
<td>GDP_per_capita_{r}(log)</td>
<td>-0.033 * (0.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop_density_{r}(log)</td>
<td>0.021 * (0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human_{r}</td>
<td></td>
<td>-0.019 ** (0.006)</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>0.086</td>
<td>0.130</td>
</tr>
</tbody>
</table>

N 1648 1648 1648
R^2 0.125 0.538 n.a

Notes: standard errors in parentheses, * p < 0.10, ** p < 0.05, *** p < 0.01.

We also used the Moran index to test the spatial autocorrelation between the industrial innovation level, relatedness, and complexity. The significantly positive coefficient of the Moran index implies that the level of industrial innovation and relatedness is spatially correlated. Next, we conducted LM tests. All four tests rejected the null hypothesis, suggesting that the sample we selected was affected by both spatial lag and spatial error autocorrelation. In response, we constructed the SAR model, SEM model, and SDM model shown in Table 5. For spatial metrology models, the SDM model was adopted because it had the smallest value. We then performed an LD test on our SDM model to verify whether it could be reduced to an SAR model and SEM model. As shown in Table 5, the results of the LR test were 98.34 and 78.62 and significant at a 1% level.
Table 5. Regression results of the spatial measurement model.

<table>
<thead>
<tr>
<th></th>
<th>Patent_{rj} (SAR)</th>
<th>Patent_{rj} (SEM)</th>
<th>Patent_{rj} (SDM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relatedness Density_{rj}</td>
<td>0.828 *** (3.48)</td>
<td>0.743 *** (2.31)</td>
<td>0.739 *** (3.54)</td>
</tr>
<tr>
<td>TCI_{j}</td>
<td>0.13</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>FDI_{r}</td>
<td>0.014 *** (1.19)</td>
<td>0.012 *** (1.31)</td>
<td>0.035 *** (0.67)</td>
</tr>
<tr>
<td>Human_{r}</td>
<td>0.323 *** (3.72)</td>
<td>0.316 *** (2.94)</td>
<td>0.030 (-0.13)</td>
</tr>
<tr>
<td>GOV_{r}</td>
<td>-1.104 ** (-2.45)</td>
<td>-1.052 ** (-2.26)</td>
<td>-5.192 *** (-4.33)</td>
</tr>
<tr>
<td>Tech_{r}</td>
<td>13.365 *** (5.34)</td>
<td>12.856 *** (2.44)</td>
<td>2.478 (0.04)</td>
</tr>
<tr>
<td>GDP_per_capita_{rj} (log)</td>
<td>0.16 (-0.44)</td>
<td>0.01 (-0.14)</td>
<td>0.94 (3.42)</td>
</tr>
<tr>
<td>Pop_density_{rj} (log)</td>
<td>-0.293 *** (-3.19)</td>
<td>-0.322 *** (-3.28)</td>
<td>-0.424 * (-1.72)</td>
</tr>
<tr>
<td>The fixed effect</td>
<td>-0.871</td>
<td>-3.335</td>
<td>-6.972</td>
</tr>
<tr>
<td>Log-L</td>
<td>-3722.787</td>
<td>-3727.551</td>
<td>-3747.934</td>
</tr>
<tr>
<td>sigma²</td>
<td>3.001 *** (31.26)</td>
<td>3.053 *** (30.32)</td>
<td>2.814 *** (33.46)</td>
</tr>
<tr>
<td>R²</td>
<td>0.973</td>
<td>0.889</td>
<td>0.893</td>
</tr>
<tr>
<td>LR test</td>
<td>98.34 ***</td>
<td>78.62 ***</td>
<td></td>
</tr>
</tbody>
</table>

Notes: standard errors in parentheses, * p < 0.10, ** p < 0.05, *** p < 0.01.

The coefficients of relatedness were also positive and significant at a level of 1%, which again indicates that a higher level of relatedness is related to a higher level of industrial innovation in this highly developed region in China. However, the coefficient of complexity was not significant. This finding may suggest a notable feature of technological complexity under the dominant milieu of digital industry development in the Yangtze River Delta region, as increasing the level of technological complexity was found to be a follow-up activity within specific industrial systems in each city. In other words, knowledge can be easily transferred and related among different digital industries and even among different cities. However, complexity, as an outcome of relatedness, is less likely to be affected by geographically nearby industries in the initial stage of development. A further decomposition of coefficients also showed a similar pattern (Table 6): The relatedness in a city not only has a positive effect on promoting the innovation level of digital industries but also has a significant spatial spillover effect on other cities.

Table 6. Spatial effect decomposition.

<table>
<thead>
<tr>
<th></th>
<th>Direct Effect</th>
<th>Indirect Effect</th>
<th>Total Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relatedness Density_{rj}</td>
<td>0.876 ***</td>
<td>2.681 **</td>
<td>3.551 ***</td>
</tr>
<tr>
<td>TCI_{j}</td>
<td>0.936</td>
<td>0.819</td>
<td>0.743</td>
</tr>
<tr>
<td>FDI_{r}</td>
<td>0.039 **</td>
<td>0.219 *</td>
<td>0.2385 *</td>
</tr>
<tr>
<td>Human_{r}</td>
<td>0.242 **</td>
<td>2.303 **</td>
<td>3.082 *</td>
</tr>
</tbody>
</table>
A threshold model was employed to find more specific reasons why the degree of digital innovation was inconsistent with the level of relatedness across different regions. First, the GDP per capita, as a well-known proxy of socio-economic development, was used as a threshold variable. In addition, government intervention (GOV) was considered a threshold proxy of political institution power because the presence of local regulations and management plays a vital and direct role during the process of system innovation. Relatedness, knowledge complexity, and their interaction terms were also analyzed as explanatory variables.

As shown in Table 7, a triple threshold effect is evident when discussing the nonlinear relationship between relatedness and innovation capacity. However, there is no threshold effect when the scenario is analyzed with complexity or an interaction term. The summarized results are presented in Table 8. Specifically, when the value of the threshold variable is below 11.795, the coefficient of relatedness is not significant. However, when the threshold value is between 11.795 and 11.934, the coefficient becomes −0.607. When the threshold value is between 11.934 and 11.973, the coefficient of relatedness is still not significant. A positive effect of 0.781 is produced when the threshold variable is higher than 11.973, which describes an inverted U-shaped trend. Regarding this result, it is believed that in regions with extremely high development, relatedness has an apparently positive impact on digital innovation. However, when economic development is relatively low, relatedness may have a negative influence on innovation at a certain point.

In comparison, when the threshold variable is the level of government intervention and has a value of less than 0.249, the coefficient of relatedness is 0.262. If the threshold value of government intervention is greater than 0.249, the relationship between relatedness and system innovation becomes negative, with a coefficient of −0.384. Similarly, considering the effect of the interactive term between relatedness and complexity, the coefficient is 0.440 if the threshold value is below 0.232. If the value is higher than 0.232, the coefficient of the interaction term is not significant. These findings suggest a moderating role of government intervention during the process of digital industrial innovation. However, the level of government intervention can neither be too high nor too low; otherwise, innovation, together with technological coordination, complexity, and improvement, are suppressed. Thus, complexity was, again, found to be irrelevant for explaining such an innovation process alone.

Table 7. Threshold effects of per capita GDP and GOV.

<table>
<thead>
<tr>
<th>Threshold Variables</th>
<th>Explanatory Variables</th>
<th>Model</th>
<th>Threshold Estimate</th>
<th>F-Value</th>
<th>p-Value</th>
<th>Number of Bootstrap</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per capita GDP</td>
<td>Relatedness</td>
<td>Single threshold</td>
<td>11.795</td>
<td>32.875</td>
<td>***</td>
<td>0.000</td>
<td>7.824</td>
<td>5.579</td>
<td>3.869</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Double threshold</td>
<td>11.934</td>
<td>7.595</td>
<td>***</td>
<td>0.010</td>
<td>7.724</td>
<td>5.442</td>
<td>4.281</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Triple threshold</td>
<td>11.973</td>
<td>17.337</td>
<td>***</td>
<td>0.000</td>
<td>13.007</td>
<td>5.523</td>
<td>3.863</td>
</tr>
</tbody>
</table>
5. Discussion

After the analysis, we sought to develop a new scenario not presented in previous studies to demonstrate how S3 policies or their theoretical components, such as systems-specific relatedness or unrelatedness, could explain global socio-economic events. In other words, if the implementation of S3 can help stimulate innovation in Europe, our goal was to determine whether the corresponding protocol and its effectiveness could also be observed in other economies with different cultural and institutional settings. We found that Chinese digital industries are more likely to follow the path of diversified relatedness, as within the category of digital industry, different innovation activities are closely linked with each other. This joint effect on the level of industrial innovation was found to be significant, which is consistent with the results of Guo and He and Li [52,53], who reported that the level of diversification is driven by relatedness at an industrial level. Meanwhile, the positive, direct effect of complexity will lead to high potential economic returns due to a high level of complex activities [7]. However, the spatial effect of complexity is neither more positive nor more significant than relatedness, which indicates that the average knowledge complexity of a city does not have a direct spillover impact on surrounding areas. Additionally, efficient cooperation between relatedness and complexity can only be achieved if the level of government intervention is moderate. Intervention that is too high or too low will only lead to a negative relationship, as suggested by Boschma [48]. With this factor in mind, the present study further proposes how a specific combination of factors contributes to the development of an innovation system instead of only highlighting which factors could be pertinent. Thus, the discussion of S3 should focus on more than the dichotomous argument between relatedness and unrelatedness. Many socio-economic factors moderate the effectiveness of these theoretical components within different innovation systems. This implication, to a certain extent, is consistent with the core concept of neo-institutional theories [54] and has been presented in many previous studies (e.g., [55]). Indeed, the reason why many organizations eventually become similar is not because of the presence of market competition but due to their obedience toward the same external institutional pressures.
6. Conclusions and Research Limitations

This study aimed to provide a meaningful analysis and references for understanding how smart specialization policy is utilized within a specific innovation system. With this goal in mind, we further sought to clearly visualize how a specific combination of factors contributes to the development of an innovation system, instead of only highlighting the potentially relevant factors. We argue that the discussion of S3 should consider more socio-economic factors. We used patent data and the data of listed companies from 70 four-digit digital industries in 27 cities in the highly developed Chinese Yangtze River Delta region and performed econometric analyses to capture the impacts of relatedness, knowledge complexity, and their interactions on digital industry innovation. Our findings revealed that Chinese digital industries are more likely to follow the path of diversified relatedness. Unsurprisingly, a positive, direct effect of complexity was also observed. However, in the spatial model, the effect of complexity was neither more positive nor significant than relatedness. Similarly, the effect of complexity was less straightforward than relatedness in the threshold model. We determined that efficient cooperation between relatedness and complexity can be only achieved if the level of government intervention is moderate.

Ultimately, this study has important policy implications. Beyond the scope of “diversification or un-diversification”, policy makers should clearly understand the role of political, cultural, socio-economic, and institutional quality within an innovation system. These factors could serve as a mirror for the implementation of the S3 strategy. Our study shows that the Chinese context (e.g., digital industries) is sensitive to institutional constraints such as the level of government intervention. Thus, increasing relatedness or complexity levels appears to be a short-term solution for the development of regional innovation systems. Attention should be given to exploiting opportunities for creating efficient coordination between relatedness and complexity within a system. Finally, it is necessary to give full play to the role of central cities in developing communication networks to jointly increase the level of knowledge complexity in the context of digital industry development in the Yangtze River Delta, as the performance of major cities in innovation and relatedness was found to significantly outperform that of surrounding cities, with an observable spillover effect. However, a great disparity and inconsistency in policy making was observed in terms of complexity.

This study also has several limitations. First, due to data convenience, the level of relatedness and complexity was only estimated based on a single-year dataset. Thus, only a cross-section regression analysis was applicable. In the future, it would be ideal to update our dataset using a panel structure. Second, this study focused only on the feature of knowledge embodied in patents. However, smart specialization involves diversification toward other functions and cannot be fully explored using only knowledge captured by patents. Finally, our presentation of applications within an innovation system applies only to the context of digital industries. Future studies should also aim at exploring the effectiveness and applicability of S3 theory in other industrial categories.

Author Contributions: Conceptualization, K.Z.; methodology, Z.Y.; software, S.Y.; validation, Z.Y.; formal analysis, S.Y.; writing—original draft preparation, S.Y.; writing—review and editing, M.Z.; supervision, K.Z.; funding acquisition, K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Social Science Post-Funded Project of China, National Office for Philosophy and Social Sciences (Funding Number: 22FJYB008)

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author [Kai Zhao] upon reasonable request.
References

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.