Papers in Evolutionary Economic Geography

# 16.27

# Industrial diversification in Europe: The differentiated role of relatedness

Jing Xiao, Ron Boschma, Martin Andersson

Utrecht University
Urban & Regional research centre Utrecht

# Industrial diversification in Europe:

## The differentiated role of relatedness

Jing Xiao*, Ron Boschma*# and Martin Andersson*$

* Center for Innovation, Research and Competence in the Learning Economy (CIRCLE), Lund University, PO Box 117, 22100 Lund, Sweden (Email: jing.xiao@circle.lu.se)

# Urban and Regional Research Centre Utrecht (URU), Utrecht University, PO Box 80 115, 3508 TC Utrecht, the Netherlands (Email: r.boschma@geo.uu.nl)

$ Department of Industrial Economics, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden (Email: martin.andersson@bth.se)

### Abstract

There is increasing interest in the drivers of industrial diversification, and how these depend on economic and industry structures. This paper contributes to this line of inquiry by analyzing the role of relatedness in explaining variations in industry diversification, measured as the entry of new industry specializations, across 173 European regions during the period 2004-2012. There are significant differences across regions in Europe in terms of industrial diversification. Relatedness has a robust positive influence on the probability that new industry specialization develops in a region. A novel finding is that the influence of relatedness on the probability of new industrial specializations depends on innovation capacity. We find that relatedness is a more important driver of diversification in regions with a weaker innovation capacity. The effect of relatedness appears to decrease monotonically as the innovation capacity of a local economy increases. This is consistent with the argument that high innovation capacity allows an economy to 'break from its past' and to develop, for the economy, truly new industry specializations. We infer from this that innovation capacity is a critical factor for economic resilience.

**Key words:** industrial diversification, related diversification, evolutionary economic geography, unrelated diversification, European regions, resilience

**JEL codes:** B52 L16 O14 O18 R11

## 1. Introduction

History shows that all economies – at level of countries, regions as well as cities – are inevitably confronted with activities (firms, industries) that face stagnation and decline. The decline of car manufacturing in Detroit is a strong reminder of the potentially devastating consequences associated with a decline of an economy's principal economic specialization. There is therefore a constant pressure to develop new economic specializations. This has led to a search for factors that stimulate processes of industrial diversification, because diversification is seen as an important way in which the industrial base of an economy is renewed and broadened.

The bulk of the literature on diversification has focused on the sub-national scale, in particular at the level of cities or regions. One reason for this is that issues associated with diversification and potential consequences of closure of plants or decline of industries are more pronounced at the sub-national level. Local economies, such as cities or regions, are for example often more specialized than whole countries. Therefore, they are more dependent on one or a few industries, firms or activities. There are also significant differences in industrial diversification even between regions or cities that operate under similar national institutional framework conditions. This has led to a general interest in the role that local economic and industrial structures play in fostering (or hampering) processes of industrial diversification. One of the most recent insights from this line of research is that industries are more likely to enter and develop in a region when they are related to pre-existing industries in that region (e.g. Neffke et al. 2011). Similarly, new technologies are more likely to occur in regions with an already established presence of related technologies (e.g. Kogler et al. 2013; Rigby 2015).

However, there is still little systematic understanding of what characteristics or structures of local economies that spur diversification, and how they influence the novelty and nature of diversification processes (Boschma 2016). For example, are local economies with higher innovation capacity in a better position to diversify? Do they have more of a tendency to diversify in more unrelated activities, i.e. are they in a better position to 'break from their past'? Is the capacity of diversification related to the density or centrality of a local economy? There is an increasing number of case studies on new path creation in single regions (see e.g. Isaksen 2015), but no studies yet exist that compare the intensity and type of diversification in different types of regions simultaneously.

This paper aims to fill this gap by doing a study on the capacity of 173 European regions to develop new industrial specializations in the period 2004-2012. We test whether relatedness is a crucial driving force behind industrial diversification across regions in Europe. More in particular, we test whether the diversification patterns differ across European regions with different economic and industrial structures. To this end, we distinguish between (i) core knowledge regions, (ii) manufacturing regions and (iii) peripheral regions in the EU. This categorization is aimed to reflect innovation capacity, industry structure and centrality, respectively, in broad terms. It allows us to analyze whether the influence that relatedness has on the entry of new industry specializations differ across different categories of regions. A basic hypothesis is that relatedness is more important in regions with lower innovation capacity and more limited overall knowledge resources. Local economies with stronger innovation and knowledge resources are supposedly in a better position to develop, for the region, truly new industries and break from their past industry specializations.

Our main findings are as follows: First, we document significant variations across regions in Europe in terms of industrial diversification processes. Second, relatedness appear to be an important determinant of the probability that a new industrial specialization develops in a local economy. Third, the effect of relatedness does indeed vary across regions in a systematic way. We find evidence in favor of that the effect that relatedness has on diversification decreases monotonically as the innovation capacity of a region increases. The effect of relatedness is weakest in the core knowledge regions of Europe. We interpret this as that high innovation capacity put local economies in a better position to 'break from their past', and to develop truly new industry specializations that are less related to their current (or previous) industry structures. Local innovation capacity is from this perspective an important determinant for resilience.

The paper is structured as follows. Section 2 describes briefly the current state-of-affairs in the regional diversification literature. Section 3 introduces the data and methodology, section 4 presents the main findings on diversification of European regions. Section 5 concludes.

## 2. Towards a territory-specific treatment of industrial diversification

In the aftermath of the economic crisis, more urgency is felt than ever to have a basic understanding of the process of industrial diversification. As economies can be hit severely by sudden or slow-burning shocks, there is a constant pressure on countries and regions to develop new economic activities that absorb redundant capital and labour and create new job opportunities. There is a rapidly expanding literature that focuses explicit attention on country- and region-specific capabilities that are considered a key source of industrial diversification (Hidalgo et al. 2007; Neffke et al. 2011; Rigby 2015). What this literature shows is that territory-specific capabilities provide opportunities to develop new industries but also set limits to this process of structural change. If a city or region does not possess the capabilities required for a new specific activity, it is almost impossible to develop it.

The literature has applied the notion of capabilities both conceptually and empirically in its relatedness concept (Boschma 2016). Relatedness refers to capabilities that different economic activities share in terms of similarity and/or complementarity (Breschi et al. 2003; Makri et al. 2010;

Broekel and Brachert 2015). Similarity means that economic activities share resources, like a common knowledge base, that may induce knowledge spillovers and interactive learning across activities. Complementarity means that an activity requires complementary resources from other activities that needs to be combined to make a (new) product, etc. This literature claims that economies are more likely to diversify into new activities that are related to existing activities, so they can draw on and exploit their underlying capabilities. As such, diversification processes in local economies are depicted as an emergent branching process (Frenken and Boschma 2007) in which new activities build on and combine related local activities (Martin and Sunley, 2006; Fornahl and Guenther 2010).

There has been a recent upsurge of studies that have confirmed the predominance of related diversification. In particular, these studies tend to focus on regions and show that industries are more likely to enter and more likely to survive in a region when related to existing industries in that region (see e.g. Neffke et al. 2011; Boschma et al. 2013; Essletzbichler 2015; He and Rigby 2015). The same is true for new technologies which are more likely to occur in regions when related technologies are locally present (e.g. Kogler et al. 2013; Colombelli et al. 2014; Heimeriks and Boschma 2014; Van den Berge and Weterings 2014; Boschma et al. 2015; Feldman et al. 2015; Rigby 2015; Tanner 2015). In other words, related diversification is a dominant pattern in many regions. This is not unexpected, as new capabilities required for related diversification are easier to acquire and less costly when being close to existing local capabilities (Saviotti and Frenken 2008). For instance, it is easier for regions to diversify into trucks when specialized already in motor bikes, as both industries build on the same engineering capability base. Unrelated diversification is a more exceptional event, as it requires the build-up of completely new capabilities that is accompanied with fundamental uncertainty and high costs. For instance, it is extremely complex to diversify into pharmaceuticals when specialized in aerospace as both activities are not related: the distance between their underlying capability bases is just too large (Neffke et al. 2015).

So, what these studies show is that relatedness is a key driver behind industrial diversification, but this finding is typically an average effect across many different regions. What these studies also tell is that relatedness is not a necessary condition for successful diversification, as unrelated diversification also happens now and then. This makes the question relevant whether relatedness is a driving force in every region. Some studies have identified notable differences between regions because their institutions differ. For instance, Cortinovis et al. (2016) found that bridging social capital (as opposed to bonding social capital) in regions is an enabling factor for regional diversification in the EU, especially where formal institutions are weak. Boschma and Capone (2015b) found that East European countries are more likely to diversify into new industries that are strongly related to their existing industries, in contrast to West European countries. And Boschma and Capone (2015a) found that institutions that coordinate less tightly labor, capital and product markets give countries more freedom to diversify in more unrelated activities. This is different from national institutions that regulate more tightly market relations because they make countries to rely more on related diversification.

Isaksen and Trippl (2014) have linked conceptually three types of regional innovation systems (RIS), after Todtling and Trippl (2005), to the question whether regions are more likely to develop new growth paths, and if so, whether regions focus on new path creation versus path renewal. Broadly speaking, new path creation reflects the unrelated diversification type, while path renewal can associated with the related type of diversification. They expect that new path creation and path renewal are more typical patterns in organizationally thick and diversified RIS, because such regions offer a rich and diverse environment. Organizationally thick and specialized RIS are dominated by highly specialized industrial and institutional structures and inward-looking networks, as is common in many old industrial regions. This type of RIS is perceived to have a weak capacity to develop new growth paths, and therefore more likely to rely on existing activities and path extension. Regions with organizationally thin RIS have a weak absorptive capacity, little local knowledge exchange, and closed social networks that tend to lead to conformity. These regions are more likely to experience path extension and, worse, path exhaustion, due to negative lock-in. In sum, Isaksen and Trippl (2014) expect both related and unrelated diversification to take place only in the first type of RIS, while the two other types of RIS are unlikely to experience diversification because of path lock-in.

So, little attention has yet been given to the intensity and nature of diversification in regions with different economic and industrial structures (Boschma 2016). What characteristics of local economies stimulate diversification? Do centrality and innovation capacity fuel a local economy’s capacity to diversify in new industries? Do the same characteristics facilitate more genuine renewal, for example by developing industry specializations in unrelated activities? There are a number of case studies in single regions (see e.g. Isaksen 2015), but no studies yet exist that compare the intensity and type of diversification between different types of regions. Using data on regions in Europe, we explore in this paper the role that economic and industrial characteristics of local economies (regions) play in explaining and stimulating processes of industrial diversification, related as well as unrelated diversification. In particular, we focus on the urban-peripheral nature of regions and their innovation capacity, and make a distinction between core knowledge regions, manufacturing regions and peripheral regions.

## 3. Data and methodology

### 3.1 Data

To measure relatedness and industrial diversification, we use employment data from Orbis database collected by Bureau Van Dijk. The Orbis database contains unique annual firm-level statistics, such as employment, industrial affiliation, and location, covering about 10 million firms across the Europe. After a substantial cleansing and geo-coding process, the original dataset was aggregated into 260 European NUTS2 regions and 615 4-digit NACE sectors (version 2) for the period 2004 to 2012. We dropped some countries that are most affected by the problem of missing values in employment or some small countries with only one NUTS2 (2010 classification) level region.[1] The final dataset contains 173 NUTS2 regions in 12 countries, including Belgium, Bulgaria, Germany, Denmark, Spain, France, Greece, Italy, the Netherlands, Poland, Portugal and Romania.[2] According to the geographical grouping by the United Nations Statistics Division, we formally distinguish these countries among: western European countries (Belgium, Germany, France and the Netherlands), eastern European countries (Bulgaria, Poland and Romania), northern European country (Denmark) and southern European countries (Spain, Greece, Italy and Portugal). As our data only contain one country in northern Europe, we combine western and northern European countries in one country group.

We aim to explore whether the urban-peripheral nature of regions and their innovation capacity have an effect on regional diversification. To indicate innovation capacity at regional level, we follow the OECD approach by Marsan and Maguire (2011) that employs a cluster analysis of regions based on socio-demographic, economic, and innovation-related factors. This typology provides a holistic assessment of multiple dimensions of regional characteristics. In this way, regions can be compared with their peers by simultaneously accounting for not only innovation capacity but also socio-demographic and economic factors.

Following this approach, we divide the regions in our dataset into seven groups[3] that are further summarized into three macro categories: knowledge hubs, industrial production zones, and non-S&T(Science and Technology)-driven regions. In order to focus on the industrial diversification process of regions with high innovation capacity, the category of knowledge hubs are further distinguished between two peer groups: knowledge-intensive city/capital districts and knowledge and technology hubs. Figure 1 displays the maps by region category. Moreover, we report the number of regions by country group or region category/group respectively in Table 1. Please notice that there are only 156 regions in our dataset matched with the typology by Marsan and Maguire (2011). The non-matched regions are grouped into "other regions". Among the 156 matched regions, about 15% of them are knowledge hubs, 50% industrial production zones, and 35% of non-S&T-driven regions.

[1] We keep countries with more than one NUTS2 regions in order to obtain variations within each country.
[2] A full list of 173 NUTS2 regions is shown in Table B1 in the appendix.
[3] Marsan and Maguire (2011) originally identify eight peer groups. In our analysis, we exclude one peer group – US states with average S&T performance as it does not apply to our context.

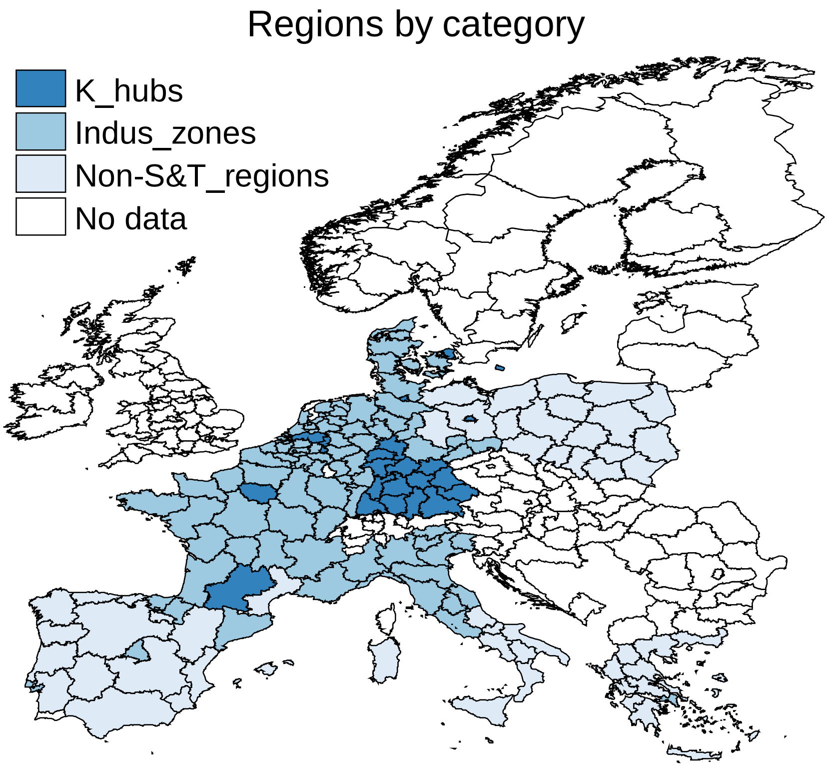


***Figure 1.*** *Map for European regions by region category or group.*

According to Marsan and Maguire (2011), knowledge hubs have the highest performance in economic- and innovation-related indicators. Within the category of knowledge hubs, the group of knowledge-intensive city/capital districts is special: they are a distinctive group of regions with extremely high levels of innovation related variables, urbanization and GDP per capita. Our dataset contains four NUTS2 regions in the group of knowledge-intensive city/capital districts: the Brussels Capital Region, Berlin, Bremen and Hamburg. In the group of knowledge and technology hubs, our dataset covers 19 regions, including Baden-Württemberg, Bavaria and Hesse regions in Germany, Ile-de-France (Paris) and Midi-Pyrénées regions in France, the southern Netherlands region and the capital region of Copenhagen in Denmark. From Figure 1 and Table 1, it is noteworthy that knowledge hubs are all located in western and northern European countries.

Industrial production zones, however, are characterized as regions with a high level of agglomeration activities of production but lagging behind to regions on the innovation frontier (Marsan and Maguire 2011). From Figure 1 and Table 1, we notice that industrial production zones are mainly located in western and northern European countries. Industrial production zones are also found in southern European countries, but only account for a small share compared to that in western and northern European countries.

Non-S&T-driven regions, by contrast, are characterized as peripheral regions with the lowest level performance on innovation-related indicators. Figure 1 and Table 1 show that non-S&T-driven regions are mainly located in southern and eastern European countries. This distribution of regions is not unexpected, as the typology by Marsan and Maguire (2011) is developed according to indicators that reflect regional levels of economic development and innovation capacity.

As this paper explores the acquisition of new specialized industries, we only include tradable sectors in the final dataset. Tradable sectors are identified as the sectors listed in Standard International Trade Classification (SITC; version 3). By matching SITC3 sectors with NACE2 sectors, we remain 323 tradable sectors in the final dataset.[4] Each sector is grouped separately into manufacturing, service or other sectors[5].

We are interested in the question whether the innovation capacity of regions has an impact on the effect of relatedness on regional diversification. This may be reflected in the industrial composition in regions. Scholars (Heidenreich, 2009; Kirner *et al.*, 2009; Santamaria *et al.*, 2009) have argued that inter-industry knowledge spillovers are especially relevant for high-tech sectors, as they rely heavily on knowledge-related inputs. Hartog et al. (2012) found that the effect of related variety on regional growth is conditioned by the technological intensity of local sectors: related variety among high-tech sectors in a region enhances regional employment growth, in contrast to related variety in medium and low-tech sectors. We investigate whether related diversification is more likely to occur in more knowledge-intensive industries, as compared to knowledge-extensive industries. We follow the OECD classification (Hatzichronoglou 1997; Eurostat 2015) and divide manufacturing and service sectors into two broad categories: high-tech manufacturing, medium high-tech manufacturing and knowledge-intensive service (HM-KIS) sectors, and medium low-tech manufacturing, low-tech manufacturing and less knowledge-intensive service sectors (LM-LKIS). Table 2 displays the number of industries by manufacturing/service/other and by industry categories.

[4] A full list of industries is shown in Table B2 in the appendix.
[5] Other sectors refer to industries in Section A (Agriculture, forestry and fishing), Section B (Mining and quarrying), Section D (Electricity, gas, steam and air conditioning supply), Section E (Water supply; sewerage, waste management and remediation activities) and Section F (Construction) in NACE2 classification.

Table 1 The number of regions by country group or region category/group

| | Knowledge hubs | | Industrial production zones | Non-S&T-driven regions | Sub-total | Other regions | Total |
|---|---|---|---|---|---|---|---|
| | Knowledge-intensive city/capital districts | Knowledge and technology hubs | | | | | |
| Western European countries | 4 | 18 | 56 | 4 | 82 | 1 | 83 |
| Estern European countries | - | - | - | 16 | 16 | 14 | 30 |
| Northern European countries | - | 1 | 4 | - | 5 | - | 5 |
| Southern European countries | - | - | 18 | 35 | 53 | 2 | 55 |
| Total | 4 | 19 | 78 | 55 | 156 | 17 | 173 |

Note: Only 156 regions in our data are matched with the OECD typology by Marsan and Maguire (2011). "Other regions" refer to the non-matched regions. The non-matched regions are mainly in Bulgaria (6 regions) and Romania (8 regions), which are not included in the country sample in Marsan and Maguire (2011). Please refer to Table B1 in the appendix for more detailed information of non-matched regions.

Table 2 The number of industries by manufacturing/service/other or by industry category

| | HM-KIS sectors | LHM-LKIS sectors | Other sectors | Total |
|---|---|---|---|---|
| Manufacturing sectors | 70 | 152 | - | 222 |
| Service sectors | 22 | 13 | - | 35 |
| Other sectors | - | - | 66 | 66 |
| Total | 92 | 165 | 66 | 323 |

Note: "Other sectors". Other sectors include industries in Section A (Agriculture, forestry and fishing), Section B (Mining and quarrying), Section D (Electricity, gas, steam and air conditioning supply), Section E (Water supply; sewerage, waste management and remediation activities) and Section F (Construction) in NACE2 classification.

## 3.2 Variables

### 3.2.1 The measure of regional specialization

Following other studies on regional diversification (e.g. Neffke et al. 2011; Boschma et al. 2013), we identify entry of a new specialized industry in a region by observing regional specialization dynamics in a 5-year interval. We compare the regional specialization status of industry *i* between year *t* and *t-5*. If industry *i*, which is not specialized in region *c* at year *t-5*, is found to be specialized in region *c* at year *t*, we identify industry *i* as a new industry that enters into the specialization portfolio in region *c* between year *t-5* and *t*. We use location quotients (LQ) to measure regional specialization, and compare the share of employment of industry *i* in region c relative to the share of overall employment of industry *i* in all regions, as in Equation (1).

$$LQ_{ic} = \left(\frac{E_{ic}/E_{*c}}{E_{i*}/E_{**}}\right) \tag{1}$$

where *E* refers to employment; the subscripts *i* and *c* refer to industry *i* and region *c* respectively; and the subscript * refers to all industries or all regions included in the analysis. A higher LQ means a comparative over-presence of industry *i* in region *c* compared to all regions. But how high of a LQ is enough to identify a specialized industry in a region? The lack of a widely accepted cut-off value of LQ is one main criticism when it comes to the use LQ to identify agglomeration activities (O'Donoghue and Gleave 2004).

In this context, we employ a bootstrap method developed by Tian (2013), to estimate the statistically significant cut-off value of LQ for each industry. Industry *i* is defined to be specialized by region *c* if the standardized location quotient (SLQ) in region *c* is higher than the statistically significant cut-off values of SLQ for industry *i*, which is obtained from the bootstrap resampling process.[6] This method solves the drawback that LQ fails to consider the absolute scale of local industries (O'Donoghue and Gleave 2004) and also does not need any assumption about the statistical distribution of LQ (Tian 2013).

### 3.2.2 The measure of relatedness with existing industries in a region

We employ a proximity approach to measure the relatedness with existing industries in a region. This approach is developed by Hidalgo et al. (2007) and widely used by studies in related diversification (see e.g., Hausmann and Klinger 2007; Hausmann and Hidalgo 2010; Neffke et al. 2011; Boschma

[6] See Cortinovis et al. (2016) for detailed description of the calculation process.

and Capone 2015b). The proximity approach is based on a co-occurrence analysis, in which relatedness between products or industries is revealed by a likelihood of the co-occurrence of two products or industries in the same region. The first step to construct the measure of relatedness with existing industries in a region is to calculate the proximity between each pair of industries. In order to rule out that the likelihood of the co-occurrence of two industries in a region is misled by the overall prevalence of employment in some regions or the large size of some industries (Hausmann and Klinger 2007), we take the minimum conditional probability that a region has a specialization of one industry given its co-specialization of another, as in Equation (2).

$$\varphi_{i,j,t} = \min\{P(x_{i,t}|x_{j,t}), P(x_{j,t}|x_{i,t})\} \tag{2}$$

where φ is the proximity index. In this way, we get a 323-by-323 matrix of proximity indexes based on the co-occurrence analysis of 173 regions. The second step is to link the proximity with the regional structure of industrial specialization by constructing a density indicator, developed by Hausmann and Klinger (2007), as in Equation (3).

$$d_{i,c,t} = \left(\frac{\sum_k \varphi_{i,k,t}\, x_{k,c,t}}{\sum_k \varphi_{i,k,t}}\right)$$

(3)

where the subscript *i* refers to the focus industry; $x_{k,c,t}$ takes a value of 1 when industry *k* is specialized in region *c*. The density indicator is the share of proximities of industry *i* to all industries *k that are* specialized in region *c* at year *t*, in the total proximities of industry *i* to all the industries *k* that are included in the analysis at year *t.* The density indicator is both industry- and region-specific and varies from 0 to 1. A higher density indicator means a higher level of relatedness of industry *i* with the industrial specialization portfolio of region *c* at year *t*.

## 3.3 Descriptive statistics

Descriptive statistics of the main variables are displayed in Table A1 in the Appendix. This paper focuses on the regional diversification process. Therefore, only industries that are already present in each region but have not yet been specialized in each region at the beginning of each 5-year interval are included in the analysis. Category dummy variables of industries and regions are time-invariant. Specialization status and density indicator are time-varying from one 5-year interval to another. Specialization status is measured at the end of each 5-year interval, and the density indicator is measured at the beginning of each 5-year interval. There are in total 135,871 industry-region observations from 2009 to 2012[7], and we have some missing values for region category dummy variables. The correlation coefficients among the main variables are reported in Table A2 in the appendix.

We divide our observations into four groups according to the quartile of density indicator. We calculate the probabilities of acquiring new specialized industries for each quartile of density by dividing the number of new specialized industries by their respective specialization opportunities (the number of industries which are already present in each region but have not yet been specialized in each region at the beginning of each 5-year interval). In this way, we depict how the probabilities of acquiring new specialized industries change as the density indicator increases and how the pattern changes over industries and regions. As shown in the left graph of Figure 2, we find that the probabilities increase as the density increases in general. The probability in the highest quartile is about three times higher than that in the lowest quartile. The pattern that the probabilities of acquiring new specialized industries increase as density increases is further confirmed over industry categories (the middle graph of Figure 2) and over region categories (the right graph of Figure 2).

To further probe the main patterns of entry across regions in Europe, we also recognize that there may be some issues with comparing average entry probabilities across regions directly. It could imply that we overestimate diversification activity in regions with a low level of entry opportunities. To account for the potential impact of the differences in industrial structures across regions and in the relative intensity of entry opportunity across industries,[8] following Audretsch and Fritsch (2002), we calculate a sector-adjusted entry number based on a shift share approach. To obtain the "pure" variation in terms of diversification intensity across regions, the raw entry number is adjusted by the expected entry number which is calculated based on the assumption of an identical industrial structure across regions (see Appendix G for details).

Based on the sector-adjusted entry number during the period of 2004-2009, the top three regions are Ile-de-France (Paris) in France, Trento in Italy and Oberbayern (München) in Germany and the bottom three regions are Mazowieckie in Poland, Weser-Ems in Germany and Rhône-Alpes in France. When we look at the average sector-adjusted entry number in the same period by region category, we find that knowledge hubs have the highest sector-adjusted entry number, which is about 1.5 times as those

[7] We focus on the years from 2009 to 2012 so that our data are long enough to compare the dynamics of regional specialization between year *t* and year *t-5*.
[8] Our formal econometric analysis in Section 4 does not suffer from the problem as it is based on industry level and we include both region-year and industry-year dummy variables to control for time-varying heterogeneity across regions or industries.

of the other two categories of regions. The average sector-adjusted entry numbers of industrial production zones and non-S&T-driven regions are quite close.

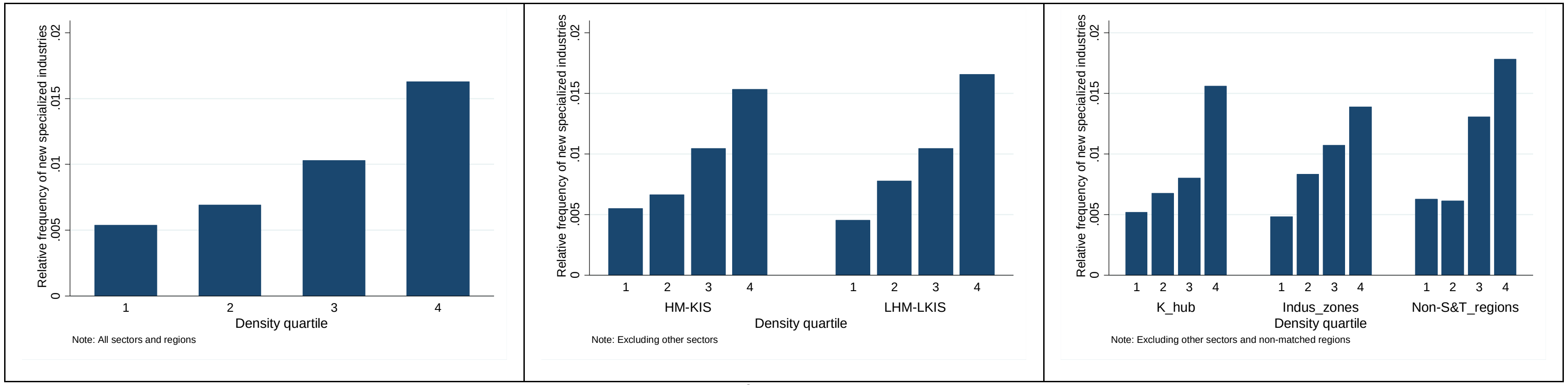


***Figure 2.*** *Probabilities of acquiring new industrial specializations.*[9]

[9] The division between HM-KIS and LHM-LKIS sectors only apply to manufacturing and service sectors. Thus, we exclude other sectors in the middle graph of Figure 1 2 where we focus on the pattern over sectors (between HM-KIS and LHM-LKIS sectors). Moreover, some regions in our dataset are not matched to any region categories by the OECD typology by Marsan and Maguire (2011). Thus, we further exclude the non-matched regions in the right graph of Figure 1 2 where we focus on the pattern over regions.

## 4. Regression analysis

We employ regression analysis to detect the effect of the density indicator on regional diversification by controlling for possible confounding factors. The basic model is displayed in Equation (4):

$$y_{i,c,t} = \alpha + \beta * d_{i,c,t-5} + \gamma_{c,t} + \theta_{i,t} + \varepsilon_{i,c,t}$$
(4)

where $y_{i,c,t}$ is specialization status of industry *i* in region *c* at year *t*, with 1 indicating that industry *i* is specialized in region *c* at year *t*, and 0 otherwise; $d_{i,c,t-5}$ is the density indicator of industry *i* in region *c* at year *t-5;* $\gamma_{c,t}$ and $\theta_{i,t}$ are region-year and industry-year dummy variables which are used to control for time-varying heterogeneity across regions or industries. We use the linear probability model instead of logit or probit models for estimation. As the logit or probit model may lead to bias or inconsistency when they estimate the model with a large amount of dummy variables (Greene 2012; Boschma et al. 2013). The density indicator is standardized before it enters into regressions. We report heteroskedasticity-consistent standard errors for each regression.

### 4.1 Results

Table 3 reports the results of the effects of density indicator on acquiring new industrial specializations. The first panel of Table 3 includes all industries and regions. As expected, density exhibits a significantly positive effect on acquiring the specialization of a new industry in the future. The positive effect of density is further confirmed across different specifications in the second panel of Table 3 where we focus on manufacturing and service sectors. It is noteworthy that the magnitude of density indictor is lower in the second panel which implies that density plays a more important role in acquiring new industries for other sectors than for manufacturing and service sectors. In Specification (2) of the second panel, we include an additional interaction term between density indicator and the dummy variable of HM-KIS sectors. The significantly positive coefficient of the interaction term shows that the positive effect of density is higher in HM-KIS sectors than in LHM-LKIS sectors. In Specification (3) of the second panel, we include additional interaction terms between density indicator and region categories, taking the interaction term between density and the dummy variable of non-S&T-driven regions as the reference group. The interaction term between density and the dummy variable of industrial production zones has a statistically negative coefficient, indicating that the effect of density is lower in industrial production regions than in non-S&T-driven regions. The coefficient of the interaction term between density and knowledge hubs is positive but not statistically significant.

Table 3 The effects of density indicator on acquiring new industrial specializations

| Variables | All | Excluding other sectors | | |
|---|---|---|---|---|
| | | (1) | (2) | (3) |
| Density | 0.00506*** | 0.00473*** | 0.00404*** | 0.00488*** |
| | (0.000444) | (0.000490) | (0.000579) | (0.00102) |
| Density*HM-KIS sectors | | | 0.00189** | |
| | | | (0.000902) | |
| Density*Knowledge hubs | | | | 1.97e-05 |
| | | | | (0.00150) |
| Density*Industrial production zones | | | | -0.00203* |
| | | | | (0.00117) |
| Region-year dummies | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes |
| Constant | 0.0500** | 0.0599** | 0.0475** | 0.0730*** |
| | (0.0211) | (0.0250) | (0.0226) | (0.0278) |
| Observations | 135,871 | 114,408 | 114,408 | 101,351 |
| R-squared | 0.022 | 0.022 | 0.022 | 0.024 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

To further confirm that density plays a differentiated role across industry and region categories, we repeat the estimations separately by manufacturing/service or by industry category (reported in Table 4) and by country group or region category (reported in Table 5). The results in Table 4 clearly show that, first, the positive coefficient of density is slightly higher in service sectors than in manufacturing sectors. Second, density has a much higher positive effect in HM-KIS sectors than in LHM-LKIS sectors: the coefficient of density for HM-KIS sectors is almost twice as large as that for LHM-LKIS sectors. From Panel "Country group" of Table 5, we find the coefficient of density is only slightly higher for western and northern European countries than for southern European countries. But the coefficient of density is much higher for eastern European countries, almost twice as large as those for the other two country groups. In Panel "Region category" of Table 5, we find the effect of density is highest for non-S&T-driven regions and lowest for knowledge hubs. But the difference of the density effects between knowledge hubs and industrial production zones is marginal.

Table 4 The effects of density indicator on acquiring new industrial specializations: by manufacturing/service or industry category

| Variables | Manufacturing/service | | Industry category | |
|---|---|---|---|---|
| | Manufacturing | Service | HM-KIS | LHM-LKIS |
| Density | 0.00444*** | 0.00497*** | 0.00633*** | 0.00333*** |
| | (0.000528) | (0.00133) | (0.000818) | (0.000612) |
| Region-year dummies | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes |
| Constant | 0.0160 | -0.00176 | 0.00721 | 0.0190 |
| | (0.0176) | (0.00319) | (0.00536) | (0.0220) |
| Observations | 96,464 | 17,944 | 42,120 | 72,288 |
| R-squared | 0.026 | 0.055 | 0.030 | 0.029 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Table 5 The effects of density indicator on acquiring new industrial specializations: by country group or region category

| Variable | Country group | | | Region category | | |
|---|---|---|---|---|---|---|
| | West and North | East | South | K_hubs | Indus_zones | Non-S&T_regions |
| Density | 0.00376*** | 0.00614*** | 0.00356*** | 0.00273** | 0.00284*** | 0.00415*** |
| | (0.000645) | (0.00117) | (0.000909) | (0.00114) | (0.000571) | (0.00104) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.0421* | 0.0155** | 0.0726*** | 0.0259*** | 0.0473** | 0.0450** |
| | (0.0223) | (0.00697) | (0.0277) | (0.00873) | (0.0221) | (0.0226) |
| Observations | 55,932 | 22,451 | 36,025 | 16,607 | 50,816 | 33,928 |
| R-squared | 0.031 | 0.059 | 0.046 | 0.065 | 0.033 | 0.054 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Why does not the role of density in developing new industrial specializations exhibit a clear decreasing pattern as the regional innovation capacity increases? One explanation could be the over-presence of HM-KIS sectors in knowledge hubs. In order to test this, we repeat the estimation in Table 5 by distinguishing the category of HM-KIS sectors from the category of LHM-LKIS sectors. The results are reported in Table 6. It still holds that the effect of density is generally stronger for HM-KIS sectors than for LHM-LKIS sectors. Moreover, the differences of the density coefficients among region categories are smaller in the category of HM-KIS sectors than those in the category of LHM-LKIS sectors. That is to say, the differences of density role among region categories is more obvious for LHM-LKIS sectors than for HM-KIS sectors. But the pattern among region categories is still ambiguous. We find that the density effect for knowledge hubs is lower than that for non-S&T-driven regions but higher than that for industrial production zones in the category of HM-KIS sectors. By contrast, in the category of LHM-LKIS sectors, we find that the coefficient of density decreases as the regional innovation capacity increases but the coefficient of density is not statistically significant for knowledge hubs.

Another explanation could be owing to the high heterogeneity within the category of knowledge hubs. According to the typology of Marsan and Maguire (2011), the category of knowledge hubs contains two groups: knowledge-intensive city/capital districts and knowledge and technology hubs. We repeat the estimation by distinguishing the two peer groups within the category of knowledge hubs. The results are reported in Table 7. From Table 7, it is interesting to notice that density has a much higher coefficient for knowledge-intensive city/capital districts than for knowledge and technology hubs, and in both categories of HM-KIS and LHM-LKIS sectors, although the coefficients are not statistically significant. That is to say, the positive effect of density in the category of knowledge hubs could be mainly attributed to the group of knowledge-intensive city/capital districts but this pattern is not statistically significant.

To sum up, we find that density plays a critical role in developing new industrial specializations in European regions. Over industries, we find that density plays a much higher effect for HM-KIS sectors than for LHM-LKIS sectors. The difference of density effect between manufacturing and service is marginal. Over regions, we find that the effect of density is much higher in eastern European countries relative to other European countries. Moreover, if we exclude the group of knowledge-intensive city/capital districts, we find in general that the density effect monotonically decreases as the regional innovation capacity increases. Over both industries and regions, the differences of the density effect among region categories is more obvious for LHM-LKIS sectors than for HM-KIS sectors.

Table 6 The effects of density indicator on acquiring new industrial specializations: by both industry and region categories

| Variable | HM-KIS | | | LHM-LKIS | | |
|---|---|---|---|---|---|---|
| | K_hubs | Indus_zones | Non-S&T_regions | K_hubs | Indus_zones | Non-S&T_regions |
| Density | 0.00536** | 0.00506*** | 0.00566*** | 0.000724 | 0.00130* | 0.00299** |
| | (0.00210) | (0.000976) | (0.00164) | (0.00122) | (0.000692) | (0.00132) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.00122 | 0.00822 | 0.00198 | -0.000635 | 0.0355 | 0.0928*** |
| | (0.00533) | (0.00591) | (0.00212) | (0.00136) | (0.0221) | (0.0329) |
| Observations | 6,302 | 19,270 | 11,912 | 10,305 | 31,546 | 22,016 |
| R-squared | 0.069 | 0.042 | 0.061 | 0.073 | 0.038 | 0.059 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Table 7 The effects of density indicator on acquiring new industrial specializations: by region group within knowledge hubs

| Variable | HM-KIS | | LHM-LKIS | |
|---|---|---|---|---|
| | K_city/capital districts | K_tec hubs | K_city/capital districts | K_tec hubs |
| Density | 0.0149 | 0.00380* | 0.00879 | -0.00129 |
| | (0.00913) | (0.00224) | (0.00557) | (0.00126) |
| Region-year dummies | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes |
| Constant | -0.00306 | 0.0170** | -0.00459 | 0.00240 |
| | (0.00830) | (0.00862) | (0.00445) | (0.00333) |
| Observations | 1,010 | 5,292 | 1,546 | 8,759 |
| R-squared | 0.280 | 0.079 | 0.292 | 0.083 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

### 4.2 Robustness checks

We conduct four robustness checks to test the sensitivity of our main results. The first check concerns the definition of regional specialization of an industry. As discussed in Section 3.2.1, we identify regional specialization based on the statistically significant cut-off values of SLQ. A potential critique is that the method of LQ may not well capture the dynamics of specialization given that the LQ is a ratio. For example, one region may acquire the specialization of an industry not owing to employment growth in this sector in this region but as a result of the employment decline in this sector in other regions. However, we believe this potential problem is highly reduced as we use the statistically significant cut-off values of LQ, which is a quite strict criterion and thus could identify the regions with highly clustered activities for each sector. Nevertheless, we still conducted a robust check by adding a new criterion when we identify regional specialization: the positive employment growth of each sector in each region within each 5-year interval.[10] Based on this new definition of regional specialization, we re-estimated the model, as reported in Tables C1 and C2 in the appendix which show that our main findings hold.

The second robustness check concerns the stability of the specialization status. For example, we observe that region *c* acquires the specialization of a new industry during the period of 2004-2009. But what if this status of the specialization of the industry is not stable over time? In order to test whether the stability of the specialization status impacts on the results, we sort each industry-region observation by year and observe them from 2004 to 2012. We delete the observations with more than one change in the specialization status during the period. We construct a new sample with only observations with no change or only one change in the status of specialization from 2004 and 2012. Based on the new sample, we re-estimate the model and report the results in Table D1 & D2 in the appendix. Based on the results in Table D1, the effect of density by manufacturing/ service or country groups is not consistent with our main findings. First, from Table D1, the effect of density is not significant in service sectors and lower than that in manufacturing sectors. However, in our main findings, the positive effect of density is higher in service than manufacturing sectors. Second, from Table D1, the magnitude of density is highest in southern European countries and very close to that in eastern European countries. From our main findings, however, the magnitude of density effect is highest in eastern European countries. From Table D2, the main findings in terms of the density effect by both industry and region categories hold.

[10] We do not include the employment growth of each sector as one criterion in the main analysis for two reasons. First, we believe that the consideration of the statistically significant cut-off values of LQ is already a quite strict criterion to identity regional specialization. Second, we use employment data from Orbis to calculate the employment growth of each sector in each region. As discussed in Section 3.1, this data mainly cover big firms and suffer from the problem of missing values.

The third robustness check concerns whether the main results are sensitive to different time periods of measuring the dynamics of regional specialization. As our data is limited in length, we re-construct a new sample with only one interval: 2004-2012. Based on this sample, we re-estimate the model and report the results in Table E1 & E2 in the appendix. It is noteworthy that the findings in terms of the effect of density by manufacturing/ service or country groups does not hold. First, from Table E1, the effect of density is non-significant in service sectors and lower than that in manufacturing sectors. However, our main findings show that the density effect is significantly positive for both manufacturing and service sectors and it plays a more important role in service than manufacturing sectors. Second, from Table E1, the magnitude of density effect in southern European countries is lower but very close to that in eastern European countries. However, according to our main findings, the density effect in southern European countries should be at a similar level with that in western and northern European countries. From Table D2, we find the main findings in terms of the density effect by both industry and region categories hold.

The fourth robustness check concerns whether the main results are sensitive if we add other time-varying control variables at regional level. We re-estimate the model by including a set of time-varying regional-level control variables, retrieved from Cambridge Econometrics regional database and Eurostat regional database, including the average growth rate of GDP per capita within each five-year interval[11], the Los-index[12], population density, levels of Gross Domestic Product (GDP), shares of workers in science and technology (S&T) in active population, and levels of gross capital formation per employee. As the regional control variables are time-varying, in the estimation we only include region dummy variables to control for constant heterogeneity across regions instead of region-year dummy variables which control for time-varying heterogeneity across regions. As reported in Table F1 and F2, our main findings hold.

From the results of the robustness checks, the differentiated role of density in acquiring new industrial specializations over industries and regions is mainly attributed to a core factor – innovation capacity. When we use innovation capacity to distinguish sectors and regions, the role of relatedness density is quite robust.

## 5. Concluding remarks

A robust finding emerging out of many recent studies is that relatedness is a strong driver of industrial diversification. However, this finding tend to be an average effect across many different types of

[11] The average growth of GDP per capital in the last interval is the average growth rate of GDP per capita in a four-year interval as the data of GDP per capita is not available for year 2012.
[12] We calculate the Los-index (Los 2000) based on the 323-by-323 matrix of proximity indexes.

(local) economies. This paper contributed by exploring heterogeneity across European regions in terms of the role of relatedness in explaining industrial diversification, measured as the entry of new industry specializations. The first finding is that relatedness has positive influence on the probability that a new industry specialization develops. This result is robust and holds across all regions under investigation; i.e. the local presence of related activities provides a powerful explanation for what type of new industrial specializations is developed in regions, no matter whether these concern core knowledge regions, manufacturing regions or peripheral regions. A second finding is that the influence of relatedness on the probability of new industrial specializations depends on innovation capacity: relatedness is a more important driver of diversification in regions with a weaker innovation capacity. The effect of relatedness appears to decrease monotonically as the innovation capacity of a local economy increases. This is consistent with the argument that high innovation capacity allows an economy to 'break from its past' and to develop, for the economy, truly new industry specializations. Still, we also find some industrial differences: while relatedness plays a more important role in knowledge-intensive industries than knowledge-extensive industries, the difference of the relatedness effect across European regions is more pronounced for knowledge-extensive sectors.

These findings clearly underscore that the effect of relatedness is not invariant to local conditions. They also call for further investigation. First, there is a need to unravel the specific capabilities that underlie related diversification in a local economy. What enabling factors make some industries more likely to grow out of other industries? Is it because the new industries can build on a similar knowledge base, draw on a shared network, make use of similar institutions, or exploit a common set of local skills? And do regions with different economic and industrial structures differ in that respect? Second, there is a need to be more precise about what makes related diversification different from unrelated diversification. Boschma (2016) have pleaded for an approach that determines what types of new combinations made between related and unrelated industries lead to new industrial specializations. Third, a crucial question is what type of diversification secures long-term economic development. Do specialized local economies need to diversify in unrelated activities to avoid lock-in in the long run, and is there a difference between regions with different economic and industrial structures? Fourth, the regional diversification literature has focused primarily on local capabilities. However, recent evidence suggests that non-local capabilities, besides local capabilities, influence regional diversification (Isaksen 2015; Trippl et al. 2015). This calls for a multi-scalar perspective to assess the relative importance of local and non-local capabilities (Binz et al. 2014). And fifth, there is a need to develop a micro-perspective that identifies the key agents that drive diversification in different types of regions, and determines which regional factors make local actors in some regions successful in inducing institutional change to enable new activities.

## Acknowledgements

We thank Nicola Cortinovis from Utrecht University for sharing the cleansed and geo-coded version of the employment data for European regions. Please see Cortinovis and Oort (2015) for more information in terms of the construction of the dataset. The financial support by the JPI (Joint Programming Initiative) Urban Europe program (the Resilient Cities project) is gratefully acknowledged.

## References


Audretsch, D. B. and Fritsch, M. (2002). Growth Regimes over Time and Space. *Regional Studies.* 36(2), 113-124.

Berge, M. van den, and A. Weterings (2014). *Relatedness in eco-technological development in European regions*, The Hague: Planbureau voor Leefomgeving

Binz, C., Truffer, B., and Coenen, L. (2014). Why space matters in technological innovation systems—Mapping global knowledge dynamics of membrane bioreactor technology. *Research Policy* 43(1), 138-155.

Boschma, R. (2016). Relatedness as driver behind regional diversification: a research agenda, *Regional Studies*, forthcoming.

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

Boschma, R. and G. Capone (2015a). Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework, *Research Policy* 44, 1902-1914.

Boschma, R. and G. Capone (2015b). Relatedness and diversification in the EU-27 and ENP countries. *Environment and Planning C: Government and Policy,* forthcoming.

Boschma, R., 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(1): 29-51.

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

Broekel, T. and M. Brachert (2015). The structure and evolution of inter-sectoral technological complementarity in R&D in Germany from 1990 to 2011 *Journal of Evolutionary Economics* 25, 755-785.

Colombelli, A., J. Krafft and F. Quatraro (2014). The emergence of new technology-based sectors in European regions: a proximity-based analysis of nanotechnology, *Research Policy* 43, 1681-1696.

Cortinovis, N., J. Xiao, R. Boschma and F. Van Oort (2016). Quality of government and social capital as drivers of regional diversification in Europe, Papers in Evolutionary Economic Geography # 16.10, Utrecht, the Netherlands: Urban & Regional Research Centre Utrecht, Utrecht University.

Cortinovis, N. and van Oort, F. (2015). Variety, economic growth and knowledge intensity of European regions: a spatial panel analysis. *The Annals of Regional Science*. 55(1), 7-32.

Essletzbichler, J. (2015). Relatedness, industrial branching and technological cohesion in US metropolitan areas, *Regional Studies* 49 (5), 752–766.

Eurostat (2015). High-tech industry and knowledge-intensive services (htec) Annex 3. Retrieved from http://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf

Feldman, M.P., D.F. Kogler and D.L. Rigby (2015). rKnowledge: The Spatial Diffusion and Adoption of rDNA Methods, *Regional Studies* 49(5), 798-817.

Fornahl, D. and C. Guenther (2010). Persistence and change of regional industrial activities. The impact of diversification in the German machine tool industry, *European Planning Studies* 18 (12), 1911-1936.

Frenken K. and R. A. Boschma (2007). A theoretical framework for economic geography: industrial dynamics and urban growth as a branching process, *Journal of Economic Geography*, 7(5), 635–649.

Greene, W. H. (2012). Econometric Analysis. 7th ed. Harlow, England: Pearson Education Limited.

Hartog, M., R. Boschma and M. Sotarauta (2012). The impact of related variety on regional employment growth in Finland 1993-2006. High-tech versus medium/low-tech, *Industry and Innovation*, 19 (6), 459-476.

Hausmann, R. and Hidalgo, C. A. (2010). Country diversification, product ubiquity, and economic divergence. Working Paper No. 201. Cambridge, Mass.: Center for International Department, Havard University.

Hausmann, R. and Klinger, B. (2007). The structure of the product space and the evolution of comparative advantage (Working paper no. 146). Working Paper No. 146. Cambridge, Mass.: Center for International Development, Harvard University.

Hatzichronoglou, T. (1997). Revision of the High-Technology Sector and Product Classification.

He, C. and D. Rigby (2015). Regional industrial evolution in China: Path dependence or path creation?, working paper.

Heidenreich, M. (2009). Innovation patterns and location of European low-and-medium-technology industries, *Research Policy*, 38, pp. 483–494.

Heimeriks, G. and R. Boschma (2014). The path- and place-dependent nature of scientific knowledge production in biotech 1986-2008, *Journal of Economic Geography* 14, 339-364.

Hidalgo, C. A., Klinger, B., Barabási, A.-L. and Hausmann, R. (2007). The Product Space Conditions the Development of Nations. *Science.* 317(5837), 482-487.

Isaksen, A. (2015). Industrial development in thin regions: trapped in path extension? *Journal of Economic Geography* 15, 585-600.

Isaksen, A. and M. Trippl (2014). Regional industrial path development in different regional innovation systems: A conceptual analysis, Papers in Innovation Studies 2014/17, Lund University, Lund.

Kirner, E., Kinkel, S. and Jaeger, A. (2009). Innovation paths and the innovation performance of low-technology firms—an empirical analysis of German industry, *Research Policy*, 38, 447–458.

Kogler, D.F., D.L. Rigby and I. Tucker (2013). Mapping Knowledge Space and Technological Relatedness in US Cities, *European Planning Studies* 21(9), 1374-1391.

Los, B. (2000). The empirical performance of a new inter-industry technology spillover measure. In P. P. Saviotti and B. Nooteboom. Technology and Knowledge. pp. 118–151. Cheltenham: Edward Elgar.

Makri, M., M.A. Hitt and P.J. Lane (2010). Complementary technologies, knowledge relatedness and invention outcomes in high technology mergers and acquisitions, *Strategic Management Journal* 31, 602-628.

Marsan, G. A. and Maguire, K. (2011). Categorisation of OECD Regions Using Innovation-Related Variables. 2011/03. Paris, France: OECD.

Martin R. and Sunley P. (2006). Path dependence and regional economic evolution, *Journal of Economic Geography* 6(4), 395–437.

Neffke, F., Hartog, M., Boschma, R. and Henning, M. (2015). Agents of structural change. The role of firms and entrepreneurs in regional diversification, Papers in Evolutionary Economic Geography, Utrecht University.

Neffke, F., Henning, M. and Boschma, R. (2011). How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography.* 87(3), 237-265.

O'Donoghue, D. and Gleave, B. (2004). A Note on Methods for Measuring Industrial Agglomeration. *Regional Studies.* 38(4), 419-427.

Rigby, D. (2015). Technological relatedness and knowledge space. Entry and exit of US cities from patent classes, *Regional Studies*, 49 (11), 1922-1937.

Santamaria, L., Nieto, M. J. and Barge-Gil, A. (2009). Beyond formal R&D: taking advantage of other sources of innovation in low- and medium-technology industries, *Research Policy*, 38, 507–517.

Saviotti, P.P. and K. Frenken (2008). Trade variety and economic development of countries. *Journal of Evolutionary Economics*, 18 (2), 201-218.

Tanner, A.N. (2015). The emergence of new technology-based industries: the case of fuel cells and its technological relatedness to regional knowledge bases, *Journal of Economic Geography*, forthcoming, doi:10.1093/jeg/lbv011.

Tian, Z. (2013). Measuring Agglomeration Using the Standardized Location Quotient with a Bootstrap Method. *The Journal of Regional Analysis & Policy.* 43(2), 186-197.

Tödtling F., Trippl M. (2005). One Size Fits All? Towards a Differentiated Regional Innovation Policy Approach. *Research Policy* 34: 1203-1219

Trippl, M., M. Grillitsch and A. Isaksen (2015). External "energy" for regional industrial change: attraction and absorption of non-local knowledge for new path development, Papers in Innovation Studies, no. 2015/47, CIRCLE, Lund University, Lund.

## Appendix Tables

Table A1 Description and summary statistics of main variables

| Variables | Description | Obs | Mean | Median | Std. Dev. |
|---|---|---|---|---|---|
| Spe_t | Dummy variable of specialization at t+5 | 135871 | 0.010 | 0.000 | 0.098 |
| Density_t-5 | Density indicator at t | 135871 | 0.027 | 0.021 | 0.028 |
| HM-KIS | Dummy variable for sectors in high-tech manufacturing, medium-high-tech manufacturing and knowledge-intensive service sectors | 114408 | 0.368 | 0.000 | 0.482 |
| LHM-LKIS | Dummy variable for sectors in medium-low-tech manufacturing, low-tech manufacturing and less knowledge-intensive service sectors | 114408 | 0.632 | 1.000 | 0.482 |
| K_hub | Dummy variable for regions in knowledge hubs | 119825 | 0.160 | 0.000 | 0.366 |
| Indus_zones | Dummy variable for regions in indsyutrial produztion zones | 119825 | 0.500 | 1.000 | 0.500 |
| Non-S&T_regions | Dummy variable for regions in non-S&T-driven regions | 119825 | 0.340 | 0.000 | 0.474 |

Table A2 Correlations of main variables

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Spe_t (1) | 1 | | | | | | |
| Density_t-5 (2) | 0.058 | 1 | | | | | |
| HM-KIS (3) | 0.002 | 0.012 | 1 | | | | |
| LHM-LKIS (4) | -0.002 | -0.012 | -1 | 1 | | | |
| K_hub (5) | -0.005 | -0.025 | 0.009 | -0.009 | 1 | | |
| Indus_zones (6) | -0.005 | 0.003 | 0.020 | -0.020 | -0.436 | 1 | |
| Non-S&T_regions (7) | 0.008 | 0.016 | -0.028 | 0.028 | -0.313 | -0.718 | 1 |

Table B1 The List of Regions (NUTS 2010 Classification)

| NUTS2 | Group | Category | NUTS2 | Group | Category | NUTS2 | Group | Category | NUTS2 | Group | Category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BE10 | 1 | 1 | DEA4 | 4 | 2 | ES53 | 6 | 3 | ITI2 | 5 | 2 |
| BE21 | 4 | 2 | DEA5 | 4 | 2 | ES61 | 6 | 3 | ITI3 | 5 | 2 |
| BE22 | 4 | 2 | DEB1 | 4 | 2 | ES62 | 6 | 3 | ITI4 | 4 | 2 |
| BE23 | 4 | 2 | DEB2 | 4 | 2 | FR10 | 2 | 1 | NL11 | 3 | 2 |
| BE24 | 4 | 2 | DEB3 | 4 | 2 | FR21 | 4 | 2 | NL12 | 3 | 2 |
| BE25 | 4 | 2 | DEC0 | 4 | 2 | FR22 | 4 | 2 | NL13 | 3 | 2 |
| BE31 | 4 | 2 | DED2 | 4 | 2 | FR23 | 4 | 2 | NL21 | 3 | 2 |
| BE32 | 4 | 2 | DED4 | 4 | 2 | FR24 | 4 | 2 | NL22 | 3 | 2 |
| BE33 | 4 | 2 | DED5 | 4 | 2 | FR25 | 4 | 2 | NL23 | 3 | 2 |
| BE34 | 4 | 2 | DEE0 | 6 | 3 | FR26 | 4 | 2 | NL31 | 3 | 2 |
| BE35 | 4 | 2 | DEF0 | 4 | 2 | FR30 | 4 | 2 | NL32 | 3 | 2 |
| BG31 | N/A | N/A | DEG0 | 4 | 2 | FR41 | 4 | 2 | NL33 | 3 | 2 |
| BG32 | N/A | N/A | DK01 | 2 | 1 | FR42 | 4 | 2 | NL34 | 3 | 2 |
| BG33 | N/A | N/A | DK02 | 3 | 2 | FR43 | 4 | 2 | NL41 | 2 | 1 |
| BG34 | N/A | N/A | DK03 | 3 | 2 | FR51 | 4 | 2 | NL42 | 2 | 1 |
| BG41 | N/A | N/A | DK04 | 3 | 2 | FR52 | 4 | 2 | PL11 | 7 | 3 |
| BG42 | N/A | N/A | DK05 | 3 | 2 | FR53 | 4 | 2 | PL12 | 7 | 3 |
| DE11 | 2 | 1 | EL11 | 7 | 3 | FR61 | 4 | 2 | PL21 | 7 | 3 |
| DE12 | 2 | 1 | EL12 | 7 | 3 | FR62 | 2 | 1 | PL22 | 6 | 3 |
| DE13 | 2 | 1 | EL13 | 7 | 3 | FR63 | 4 | 2 | PL31 | 7 | 3 |
| DE14 | 2 | 1 | EL14 | 7 | 3 | FR71 | 4 | 2 | PL32 | 7 | 3 |
| DE21 | 2 | 1 | EL21 | 7 | 3 | FR72 | 4 | 2 | PL33 | 7 | 3 |
| DE22 | 2 | 1 | EL22 | 7 | 3 | FR81 | 6 | 3 | PL34 | 7 | 3 |
| DE23 | 2 | 1 | EL23 | 7 | 3 | FR82 | 4 | 2 | PL41 | 7 | 3 |
| DE24 | 2 | 1 | EL24 | 7 | 3 | FR83 | N/A | N/A | PL42 | 6 | 3 |
| DE25 | 2 | 1 | EL25 | 7 | 3 | ITC1 | 5 | 2 | PL43 | 7 | 3 |
| DE26 | 2 | 1 | EL30 | 4 | 2 | ITC2 | N/A | N/A | PL51 | 6 | 3 |
| DE27 | 2 | 1 | EL41 | 7 | 3 | ITC3 | 4 | 2 | PL52 | 7 | 3 |
| DE30 | 1 | 1 | EL42 | 7 | 3 | ITC4 | 5 | 2 | PL61 | 7 | 3 |
| DE40 | 6 | 3 | EL43 | 7 | 3 | ITF1 | 6 | 3 | PL62 | 7 | 3 |
| DE50 | 1 | 1 | ES11 | 6 | 3 | ITF2 | 6 | 3 | PL63 | 6 | 3 |
| DE60 | 1 | 1 | ES12 | 6 | 3 | ITF3 | 6 | 3 | PT11 | 7 | 3 |
| DE71 | 2 | 1 | ES13 | 6 | 3 | ITF4 | 6 | 3 | PT15 | N/A | N/A |
| DE72 | 2 | 1 | ES21 | 4 | 2 | ITF5 | 6 | 3 | PT16 | 7 | 3 |
| DE73 | 2 | 1 | ES22 | 4 | 2 | ITF6 | 6 | 3 | PT17 | 4 | 2 |
| DE80 | 6 | 3 | ES23 | 6 | 3 | ITG1 | 6 | 3 | PT18 | 7 | 3 |
| DE91 | 4 | 2 | ES24 | 6 | 3 | ITG2 | 6 | 3 | RO11 | N/A | N/A |
| DE92 | 4 | 2 | ES30 | 4 | 2 | ITH1 | 5 | 2 | RO12 | N/A | N/A |
| DE93 | 4 | 2 | ES41 | 6 | 3 | ITH2 | 5 | 2 | RO21 | N/A | N/A |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DE94 | 4 | 2 | ES42 | 6 | 3 | ITH3 | 5 | 2 | RO22 | N/A | N/A |
| DEA1 | 4 | 2 | ES43 | 6 | 3 | ITH4 | 5 | 2 | RO31 | N/A | N/A |
| DEA2 | 4 | 2 | ES51 | 4 | 2 | ITH5 | 5 | 2 | RO32 | N/A | N/A |
| DEA3 | 4 | 2 | ES52 | 6 | 3 | ITI1 | 5 | 2 | RO41 | N/A | N/A |
| | | | | | | | | | RO42 | N/A | N/A |

Note: Region groups and categories are divided according to the OECD typology by Marsan and Maguire (2011). Group: 1 is "Knowledge-intensive city/capital districts" 2"Knowledge and technology hubs" 3 "Service and natural resource regions in knowledge-intensive countries" 4 "Medium-tech manufacturing and service providers" 5 "Traditional manufacturing regions" 6 "Structural inertia or de-industrializing regions" 7 "Primary-sector-intensive regions". Category: 1 "Knowledge hubs"; 2 "Industrial production zones" 3 "Non-S&T-driven regions".

Table B2 The List of Industries (NACE Version 2)

| NACE4 | Category | NACE4 | Category | NACE4 | Category | NACE4 | Category | NACE4 | Category | NACE4 | Category | NACE4 | Category | NACE4 | Category |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0111 | 5 | 0729 | 5 | 1107 | 2 | 1920 | 2 | 2361 | 2 | 2630 | 1 | 2932 | 1 | 3822 | 5 |
| 0112 | 5 | 0811 | 5 | 1200 | 2 | 2011 | 1 | 2362 | 2 | 2640 | 1 | 3011 | 2 | 3831 | 5 |
| 0113 | 5 | 0812 | 5 | 1310 | 2 | 2012 | 1 | 2363 | 2 | 2651 | 1 | 3012 | 2 | 3832 | 5 |
| 0114 | 5 | 0891 | 5 | 1320 | 2 | 2013 | 1 | 2364 | 2 | 2652 | 1 | 3020 | 1 | 4120 | 5 |
| 0115 | 5 | 0892 | 5 | 1391 | 2 | 2014 | 1 | 2365 | 2 | 2660 | 1 | 3030 | 1 | 4322 | 5 |
| 0116 | 5 | 0893 | 5 | 1392 | 2 | 2015 | 1 | 2369 | 2 | 2670 | 1 | 3040 | 1 | 4329 | 5 |
| 0119 | 5 | 0899 | 5 | 1393 | 2 | 2016 | 1 | 2370 | 2 | 2680 | 1 | 3091 | 1 | 4332 | 5 |
| 0121 | 5 | 0910 | 5 | 1394 | 2 | 2017 | 1 | 2391 | 2 | 2711 | 1 | 3092 | 1 | 4391 | 5 |
| 0122 | 5 | 0990 | 5 | 1395 | 2 | 2020 | 1 | 2399 | 2 | 2712 | 1 | 3099 | 1 | 5221 | 4 |
| 0123 | 5 | 1011 | 2 | 1396 | 2 | 2030 | 1 | 2410 | 2 | 2720 | 1 | 3101 | 2 | 5222 | 4 |
| 0124 | 5 | 1012 | 2 | 1399 | 2 | 2041 | 1 | 2420 | 2 | 2731 | 1 | 3102 | 2 | 5811 | 3 |
| 0125 | 5 | 1013 | 2 | 1411 | 2 | 2042 | 1 | 2431 | 2 | 2732 | 1 | 3103 | 2 | 5812 | 3 |
| 0126 | 5 | 1020 | 2 | 1412 | 2 | 2051 | 1 | 2432 | 2 | 2733 | 1 | 3109 | 2 | 5813 | 3 |
| 0127 | 5 | 1031 | 2 | 1413 | 2 | 2052 | 1 | 2433 | 2 | 2740 | 1 | 3211 | 2 | 5814 | 3 |
| 0128 | 5 | 1032 | 2 | 1414 | 2 | 2053 | 1 | 2434 | 2 | 2751 | 1 | 3212 | 2 | 5819 | 3 |
| 0129 | 5 | 1039 | 2 | 1419 | 2 | 2059 | 1 | 2441 | 2 | 2752 | 1 | 3213 | 2 | 5911 | 3 |
| 0130 | 5 | 1041 | 2 | 1420 | 2 | 2060 | 1 | 2442 | 2 | 2790 | 1 | 3220 | 2 | 5912 | 3 |
| 0141 | 5 | 1042 | 2 | 1431 | 2 | 2110 | 1 | 2443 | 2 | 2811 | 1 | 3230 | 2 | 5913 | 3 |
| 0142 | 5 | 1051 | 2 | 1439 | 2 | 2120 | 1 | 2444 | 2 | 2812 | 1 | 3240 | 2 | 5920 | 3 |
| 0143 | 5 | 1052 | 2 | 1511 | 2 | 2211 | 2 | 2445 | 2 | 2813 | 1 | 3250 | 1 | 6209 | 3 |
| 0145 | 5 | 1061 | 2 | 1512 | 2 | 2219 | 2 | 2446 | 2 | 2814 | 1 | 3291 | 2 | 6399 | 3 |
| 0146 | 5 | 1062 | 2 | 1520 | 2 | 2221 | 2 | 2451 | 2 | 2815 | 1 | 3299 | 2 | 7022 | 3 |
| 0147 | 5 | 1071 | 2 | 1610 | 2 | 2222 | 2 | 2511 | 2 | 2821 | 1 | 3311 | 2 | 7111 | 3 |
| 0149 | 5 | 1072 | 2 | 1621 | 2 | 2223 | 2 | 2512 | 2 | 2822 | 1 | 3312 | 2 | 7112 | 3 |
| 0163 | 5 | 1073 | 2 | 1622 | 2 | 2229 | 2 | 2521 | 2 | 2823 | 1 | 3313 | 2 | 7410 | 3 |
| 0164 | 5 | 1081 | 2 | 1623 | 2 | 2311 | 2 | 2529 | 2 | 2825 | 1 | 3314 | 2 | 7420 | 3 |
| 0210 | 5 | 1082 | 2 | 1624 | 2 | 2312 | 2 | 2530 | 2 | 2829 | 1 | 3315 | 2 | 7490 | 3 |
| 0220 | 5 | 1083 | 2 | 1629 | 2 | 2313 | 2 | 2540 | 1 | 2830 | 1 | 3316 | 2 | 7740 | 4 |
| 0230 | 5 | 1084 | 2 | 1711 | 2 | 2314 | 2 | 2561 | 2 | 2841 | 1 | 3317 | 2 | 7990 | 4 |
| 0240 | 5 | 1085 | 2 | 1712 | 2 | 2319 | 2 | 2571 | 2 | 2849 | 1 | 3319 | 2 | 8230 | 4 |
| 0311 | 5 | 1086 | 2 | 1721 | 2 | 2320 | 2 | 2572 | 2 | 2891 | 1 | 3320 | 2 | 8291 | 4 |
| 0312 | 5 | 1089 | 2 | 1722 | 2 | 2331 | 2 | 2573 | 2 | 2892 | 1 | 3511 | 5 | 8299 | 4 |
| 0321 | 5 | 1091 | 2 | 1723 | 2 | 2332 | 2 | 2591 | 2 | 2893 | 1 | 3512 | 5 | 8551 | 3 |
| 0322 | 5 | 1092 | 2 | 1724 | 2 | 2341 | 2 | 2592 | 2 | 2894 | 1 | 3513 | 5 | 9001 | 3 |
| 0510 | 5 | 1101 | 2 | 1729 | 2 | 2342 | 2 | 2593 | 2 | 2895 | 1 | 3514 | 5 | 9002 | 3 |
| 0520 | 5 | 1102 | 2 | 1811 | 2 | 2343 | 2 | 2594 | 2 | 2896 | 1 | 3521 | 5 | 9003 | 3 |
| 0610 | 5 | 1103 | 2 | 1812 | 2 | 2344 | 2 | 2599 | 2 | 2899 | 1 | 3522 | 5 | 9004 | 3 |
| 0620 | 5 | 1104 | 2 | 1813 | 2 | 2349 | 2 | 2611 | 1 | 2910 | 1 | 3523 | 5 | 9512 | 4 |
| 0710 | 5 | 1105 | 2 | 1814 | 2 | 2351 | 2 | 2612 | 1 | 2920 | 1 | 3812 | 5 | 9522 | 4 |
| 0721 | 5 | 1106 | 2 | 1910 | 2 | 2352 | 2 | 2620 | 1 | 2931 | 1 | 3821 | 5 | 9524 | 4 |
| | | | | | | | | | | | | | | 9529 | 4 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | 9602 | 4 |
| | | | | | | | 9609 | 4 |

Note: Industry categories are divided according to the industrial classification of OECD. Category: 1 "High-tech and medium high-tech manufacturing sectors"; 2 "Medium low-tech and low-tech manufacturing sectors"; 3 "Knowledge-intensive service sectors"; 4 "Less knowledge-intensive service sectors"; 3 "Other sectors". Other sectors include industries in Section A (Agriculture, forestry and fishing), Section B (Mining and quarrying), Section D (Electricity, gas, steam and air conditioning supply), Section E (Water supply; sewerage, waste management and remediation activities) and Section F (Construction) in NACE classification.

Table C1 The effects of density indicator: by manufacturing/service or country group - robustness check for adding the criterion of positive employment growth of regional specialization

| Variable | Manufacturing/service | | Counrty group | | |
|---|---|---|---|---|---|
| | Manu | Service | West and North | East | South |
| Density | 0.00339*** | 0.00416*** | 0.00291*** | 0.00489*** | 0.00284*** |
| | (0.000472) | (0.00128) | (0.000596) | (0.00107) | (0.000830) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.0169 | -0.00147 | 0.0428* | 0.0159** | 0.0490** |
| | (0.0175) | (0.00298) | (0.0223) | (0.00691) | (0.0230) |
| Observations | 96,464 | 17,944 | 55,932 | 22,451 | 36,025 |
| R-squared | 0.024 | 0.056 | 0.030 | 0.055 | 0.045 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Table C2 The effects of density indicator: by both industry and region categories - robustness check for adding the criterion of positive employment growth of regional specialization

| Variable | HM-KIS | | | | LHM-LKIS | | | |
|---|---|---|---|---|---|---|---|---|
| | K_hubs | | Indus_zones | Non-S&T_regions | K_hubs | | Indus_zones | Non-S&T_regions |
| | K_city/capital | K_tec hubs | | | K_city/capital | K_tec hubs | | |
| Density | 0.0118 | 0.00349 | 0.00329*** | 0.00535*** | 0.00879 | -0.00192 | 0.000645 | 0.00316*** |
| | (0.00853) | (0.00219) | (0.000779) | (0.00160) | (0.00557) | (0.00119) | (0.000621) | (0.00121) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | -0.00146 | 0.00769 | 0.00342 | 0.00188 | -0.00459 | 0.00242 | 0.0333 | 0.0759*** |
| | (0.00796) | (0.00583) | (0.00454) | (0.00204) | (0.00445) | (0.00331) | (0.0220) | (0.0292) |
| Observations | 1,010 | 5,292 | 19,270 | 11,912 | 1,546 | 8,759 | 31,546 | 22,016 |
| R-squared | 0.283 | 0.078 | 0.041 | 0.061 | 0.292 | 0.083 | 0.037 | 0.057 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Table D1 The effects of density indicator: by manufacturing/service or country group - robustness check for the stability of the specialization status

| Variable | Manufacturing/service | | Counrty group | | |
|---|---|---|---|---|---|
| | Manu | Service | West and North | East | South |
| Density | 0.00214*** | 0.00177 | 0.00129*** | 0.00240*** | 0.00250*** |
| | (0.000397) | (0.00110) | (0.000441) | (0.000923) | (0.000762) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes |
| Constant | -0.00153 | 0.00177 | 0.0333 | 0.00744 | 0.119*** |
| | (0.00108) | (0.00193) | (0.0203) | (0.00486) | (0.0358) |
| Observations | 95,377 | 17,737 | 55,294 | 22,162 | 35,658 |
| R-squared | 0.028 | 0.059 | 0.030 | 0.055 | 0.052 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Table D2 The effects of density indicator: by both industry and region categories - robustness check for the stability of the specialization status

| Variable | HM-KIS | | | | LHM-LKIS | | | |
|---|---|---|---|---|---|---|---|---|
| | K_hubs | | Indus_zones | Non-S&T_regions | K_hubs | | Indus_zones | Non-S&T_regions |
| | K_city/capital | K_tec hubs | | | K_city/capital | K_tec hubs | | |
| Density | 0.00551 | 0.000689 | 0.00185*** | 0.00398*** | 0.00703 | -0.00175* | 0.000770 | 0.00115 |
| | (0.00752) | (0.00129) | (0.000618) | (0.00139) | (0.00477) | (0.00106) | (0.000469) | (0.00113) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | -0.00471 | 0.00251 | 0.00196 | 0.000499 | -0.00171 | -0.000188 | 0.0505* | 0.132*** |
| | (0.00554) | (0.00341) | (0.00174) | (0.00153) | (0.00326) | (0.000945) | (0.0276) | (0.0390) |
| Observations | 992 | 5,205 | 19,006 | 11,801 | 1,536 | 8,706 | 31,216 | 21,732 |
| R-squared | 0.256 | 0.078 | 0.044 | 0.067 | 0.277 | 0.085 | 0.037 | 0.064 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Table E1 The effects of density indicator: by manufacturing/service or country group - robustness check for the different time period

| Variable | Manufacturing/service | | Counrty group | | |
|---|---|---|---|---|---|
| | Manu | Service | West and North | East | South |
| Density | 0.00383*** | 0.00254 | 0.00183 | 0.00569** | 0.00404** |
| | (0.00106) | (0.00251) | (0.00114) | (0.00240) | (0.00187) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.00938 | -0.00136 | 0.00754 | 0.0205** | 0.00788* |
| | (0.0103) | (0.00189) | (0.00825) | (0.0101) | (0.00475) |
| Observations | 24,108 | 4,488 | 13,977 | 5,612 | 9,007 |
| R-squared | 0.025 | 0.055 | 0.032 | 0.061 | 0.047 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Table E2 The effects of density indicator: by both industry and region categories - robustness check for the different time period

| Variable | HM-KIS | | | | LHM-LKIS | | | |
|---|---|---|---|---|---|---|---|---|
| | K_hubs | | Indus_zones | Non-S&T_regions | K_hubs | | Indus_zones | Non-S&T_regions |
| | K_city/capital | K_tec hubs | | | K_city/capital | K_tec hubs | | |
| Density | 0.00693 | -0.00289 | 0.00520*** | 0.00600* | 0.00974 | -0.00273 | 0.00120 | 0.00254 |
| | (0.0175) | (0.00411) | (0.00195) | (0.00355) | (0.0145) | (0.00272) | (0.00133) | (0.00258) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.0155 | 0.0147 | 0.0305 | -0.000723 | -0.00535 | -0.00123 | 0.0175 | 0.000615 |
| | (0.0224) | (0.0127) | (0.0209) | (0.00222) | (0.00927) | (0.00182) | (0.0111) | (0.00184) |
| Observations | 252 | 1,324 | 4,818 | 2,984 | 388 | 2,190 | 7,881 | 5,500 |
| R-squared | 0.239 | 0.096 | 0.042 | 0.062 | 0.252 | 0.078 | 0.038 | 0.056 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Table F1 The effects of density indicator: by manufacturing/service or country group - robustness check for adding time-varying control variables

| Variable | Manufacturing/service | | Counrty group | | |
|---|---|---|---|---|---|
| | Manu | Service | West and North | East | South |
| Density | 0.00415*** | 0.00419*** | 0.00348*** | 0.00583*** | 0.00327*** |
| | (0.000534) | (0.00130) | (0.000644) | (0.00116) | (0.000917) |
| Average GDP per capita | 0.000517 | -0.000859 | 0.00107 | -0.00249 | -0.00219 |
| | (0.00175) | (0.00360) | (0.00459) | (0.00413) | (0.00326) |
| LOS-Index | 0.000658 | 0.0117 | 0.00288 | 0.0133 | -0.00344 |
| | (0.00238) | (0.00745) | (0.00279) | (0.0157) | (0.00444) |
| Population density (log) | -0.00436 | 0.00577 | -0.0929 | 0.182 | 0.0428 |
| | (0.0300) | (0.0780) | (0.0800) | (0.211) | (0.0550) |
| GDP (log) | 0.0120 | -0.0651 | -0.0376 | -0.0384 | -0.0455 |
| | (0.0221) | (0.0579) | (0.0541) | (0.0497) | (0.0542) |
| Share_S&T (log) | -5.79e-05 | 0.00486 | 5.81e-05 | 0.00784 | -0.00380 |
| | (0.00197) | (0.00476) | (0.00261) | (0.00486) | (0.00332) |
| Gross capital/employee (log) | -0.00484 | 0.00300 | -0.00791 | -0.00169 | -0.00777 |
| | (0.00382) | (0.00807) | (0.00551) | (0.00649) | (0.00738) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.0210 | 0.00267 | 0.387 | 0.129 | 0.0482 |
| | (0.114) | (0.298) | (0.312) | (0.179) | (0.0451) |
| Observations | 92,855 | 17,123 | 52,810 | 22,451 | 34,717 |
| R-squared | 0.023 | 0.043 | 0.030 | 0.057 | 0.044 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Table F2 The effects of density indicator: by both industry and region categories - robustness check for adding time-varying control variables

| Variable | HM-KIS | | | | LHM-LKIS | | | |
|---|---|---|---|---|---|---|---|---|
| | K_hubs | | Indus_zones | Non-S&T_regions | K_hubs | | Indus_zones | Non-S&T_regions |
| | K_city/capital | K_tec hubs | | | K_city/capital | K_tec hubs | | |
| Density | 0.0147 | 0.00270 | 0.00454*** | 0.00556*** | 0.00871 | -0.00145 | 0.000966 | 0.00273** |
| | (0.00920) | (0.00229) | (0.000933) | (0.00162) | (0.00555) | (0.00127) | (0.000705) | (0.00130) |
| Average GDP per capita | -0.161 | -0.0105 | 0.000781 | -0.000707 | 0.0944 | 0.00228 | -0.00162 | 0.00129 |
| | (0.141) | (0.0236) | (0.00648) | (0.00270) | (0.0940) | (0.0153) | (0.00342) | (0.00399) |
| LOS-Index | -0.217 | -0.000606 | 0.00858 | -0.0115 | 0.0813 | -0.00645* | 0.00972** | 0.000727 |
| | (0.192) | (0.00815) | (0.00749) | (0.00837) | (0.108) | (0.00386) | (0.00477) | (0.00784) |
| Population density (log) | -0.632 | -0.679* | -0.00330 | 0.0823 | 1.119 | -0.0184 | -0.0690 | -0.0367 |
| | (1.707) | (0.382) | (0.118) | (0.0636) | (1.366) | (0.179) | (0.0619) | (0.0535) |
| GDP (log) | -4.557 | -0.0291 | -0.0226 | -0.0532 | 2.062 | -0.100 | -0.0310 | -0.0106 |
| | (3.637) | (0.318) | (0.0904) | (0.0349) | (2.446) | (0.195) | (0.0717) | (0.0410) |
| Share_S&T (log) | 0.103 | -0.00667 | -0.00414 | 0.00838* | -0.0487 | 0.00785 | -0.00422 | -0.00215 |
| | (0.0881) | (0.00860) | (0.00434) | (0.00504) | (0.0584) | (0.00579) | (0.00348) | (0.00510) |
| Gross capital/employee (log) | -0.173 | -0.0206 | -0.0135 | 0.00249 | 0.0921 | -0.0114 | -0.0103 | -0.00531 |
| | (0.185) | (0.0265) | (0.0115) | (0.00809) | (0.116) | (0.0111) | (0.00887) | (0.00889) |
| Region-year dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry-year dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | 4.836 | 0.682 | 0.0496 | 0.0605 | -5.351 | 0.156 | 0.137 | -0.0175 |
| | (7.405) | (0.543) | (0.164) | (0.0436) | (5.969) | (0.315) | (0.0919) | (0.0399) |
| Observations | 1,010 | 5,122 | 17,794 | 11,912 | 1,546 | 8,519 | 29,182 | 22,016 |
| R-squared | 0.279 | 0.074 | 0.038 | 0.053 | 0.292 | 0.082 | 0.037 | 0.055 |

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Excluding other sectors.

Appendix G

The shift-share approach of calculation of sector-adjusted entry numbers

Due to data availability, we use empfeldloyment for each industry to measure industrial structure across regions.[13] First, we calculate the expected number of employment for each industry/region if all regions follow an identical industrial structure, see Equation (A1).

$$HE_{ij} = E_j * S_i \; with \; E_j = \sum_i E_{ij} \;\; and \; S_i = \frac{E_i}{E} = \frac{\sum_j E_{ij}}{\sum_{ij} E_{ij}}$$

(A1)

where subscripts *i* and *j* refer to industry *i* and region *j* respectively; $E_{ij}$ represents the number of employment for each industry *i* and region *j*; $E_j$ is the total employment number for region *j* and $S_i$ *is* the share of the total employment in industry *i* in the total employment of all industries and regions in the analysis.

Second, we calculate the expected entry number for each region, as shown in Equation (A2).

$$HNEN_j = \sum_i HE_{ij} * ENR_i \;\; with \; ENR_i = \frac{EN_i}{E_i} = \frac{\sum_j EN_{ij}}{\sum_j E_{ij}}$$

(A2)

The expected entry number for each region is calculated by summing up the product of the expected number of employment for each industry/region and the average entry rate of respective industry in all regions ($ENR_i$).

Third, we calculate the entry number caused by the differences between the industrial structure of each region and the average industrial structure of all regions, see Equation (A3).

$$HIEN_j = \sum_i \left(E_{ij} - HE_{ij}\right) * ENR_i \qquad \text{(A3)}$$

Fourth, the sector-adjusted entry number is obtained by subtracting $HIEN_j$ from the observed entry number for each region, as shown in Equation (A4).

$$AEN_j = \sum_i REN_{ij} - HIEN_j \qquad \text{(A4)}$$

where $REN_{ij}$ refers to the real entry number for each industry/region and $AEN_j$ is sector-adjusted entry number for each region. The sector-adjusted entry number is assumed to filter out the differences caused by differences of industrial structures across regions.

[13] Audretsch and Fritsch (2002) use the number of establishments for each industry to measure industrial structure.