Papers in Evolutionary Economic Geography

12.01

The emergence of new industries at the regional level in Spain
A proximity approach based on product-relatedness

Ron Boschma and Asier Minondo and Mikel Navarro

Utrecht University
Urban & Regional research centre Utrecht

http://econ.geo.uu.nl/peeg/peeg.html

The emergence of new industries at the regional level in Spain
A proximity approach based on product-relatedness

Ron Boschma[1], Asier Minondo[2] and Mikel Navarro[2,3]

1 Urban and Regional research centre Utrecht (URU), Utrecht University, Utrecht
2 Deusto Business School, University of Deusto, San Sebastian
3 Orkestra - Basque Institute of Competitiveness, San Sebastian

Abstract
How do regions diversify over time? Inspired by recent studies, we argue that regions diversify into industries that make use of capabilities in which regions are specialized. As the spread of capabilities occurs through mechanisms that have a strong regional bias, we expect that capabilities available at the regional level play a larger role than capabilities available at the country level for the development of new industries. To test this, we analyze the emergence of new industries in 50 Spanish regions at the NUTS 3 level in the period 1988-2008. We calculate the capability-distance between new export products and existing export products in Spanish regions, and provide econometric evidence that regions tend to diversify into new industries that use similar capabilities as existing industries in these regions. We show that proximity to the regional industrial structure plays a much larger role in the emergence of new industries in regions than proximity to the national industrial structure. This suggests that capabilities at the regional level enable the development of new industries.

Key words: Regional branching, diversification, new industries, capabilities, Spain, proximity index

JEL-codes: R11, N94, O14

Acknowledgements: Asier Minondo acknowledges financial support from the Spanish Ministry of Science and Innovation (ECO2010-21643/ECON)

1. Introduction

Regional diversification is high on the scientific as well as the political agenda. As many regions are currently facing severe economic decline, there is increasing awareness that there is a need to develop new economic activities, in order to compensate for decline and losses in other economic activities. The question what determines regional diversification and how regions develop new growth paths is a fundamental one (Feldman et al, 2005; Simmie and Carpenter 2007; Fornahl et al., 2010), but has drawn surprisingly little attention from scholars

so far. Case-study evidence shows that new industries and technologies do not start from scratch but evolve out of regional structures that provide related competences and assets (see e.g. Klepper, 2007; Boschma and Wenting 2007; Buenstorf et al. 2010; Bathelt et al. 2011; Buerger and Cantner 2011; Tanner 2011). However, this process of regional diversification based on relatedness has not been investigated systematically until very recently.

Recent studies (Hidalgo et al. 2007; Hausmann and Klinger 2007; Hausmann and Hidalgo 2010) have demonstrated that the current industrial structure at the country level affects the future state of the industrial structure of countries to a considerable degree. These scholars argue that a country needs a specific set of local capabilities to manufacture a good. As Hidalgo (2009) explains, capabilities could be tangible inputs, such as bridges, ports and highways, or intangibles, such as norms, institutions, skills or the existence of particular networks. If a country already has most of the capabilities that are needed to produce a new good, it will have few barriers to become competitive in that good. In contrast, if the country does not possess the capabilities required to manufacture the good it will be very difficult to develop that industry. Hence, the existing set of capabilities determines which new industries will be feasible to develop in the future.

The aim of this paper is to provide evidence for this process of branching at the regional (i.e. the sub-national) rather than the national scale, because we believe that the mechanisms through which capabilities are transferred between new and existing industries operate mainly (though not exclusively) at the regional scale. To our knowledge, only one study (Neffke et al., 2011) has provided systematic evidence for branching in related activities at the regional scale. Our paper makes two contributions to this literature. First, we aim to determine at which spatial scale (the national versus the regional scale) this process of related diversification is more manifest. Second, we investigate the process of related diversification at the regional level by means of the proximity index developed by Hausmann, Hidalgo and Klinger, which determines the degree of relatedness between products. We use that information to determine the extent to which Spanish regions (at the NUTS 3 level) have diversified in related products in the period 1988-2008. Our study confirms that territories diversify into industries that are related to the existing set of industries. As expected, we found that this process of related diversification is stronger and more manifest at the regional than the national scale. As such, we claim that diversification is subject to a path-dependent branching process that occurs mainly at the regional, rather than the national scale.

The structure of the paper is as follows. Section 2 briefly introduces the literature on regional diversification and relatedness. Section 3 introduces the data, the methodology and the variables included. Section 4 present some descriptives. Section 5 present the main outcomes of the analyses conducted. Section 6 presents the main conclusions.

2. Regional diversification based on relatedness

In the 1980s and 1990s, there was a lot attention for the degree of relatedness between technologies used in sectors, because this was believed to affect the scope of knowledge spillovers and inter-industry learning (Rosenberg and Frischtak 1983; Carlsson and Stankiewicz 1991). Cluster of industries were identified on the basis of technological complementarities, and various ways to measure technological relatedness between industries were developed and applied (see e.g. Farjoun 1994; Teece et al. 1994). Nooteboom (2000) referred to the notion of optimal cognitive proximity between economic agents. With that

notion, he meant that the cognitive distance should neither be too large (to ensure effective communication), nor too small (to avoid lock-in), because both would harm the interactive learning process between agents.

In the 2000s, this idea of industry relatedness was combined with the empirical observation made by economic geographers that knowledge spillovers were often geographically bounded. One became increasingly aware of the fact that the extent to which the variety of industries or technologies present in a region is related might positively affect the scope for knowledge spillovers and learning, as local firms in different but related activities can profit more from mutual spillovers than local firms in unrelated industries. Porter (2003) was one of the first to recognize the importance of spatial externalities across related industries and incorporated this idea into his cluster concept. Porter made the claim that specialization in clusters of related industries, not in industries *per se*, is beneficial for regional development.

Frenken et al. (2007) positioned this concept of industry relatedness in the spatial externalities and regional growth literature. They referred to the related variety effect which concerned externalities that come from a diversity of related industries in a region. Close to the idea of optimal cognitive distance of Nooteboom, the notion of regional related variety captures a delicate balance between cognitive proximity and cognitive distance across sectors in a region that enables knowledge to spill over effectively between sectors. Thus, the more variety across related sectors in a region, the more learning opportunities there are for local industries, the more inter-sectoral knowledge spillovers are likely to take place, and the higher the economic performance of regions. Overall, studies have found empirical support for the significance of related variety for regional growth in the Netherlands (Frenken et al., 2007), Italy (Boschma and Iammarino 2009; Quatraro 2010), Great Britain (Bishop and Gripaios, 2010), Spain (Boschma et al., 2011) and Germany (Brachert et al., 2011).

Apart from the fact that relatedness among industries in a region may drive regional growth, it may also drive the entry of new industries in a region. Recent studies (Hidalgo et al. 2007; Hausmann and Klinger 2007; Hausmann and Hidalgo 2010) have argued that the current industrial structure of a country affects its future state, because the existing set of capabilities in a country determines which new industries will be feasible to develop in the near future. According to Hausmann and Hidalgo (2010), capabilities refer to those productive inputs that are not internationally tradable. If they were, firms could acquire those capabilities, and their absence at the national level would not affect the possibilities of countries to develop new industries. Capabilities could be tangible inputs, such as bridges, ports and highways, or intangibles, such as norms, institutions, skills or networks (Hidalgo 2009). When a country already has most of the capabilities that are needed to produce a new good, it will have few barriers to become competitive in that good. In contrast, if the country does not possess the capabilities required to manufacture the good, it will be very difficult to develop that industry.

Focusing on shifts in export portfolios of countries over time, Hausmann and Klinger (2007) demonstrated that countries predominantly expand their export activities by moving into products that are related to their current export portfolio. In addition, rich countries that have a wide range of related export products have more opportunities to diversify into new related export products, and thus have more opportunities to sustain economic growth, in comparison to poorer countries. Hidalgo et al. (2007) pointed out that the location of a country in the product map is crucial here. That is, some products are positioned in the more dense parts of the product-map (where they are related to many other products), whereas other products have few links with other products, and are located in the periphery of the product map. The location of a country in the product map then determines its diversification possibilities. If a

country is specialized in products located more in the periphery of the product map, it will have fewer possibilities to deploy its capabilities in new goods. In contrast, if a country is specialized in a dense part of the product map, it implies that its capabilities can be deployed in a larger number of products and, hence, diversification possibilities will be larger.

Not denying the importance of the nation state (see e.g. Bathelt 2003), there are very good reasons to believe that the regional (i.e. the intra-national) scale is even more important for the process of related diversification. We believe that many capabilities do not move easily within countries as well, and therefore, regions need to possess certain capabilities at the local level to develop new industries. If capabilities were fully mobile, regions could acquire those in other regions and hence, the development of new industries would not be heavily constrained by the lack of capabilities at the regional level. And because capabilities do not transfer easily within a country, we believe that regions located in the periphery of the product map will have more difficulties to diversify than regions located in the more central parts of the product map.

There is a huge literature in economic geography that stresses the importance of region-specific assets for regional competitiveness (Storper 1992; Markusen 1996; Lawson 1999; Maskell and Malmberg 1999; Maskell 2001). Storper (1995) talked about 'untraded interdependencies' such as practices and conventions, which determine the competitiveness of regions in a globalized world to an increasing extent. Malmberg and Maskell (1999) referred to 'localized capabilities' which are associated with a particular knowledge and competence base and a surrounding institutional environment that accumulate at the regional level. These 'localized capabilities' are intangibles with a high degree of tacitness which form a crucial asset for regions because they cannot be easily imitated by other regions. For instance, this literature claims that institutional settings at the regional level are the outcome of a long history that are neither for sale on the market nor can they be designed easily through public intervention (Gertler 2003). Therefore, regions tend to evolve along particular trajectories, and inter-regional variety is a persistent feature of every capitalist economy.

We claim that these region-specific, localized capabilities also operate as sources of diversification. Boschma and Frenken (2011) describe this process of regional diversification by which new industries arises from technologically related industries in regions in which existing competences are recombined in new economic activities as 'regional branching'. They claim that the regional scale is crucial here, because related diversification into new industries tends to occur through knowledge transfer mechanisms like entrepreneurial spinoffs, firm diversification, labour mobility and social networking, all of which have a strong local bias. Successful new firms are often founded by entrepreneurs that come from related activities in the same region. There is substantial evidence that these types of spinoffs contribute indeed to the process of old sectors giving birth to new sectors within a region. Longitudinal case-studies of industries have demonstrated that the most successful entrepreneurs in new industries are those that exploit competences they acquired in technologically related industries (e.g. Klepper and Simon 2000; Boschma and Wenting 2007; Klepper 2007). Labour mobility, as another key mechanism through which knowledge is transferred across firms and sectors, is also expected to be a major source of regional branching. This is because most labour mobility occurs mainly (but not exclusively) at the regional level within labour market areas (Eriksson 2011). However, empirical evidence on its importance for regional branching is (yet) lacking.

Case studies show that new local industries are deeply rooted in related activities in the region (Glaeser, 2005; Boschma and Wenting 2007; Klepper, 2007; Bathelt et al. 2011). To our knowledge, the study of Neffke et al. (2011) is yet the only one that has provided systematic evidence that regions are more likely to expand and diversify into industries that are closely related to their existing activities. They looked at the probability of new industries entering a region, and how that is affected by the degree of technological relatedness with other industries in the region. They followed the evolution of the industrial structure in 70 Swedish regions during the period 1969-2002. Neffke et al found that a new industry is more likely to enter a region when it is technologically related to other industries in that region. Another interesting finding was that an existing industry had a higher probability to exit a region when that industry was not, or weakly technologically related to other industries in the region. Consequently, they found strong and persistent evidence that the rise and fall of industries is subject to a path-dependent process at the regional level.

This paper investigates the extent to which new industries that emerged in Spain in the period 1988-2007 is related to existing industries, and if so, whether the regional scale (at the NUTS 3 level) is more important for this process of related diversification than the national scale.

3. Data and methodology

The first step to take is to develop a relatedness indicator, in order to investigate the degree of relatedness between new and existing industries. Various measures of technological relatedness between industries have been developed and applied (see e.g. Farjoun 1994; Teece et al. 1994; Porter 2003; Frenken et al., 2007; Neffke 2009). We use the proximity index of Hidalgo et al (2007) to determine the extent to which two products share a similar set of capabilities. Boschma, Minondo and Navarro (2011) have demonstrated that the relatedness indicator based on the proximity product index by Hidalgo et al. delivered better results concerning the relationship between related variety and regional growth than alternative measures, like the conventional *ex ante* measure of related variety (Frenken et al., 2007) and the cluster based *ex-post* measure of relatedness developed by Porter (2003). This approach comes close to the revealed industry relatedness measure developed by Neffke and Svensson Henning (2008) who made use of co-occurrence analysis of products at the plant level, which is based on the frequency of combinations of two products in the product portfolio of plants.

The proximity index of Hidalgo et al (2007) makes use of how often countries have comparative advantage in two goods simultaneously. Following Balassa, a country has comparative advantage when the share of a product in its exports is larger than the share of that product in world exports. If countries with a comparative advantage in good A also have comparative advantage in product B, this implies that the products A and B demand the same capabilities and, hence, are close to each other. To calculate proximity between each pair of products i and j, first, we determine whether countries have revealed comparative advantage in product i. To do so, we divide the share of product i in country's exports by the share of product i in world exports. If this ratio is 1 or above 1 we consider that the country has revealed comparative advantage. Second, we calculate the probability of having comparative advantage in product i, by dividing the number of countries that have comparative advantage in product i by the number of countries in the sample. Third, we calculate the joint probability of having comparative advantage in product i and product j, dividing the number of countries that have comparative advantage in product i and in product j by the number of countries in the sample. Fourth, we calculate the probability of having comparative advantage in product i

conditional on having comparative advantage in product *j*. Conditional probability is calculated by dividing the joint probability of having comparative advantage in product *i* and product *j* by the probability of having comparative advantage in product *j*. Following the same steps, we also calculate the probability of having comparative advantage in product *j* conditional on having comparative advantage in product *i*. Hence, for each pair of products (*i*, *j*), we end-up with two conditional probabilities: the probability of having comparative advantage in product *i* conditional on having comparative advantage in product *j*, and the probability of having comparative advantage in product *j*, conditional on having comparative in product *i*. Proximity is equal to the lowest value of the two conditional probabilities. Algebraically, proximity (φ) between product *i* and product *j* at year *t* is defined as:

$$\varphi_{ijt} = \min \left\{ P(x_{i,t} \mid x_{j,t}), P(x_{j,t} \mid x_{i,t}) \right\} \qquad (1)$$

where $P(x_{i,t} \mid x_{j,t})$ is the conditional probability of having revealed comparative advantage in product *i* given that the country has revealed comparative advantage in product *j* and $P(x_{j,t} \mid x_{i,t})$ is the conditional probability of having revealed comparative advantage in product *j* given that the country has revealed comparative advantage in product *i*.

According to Hausmann and Klinger (2007), countries will tend to diversify into goods that are close to the country's current productive structure. To test this hypothesis, they develop a density indicator to measure how close a potential new product is from the country's current productive structure. These authors argue that if the country has comparative advantage in most of the goods that are close to the product that has not been developed yet, density will be high and the probability of developing comparative advantage in that industry in the future will be high. In contrast, if density around the new industry is low, the country will be unlikely to develop comparative advantage in that product in the future. Algebraically, density is defined as:

$$d_{i,c,t} = \left(\frac{\sum_k \varphi_{i,k,t} x_{k,c,t}}{\sum_k \varphi_{i,k,t}} \right) \qquad (2)$$

where $x_{k,c,t}$ takes the value of 1 if country *c* has revealed comparative advantage in product *k* at time *t* and zero otherwise. The density around product *i*, for which country *c* has not yet developed comparative advantage at time *t*, is the sum of proximities from product *i* to all products in which country *c* has comparative advantage at time *t*, divided by the sum of proximities from good *i* to all products. If country *c* happens to have comparative advantage in all products at proximity higher than zero to product *i*, density will equal 1. In contrast, if country *c* does not have comparative advantage in any of the products related to product *i* density will equal zero.

In contrast to Hausmann and Klinger (2007), we want to analyze in this paper whether the emergence of a new industry in a Spanish province is determined by its distance to the province productive structure (province-level density), or by its distance to the rest of Spain's productive structure (country-level density).

To perform this analysis, first, we have to calculate the proximity between products. Following Hidalgo et al. (2007), we use country-level data from the NBER World Trade Database to calculate product proximity indexes (Feenstra et al., 2005). This database, which is available from www.nber.org/data, offers data on bilateral trade flows for the period 1962-2000. The trade data is disaggregated for 775 products included in the SITC Rev.2, 4-digit

classification. The main advantage of this database is that it amends the original data provided by the United Nations in two ways. First, it gives primacy to the trade flows reported by the importers, as it is assumed that trade flows are better recorded in the importing country than the exporter country. Second, corrections and additions are made to the United Nations data for trade flows to and from the United States, exports from Hong Kong and China, and imports into many other countries. From 1960 to 1983, the database offers data for all trading partners. From 1984 onwards, the database offers data for 72 countries, which account for 98% of world exports.[1]

The second step is to calculate two density measures for each province in Spain: (1) a province-level density; and (2) a country-level density for the rest of Spain. The province-level density measures how far a new industry is from the productive structure of the respective province. The country-level density measures how far a new industry is from the productive structure of the country (minus the respective province).

To calculate the province-level density we change equation (2). Now, $x_{k,c,t}$ takes the value of 1 if province c has revealed comparative advantage in product k at time t and zero otherwise. To determine the revealed comparative advantage of provinces, we use export data for Spanish provinces for the period 1988-2008 from the Spanish Dirección General de Aduanas - Agencia Tributaria database.[2] Following Balassa again, revealed comparative advantage is determined by dividing the share of a product in the exports of a province by the share of the product in world exports.

To calculate the country-level (rest of Spain) density, we change equation (2) again. Now, $x_{k,c,t}$ takes the value of 1 if the rest of Spain has revealed comparative advantage in product k at time t and zero otherwise. Revealed comparative advantage of the rest of Spain is computed by adding the exports of all other provinces that form this group, obtained from the Dirección General de Aduanas - Agencia Tributaria database, and applying Balassa's formula. Note that the country-level index will differ across provinces. We calculate the province-level density and the country-level (rest of Spain) density for 47 Spanish provinces. Due to their special geographical location, we exclude the provinces located in the islands and the two autonomous cities located in Africa. Due to the time-coverage of exports of provinces, the period of analysis is 1988-2008.

As set out before, Neffke et al (2011) was the first to provide systematic evidence for the occurrence of regional branching. Our study differs in at least two respects. The first is that our study tests whether regional branching is more important at the regional scale than the national scale. For that purpose, we test the effect of relatedness at both geographical levels simultaneously. In addition to this regional/national comparison, the co-occurrence measure used in our paper is different to that used by Neffke et al. (2011). First, we use a large pool of countries to build the proximity measures, while the co-occurrence data built by Neffke et al are based on data on Sweden. Second, the number of products for which co-occurrence is calculated is larger in our study (i.e. 774) than in the Swedish study (i.e. 174). Finally, Neffke et al. do not control for the fact whether plants are competitive in the products they co-

[1] To calculate proximity indexes, we use a sample of countries that provide data in all the years that are included in the analysis. We exclude countries with a population of less than 3 million. To avoid endogeneity problems, we excluded Spain from the sample.

[2] These data are in the Combined Nomenclature 8-digit classification. These needed to be transformed to SITC Rev 2. 4-digits, to match the classification used to calculate the proximity indexes.

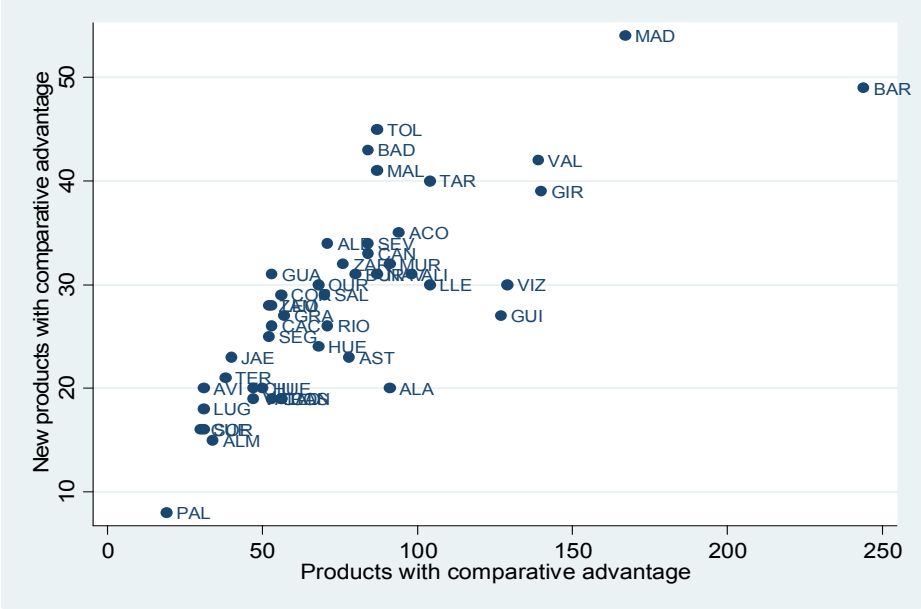
manufacture. In our study, we established a revealed comparative advantage threshold, in order to ensure that the co-occurrence event is an economically relevant and significant one.

4. The relationship between density and entry of new products in Spanish regions

In this section, we explore the relationship between density and the emergence of new products in Spanish provinces. First, we analyze the relationship between the number of products in which Spanish provinces have comparative advantage and the number of new products in which Spanish provinces have developed comparative advantage. As explained before, we consider that a province has comparative advantage in a product if the revealed comparative advantage index is one or above one. Following Hausmann and Klinger (2007), we divide the period of analysis, 1988-2008, in 5-year window intervals. This interval seems long enough to allow new industries to emerge, and short enough to provide a large set of observations for the non-parametric and parametric analyses. We also repeated the analyses with shorter (4-year) and longer (10-year) intervals to test for robustness, with no change in results. The number of products with comparative advantage is an average of the years 1988, 1993, 1998, 2003 and 2008. To calculate the number of new products, we took the average of the number of products in which provinces did not have comparative advantage in the years 1988, 1993, 1998 and 2003 but developed comparative advantage five years later.

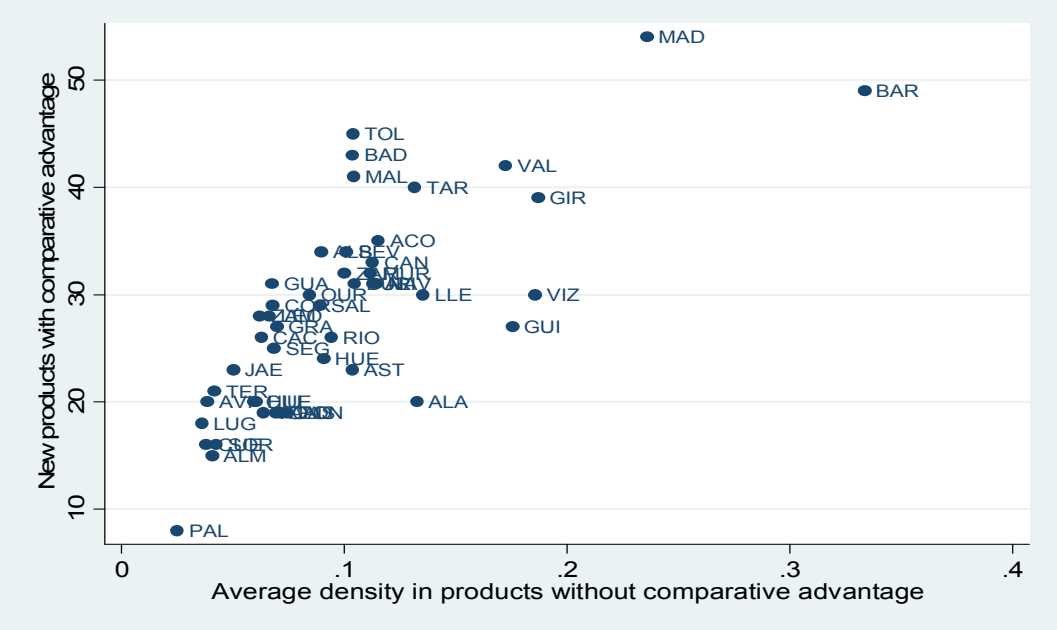
As can be seen in Figure 1, there is a clear positive relationship between the number of products in which a province has comparative advantage and the number of new products in which a province develops comparative advantage five years later. For example, we can observe that Barcelona and Madrid, the Spanish provinces with the largest number of products with comparative advantage (245 and 168 respectively) are the provinces that develop comparative advantage in the largest number of products (49 and 55 respectively). In contrast, the Spanish provinces with the lowest number of products with comparative advantage, i.e. Palencia and Cuenca (19 and 30 respectively) also developed comparative advantage in the lowest number of products (8 and 16 respectively).

Figure 1. Relationship between products with comparative advantage at time *t* and new products with comparative advantage at time *t+5* in Spanish provinces (1988-2008 average; 5-year intervals)



The question is whether this relationship between these two variables is explained by the average density around the products without comparative advantage. As shown in Figure 2, there is a strong positive relationship between the average density in the products without comparative advantage and the number of new products in which a province develops comparative advantage five years later.[3]

Figure 2. The relationship between the average density in products without comparative advantage at time *t* and new products with comparative advantage at time *t+5* in Spanish provinces (1988-2008 average; 5-year intervals)



[3] As the NBER database ends in 2000, we use proximity indexes in year 2000 to calculate densities in year 2003.

Then, we analyzed whether the probability of transitioning into a new product increases with density. We performed separate analyses for density calculated at the country level and density calculated at the province level. As explained before, we expect province-level density to have a larger impact than country-level density in determining the probability of Spanish provinces to transition into new products. Figure 3a and 3b present the probability of transitioning into a new product for different density ranges. As before, we use data for the period 1988-2008, divided in five year intervals. Note that in both graphs the maximum density range is 0.5, as there is no product with a larger density neither in the country-level indicator nor in the province-level indicator. The figures show that the probability to develop comparative advantage in a new product increases gently with density.[4] For example, if we look at province-level density in Figure 3b, a Spanish province has almost six times more probability to transition into a new product if density is between 0.4 and 0.5, than if density is between 0.0 and 0.1. We also observe that this probability is always larger for density calculated at the province level than for density calculated at the country level. This result suggests that the presence of capabilities at the province level has a larger impact on the probability of developing comparative advantage into a new product than the presence of capabilities at the country level.

[4] The unconditional probability of developing comparative advantage in a new product is 2.7 per cent. The unconditional probability of keeping comparative advantage in a product is 69.1 per cent.

Fig. 3a Probability of transitioning into new products for Spanish provinces. Country-level density (Period 1988-2008; 5-year intervals)

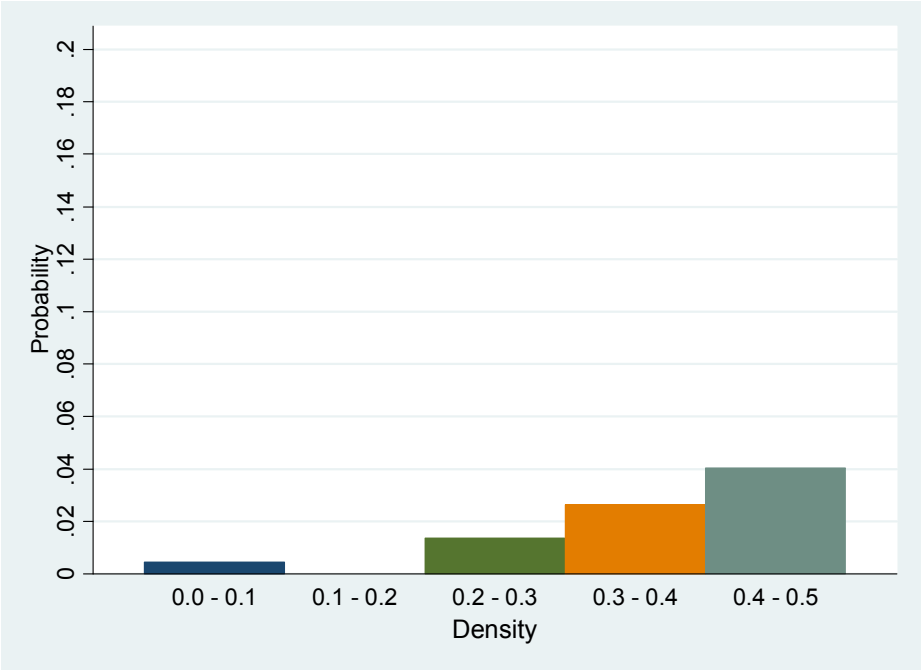
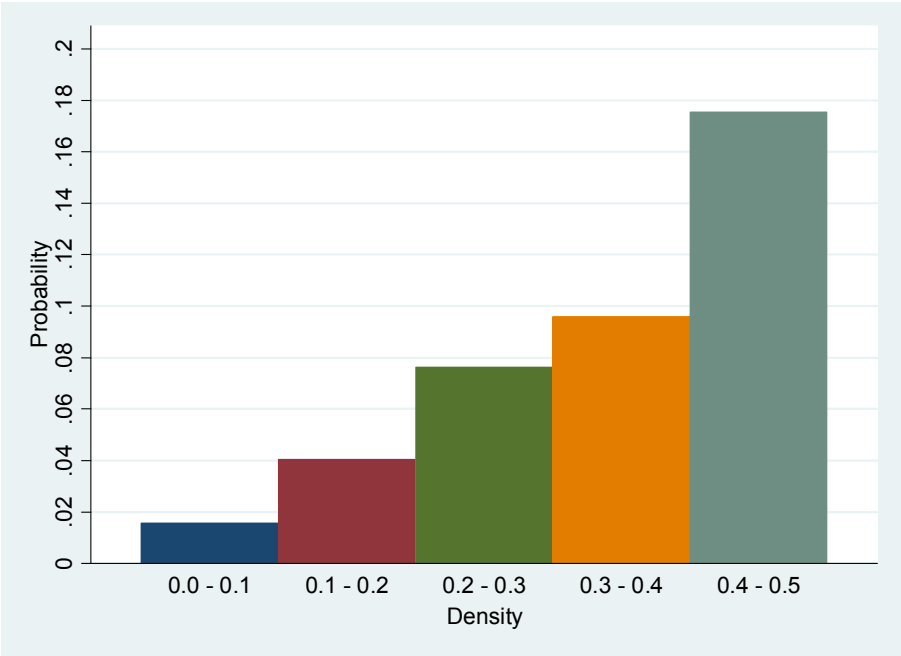


Fig. 3b Probability of transitioning into new products for Spanish provinces. Province-level density (Period 1988-2008; 5-year intervals)



There is an alternative way to study whether higher density facilitates the emergence of new industries. It consists of comparing the distribution along the density of products that remain without comparative advantage with the distribution along the density of products that develop comparative advantage. As the probability of transitioning into new products rises

with density, we expect the bulk of products that remained without comparative advantage to be concentrated at lower density levels, and the majority of products that develop comparative advantage to be concentrated at higher density levels. That is, we expect the distribution of products that develop comparative advantage to be positioned more to the right of the distribution of products that do not develop comparative advantage. Moreover, as the presence of capabilities at the province level are more important to develop comparative advantage than the presence of capabilities in the rest of the country, we expect the separation between the distribution curves to be more pronounced for density calculated at the province level than for density calculated at the country level. To test these hypotheses, in Figure 4a and Figure 4b we compare the probability density function of products that remained without comparative advantage (dashed lined - no CA at t+5) with the probability density function of products that developed comparative advantage (solid line - new industry at t+5).[5] The horizontal axis measures the density around products that had not developed comparative advantage at time *t*. In Figure 4a density refers to country-level density and in Figure 4b to province-level density. The vertical axis measures the probability that a product has a certain density level.

As expected, we find that for low density levels, the share of products that remained without comparative advantage is larger than the share of products that developed comparative advantage. When higher product density levels are reached, the reverse is the case.[6] We also observe that the difference between the probability curves to be more pronounced when density is calculated at the province level (Figure 4b) than when density is calculated at the country level (Figure 4a). We can see that the country-level density function is skewed to the right, whereas the province-level density function is skewed to the left. This can be explained by the average density of products at the country and the province level. As a country is a larger geographical unit than a province, there will be a larger set of products with revealed comparative advantage at the country than the province level. As density in new products is determined by the number of neighbour products that have revealed comparative advantage, density will be higher at the country-level than at the province-level. It is also interesting to note that transitioning into new products happens at lower density levels in Figure 4b than in Figure 4a. This result might point out that the capabilities needed to develop new industries are transferred more easily within provinces, than from the rest of the country to the province. It is only when the level of capabilities in the rest of the country is very high that the transfer of capabilities from the rest of the country to the province is more likely.

Figure 4a. Probability density function for new products vs probability density function for products with no comparative advantage. Density calculated at the country level (period 1988-2008; 5-year intervals)

[5] To draw the figures, we use the epanechnikov kernel estimator in STATA, calculated on n=50 points, and allowing the estimator to calculate the optimal bandwidth.

[6] Both in Figure 4a and in Figure 4b, the t-test for equality of average density between products that remained without comparative advantage and products that developed comparative advantage was strongly rejected.

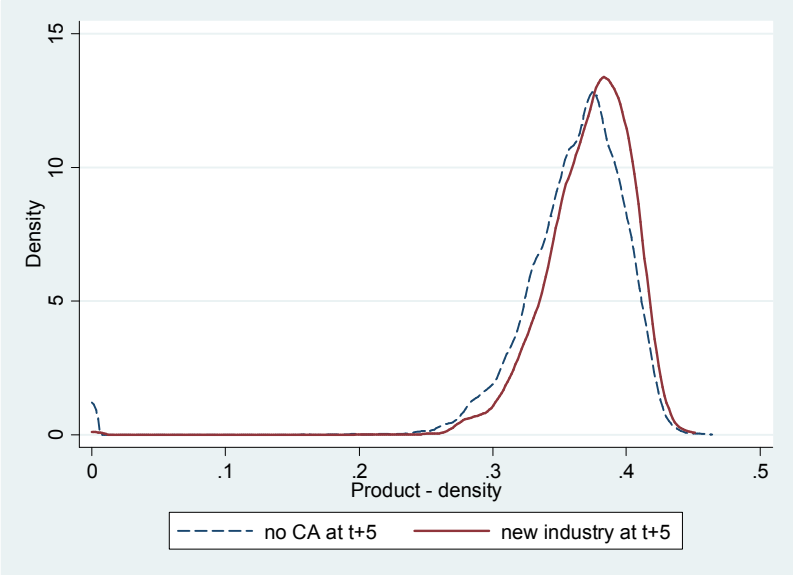
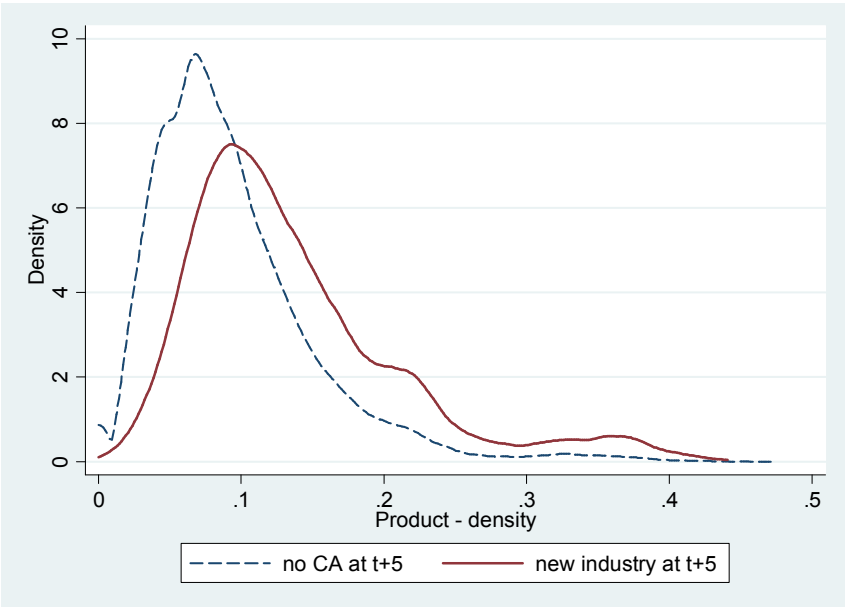


Figure 4b. Probability density function for new products vs probability density function for products with no comparative advantage. Density calculated at the province level (period 1988-2008; 5-year intervals)



5. Econometric analyses

To test formally whether province-level density plays a larger role than country-level density in developing comparative advantage in new products, we follow Hausmann and Klinger (2007) and estimate the following econometric equation:

$$x_{i,p,t+5} = \alpha + \gamma x_{i,p,t} + \alpha_c d^c_{i,p,t} + \alpha_p d^p_{i,p,t} + \beta_{i,t} + \beta_{p,t} + e_{i,p,t} \quad (3)$$

where $x_{i,p,t+5}$ takes the value of 1 if province p has revealed comparative advantage in product i at year $t+5$ and zero otherwise; $d^c_{i,p,t}$ denotes country-level density around product i in province p at year t; $d^p_{i,p,t}$ denotes province-level density around product i in province p at year t; and $e_{i,p,t}$ is the error term. Note that the coefficient γ captures the contribution of having comparative advantage at time t to keeping comparative advantage at time $t+5$, once we have removed the influence of other factors that may also affect the persistence of comparative advantage, such as density, product and province characteristics. α_p and α_c capture the impact of province-level and country-level density respectively on developing comparative advantage in new products. As capabilities should be available to develop new industries, we expect both coefficients to be positive and statistically significant. And as we claim that capabilities are more easily deployed in new industries if they are locally available, we also expect α_p to be larger than α_c. Finally, the equation includes fixed effects, $\beta_{i,t}$ and $\beta_{p,t}$, to control for time-varying product characteristics and time-varying province characteristics respectively. As in the descriptive analyses, the period of analysis (1988-2008) is divided in five year intervals.[7]

When the dependent variable is binary, scholars usually estimate the econometric equation using Logit or Probit models. However, when the econometric equation includes a large number of dummy variables, as in our case, Probit and Logit models might lead to coefficients that are biased and inconsistent (Greene, 2008). To overcome this problem, known as the incidental parameter problem, we estimate the equation with a linear probability-OLS model. However, this model is not free of limitations. As Bernard and Jensen (2004) point out, when the lagged dependent variable is included in the right hand-side of the econometric equation (revealed comparative advantage), the fixed-effects model may lead to inconsistent estimates, especially for the lagged variable coefficient. To overcome this problem, we use a System-GMM technique (Blundell and Bond, 1998) which is designed to estimate dynamic panel-data models. The advantage of this model, in addition to ensuring the consistency of estimates, is that it addresses the endogeneity problems that might be present in the model. The System-GMM model sets a system of equations in differences and in levels and uses lagged levels and the first lag of the first-differences of the endogenous variables as instruments in these equations respectively, in addition to allowing for fixed effects. We estimate the equation (3) both with the linear probability model and the System-GMM model.

Table 1 presents the results of the econometric analyses. Columns 1-3 present the results of the linear probability model, and columns 4-6 the results of the System-GMM model. In this latter model, all covariates are treated as potentially endogenous, and all available lags are used as instruments. We estimate equation (3) with one density variable (province or country), and with both density variables. To facilitate comparability, the density variable is normalized by subtracting the mean and dividing by the standard deviation. A first outcome in all estimations is that having comparative advantage at the beginning of the period raises considerably the probability of having comparative advantage at the end of the period. However, the size of the coefficient is larger in the linear probability model than in the System-GMM model. Another key finding is that density at the province level has a larger

[7] To avoid the influence of outliers, we removed from the sample observations below the 1st percentile and above the 99th percentile in both density calculated at the province level and density calculated at the country level. To control for possible errors in the recording of export data, we also excluded those observations in which province p exports of 8-digit product i were zero in year t, but revealed comparative advantage for the same product and province was 1 or above 1 in year $t+5$.

Table 1. Regression results on the determinants of having comparative advantage in the future

	(1)	(2)	(3)	(4)	(5)	(6)
Model	Linear probability	Linear probability	Linear probability	System-GMM	System-GMM	System-GMM
Comparative advantage	0.626 (0.013)***	0.618 (0.013)***	0.618 (0.013)***	0.387 (0.011)***	0.378 (0.011)***	0.380 (0.011)***
Density at country level	0.005 (0.003)		0.011 (0.003)***	0.020 (0.002)***		0.003 (0.002)
Density at province level		0.033 (0.005)***	0.034 (0.004)***		0.043 (0.003)***	0.045 (0.003)***
Adj. R-squared	0.470	0.472	0.473			
Hansen J statistic p-value				0.000	0.000	0.000
AB test for AR(1) in first differences p-value				0.000	0.000	0.000
AB test for AR(2) in first differences p-value				0.000	0.000	0.000
Obs.	132,790	132,790	132,790	132,790	132,790	132,790

Note: In linear probability model estimations province clustered standard errors in parentheses. In System-GMM estimations robust standard errors in parentheses. Linear probability regressions include province+period and product+period dummies. *** statistically significant at 1%.

impact on the evolution of comparative advantage than density at the country level. In the linear probability model, when estimated one by one, the density coefficient calculated at the province level (0.033) is almost seven times larger than the coefficient calculated at the country level (0.005); however, this latter coefficient is not statistically significant. When we introduce both densities in the equation (column 3), we find that the density coefficient calculated at the province level (0.034) is three times larger than the coefficient calculated at the country level (0.011). It is interesting to observe that the coefficient of density at the country level rises when both densities enter simultaneously in the linear probability model estimation.

In the System-GMM model, the effect of density at the province level is also larger than the effect of density at the country level. Moreover, when both densities enter simultaneously in the equation (column 6), the coefficient for country-level density is no longer statistically significant. At the bottom of the table, we present different tests for the System-GMM estimations. First, the Hansen test rejects the hypothesis that the model is saturated by a large number of instruments. Second, we test for the presence of first-order autocorrelation in errors first differences. As expected, the null hypothesis on the existence of autocorrelation is not rejected, because first differences in errors share an error level component. However, we cannot reject either the null hypothesis on the existence of second-order autocorrelation in error first differences. The non-rejection of this hypothesis means that the instruments that are used in the model are weak. Trying to overcome this limitation, we re-estimated the model introducing only third and higher order lags. However, autocorrelation remained.

Notwithstanding the limitations in the System-GMM estimations, the results of the econometric analyses clearly confirm that the productive structure at the regional level plays a larger role than the productive structure at the country level in determining the emergence of new industries at the regional level. This suggests that regions should possess certain capabilities at the regional level to ensure the development of new industries.

Following Hausmann and Klinger (2007), we performed an additional analysis to distinguish the role that density plays in developing comparative advantage in new products at the regional level from its contribution in keeping comparative advantage in current products. The estimated equation is the following:

$$x_{i,p,t+3} = \alpha + \gamma x_{i,p,t} + \alpha_c^n(1 - x_{i,p,t})d_{i,p,t}^c + \alpha_c^o(x_{i,p,t})d_{i,p,t}^c + \alpha_p^n(1 - x_{i,p,t})d_{i,p,t}^p + \alpha_p^o(x_{i,p,t})d_{i,p,t}^p + \beta_{i,t} + \beta_{p,t} + u_{p,c} \quad (4)$$

Now, α_c^n captures the impact of country-level density in developing comparative advantage in new products, whereas α_c^o captures the impact of country-level density in keeping comparative advantage in product *i*. And α_p^n captures the impact of province-level density in developing comparative advantage in new products, whereas α_p^o captures the impact of province-level density in keeping comparative advantage in product *i*.

As shown in Table 2, in all estimations the density coefficient on current products is much larger than the density coefficient on new products. This result suggests that density plays a larger role in keeping comparative advantage in current products than developing comparative advantage in new products. We still find that province-level density has a much larger impact than country-level density in determining the emergence of new industries. In the linear probability model, when the density coefficients enter simultaneously (column 3), the province-level coefficient is almost three times larger than the country-level coefficient. In the System-GMM estimation, we also obtain a large and positive coefficient for density on new

products at the province level. The country-level coefficient for density on new products is positive when only country-level variables are introduced in the model (column 5). However, it becomes negative when province and country level densities are introduced simultaneously. Finally, in all estimations the province-level density plays a larger role than country-level density in keeping comparative advantage in current products.

The results of the linear probability model showed that country-level density also influences the development of new industries in Spanish provinces. We also analyzed whether the positive impact of country-level density stems from the capabilities located in neighbouring provinces, or from the capabilities that are located in the rest of Spain. To do so, we constructed a new variable that captures the density in adjacent provinces. We also re-computed the (rest of Spain) country-level density, which captures now the capabilities that are present in non-adjacent provinces. Table 3 presents the results of the regressions. First, we estimated the linear probability and System-GMM models without distinguishing between the effects of density on new and current products (columns 1 and 3). Then, we estimated both models making a distinction between these effects (columns 2 and 4). In the linear probability model (column 1), density in adjacent provinces has a similar effect as country-level density in determining the emergence of new industries at the province level. Hence, the magnitude of spillover effects from adjacent provinces is comparable to the spillover effects from the rest of the country. In contrast, in the System-GMM estimation (column 3), density in adjacent provinces has no effect on the development of new industries, while density in the rest of the country (in non-adjacent provinces) has an effect, though much smaller than density at the province level. When we include the effect of density on both new and current products (equation 4), we find in the linear probability model (column 2) that country-level density has a larger impact than density in adjacent provinces on the development of new industries. In contrast, the effect on current industries is similar, though the coefficient for adjacent provinces is statistically not significant. In the System-GMM model (column 4), as was the case in Table 1, the density coefficients of adjacent provinces and the rest of the country are both negative. Now, the effect of adjacent provinces on current products is larger than the effect of the rest of Spain. In sum, we did not find evidence that neighbouring provinces exerted a larger influence than the rest of the country in developing new industries.

To check the robustness of our results, we performed five additional econometric tests. First, following Hausmann and Klinger (2007), we re-estimated equation (3) using the revealed comparative advantage index directly[8], instead of using a dichotomic variable for revealed comparative advantage that takes a value of 0 for no revealed comparative advantage, and a value of 1 for revealed comparative advantage. This means that the $x_{i,p,t}$ coefficient in equation (3) now takes the value of the revealed comparative index which is not upper bounded. This change in the $x_{i,p,t}$ variable precludes the estimation of equation (4), and hence we report regression results only for equation (3). As shown in Table A1 in the Appendix, we find, as expected, that density at the province level has a larger impact on the revealed comparative advantage index than density at the country level. Second, following Hausmann and Hidalgo (2010), we also used a lower threshold (0.5) to determine comparative advantage. Now, those products with a revealed comparative advantage index equal or above 0.5 get a value of one, and the rest of the products get a value of zero. As shown in Table A1, the results of the previous analyses are robust to this reduction in the revealed comparative advantage threshold. Third, we re-estimated the model using a conditional (fixed-effects) logit specification. In this model, each province has the alternative of developing comparative

[8] To avoid the influence of outliers, we remove those observations with a revealed comparative advantage index above the 99th percentile.

Table 2. Regression results on the impact of density in moving to new products and keeping comparative advantage on current products

	(1)	(2)	(3)	(4)	(5)	(6)
Model	Linear probability	Linear probability	Linear probability	System-GMM	System-GMM	System-GMM
Comparative advantage	0.620 (0.014)***	0.600 (0.011)***	0.590 (0.011)***	0.607 (0.009)***	0.388 (0.011)***	0.541 (0.010)***
Density on new products. Country	0.004 (0.003)		0.010 (0.002)***	0.012 (0.001)***		-0.013 (0.001)***
Density on current products. Country	0.028 (0.012)*		0.044 (0.010)***	0.071 (0.013)***		0.021 (0.013)
Density on new products. Province		0.027 (0.003)***	0.028 (0.003)***		0.028 (0.003)***	0.065 (0.003)***
Density on current products. Province		0.068 (0.020)**	0.070 (0.019)***		0.148 (0.010)***	0.150 (0.010)***
Adj. R-squared	0.470	0.474	0.475			
Hansen J statistic p-value				0.000	0.000	0.000
AB test for AR(1) in first differences p-value				0.000	0.000	0.000
AB test for AR(2) in first differences p-value				0.000	0.000	0.000
Obs.	132,790	132,790	132,790	132,790	132,790	132,790

Note: In linear probability model estimations province clustered standard errors in parentheses. In System-GMM estimations robust standard errors in parentheses. Linear probability regressions include province+period and product+period dummies. ***, **, * statistically significant at 1%, 5% and 10% respectively.

Table 3. Regression results on spillover effects

	(1)	(2)	(3)	(4)
Model	Linear probability	Linear probability	System-GMM	System-GMM
Comparative advantage	0.618 (0.004)***	0.592 (0.012)***	0.381 (0.011)***	0.538 (0.010)***
Density in rest of country (non-adjacent provinces)	0.008 (0.002)**		0.005 (0.002)**	
Density in adjacent provinces	0.006 (0.003)*		0.000 (0.010)	
Density in province	0.033 (0.005)***		0.042 (0.003)***	
Density on new products. Country (non-adjacent provinces)		0.009 (0.002)***		-0.006 (0.002)***
Density on current products. Country (non-adjacent provinces)		0.028 (0.008)**		0.026 (0.016)**
Density on new products. Adjacent provinces		0.005 (0.002)*		-0.017 (0.003)***
Density on current products. Adjacent provinces		0.029 (0.015)		0.041 (0.011)***
Density on new products. Province		0.028 (0.003)***		0.054 (0.003)***
Density on current products. Province		0.066 (0.019)**		0.128 (0.011)***
Adj. R-squared	0.473	0.475		
Hansen J statistic p-value			0.000	0.000
AB test for AR(1) in first differences p-value			0.000	0.000
AB test for AR(2) in first differences p-value			0.000	0.000
Obs.	132,790	132,790	132,790	132,790

Note: In linear probability model estimations province clustered standard errors in parentheses. In System-GMM estimations robust standard errors in parentheses. Linear probability regressions include province+period and product+period dummies. ***, **, * statistically significant at 1%, 5% and 10% respectively.

advantage in each of the products included in the database. The probability of developing comparative advantage depends on the (product) characteristics of each alternative: revealed comparative advantage, density at the country-level and density at the province-level. As shown in Table A1, province-level density remains much more important than country-level density in developing comparative advantage in new products. However, we now found that density plays a more significant role in developing new products than in keeping revealed comparative advantage in current products. Fourth, we checked whether our results changed when we use alternative time-intervals. We re-estimated the equations with a longer interval (10 years) and with a shorter interval (4 years), but the results remained the same. And fifth, we recalculated proximity indexes excluding from the sample countries that were very far from Spain's level of economic development. More specifically, in all years, we excluded those countries classified as low-income by the World Bank. Again, our findings remained the same.

6. Conclusions

In this paper, we investigated the process of related diversification at the regional level (i.e. regional branching) by means of the proximity product index developed by Hausmann, Hidalgo and Klinger. Our findings showed that Spanish regions diversified into new industries that are related to the existing set of industries at the regional level, suggesting that new industries make use of capabilities in which regions are specialized. We also found strong evidence that capabilities available at the regional level played a larger role than capabilities available at the country level for the emergence and development of new industries in Spanish regions during the period 1988-2008. Our findings suggest that these capabilities are mobile across regions only to a very limited extent.

In the theoretical part, we explained that this finding is in line with expectation, because we argued, among other reasons, that crucial knowledge transfer mechanisms (like spinoffs, firm diversification, labor mobility and inter-firm networks) that connect new with existing industries, and that may provide a major input for successful regional diversification, are more likely to operate at the regional scale. However, we did not investigate and measure these mechanisms in this study. There is some evidence from studies on the rise of industries that entrepreneurship might be one of the driving forces behind regional branching. These studies show that old sectors give birth to new sectors through entrepreneurship, and that new firms in new industries have a higher survival rate when the entrepreneurs originate from related industries (Klepper, 2007; Boschma and Wenting 2007). How that works out for the other transfer mechanisms is still relatively unexplored. A major challenge for future research is to determine which of these mechanisms actually drive this process of regional branching.

Another issue is the selection of the relatedness indicator to study regional diversification. We made use of the proximity product indicator developed by Hidalgo and others. Other studies have used different approaches, like the Neffke et al. (2011) study on inter-industry relatedness which was based on co-occurrence analysis of products that belong to different industries in the portfolios of plants. Another good alternative is the skill-relatedness measure developed by Neffke and Svensson Henning (2008) who make use of the intensity of labour flows between industries to trace and identify the degree of revealed relatedness between industries. To make use of alternative measures of industry relatedness may provide other robustness checks for the findings found here, but such an exercise could also shed light on different forms of industry relatedness that might have different effects on regional branching.

In this paper, we explored the extent to which the entry of a new industry depends on (a variety of) industries to which it is related. Doing so, we left out other dimensions that might be considered crucial in the process of regional diversification, such as institutional reforms (Hassink, 2005; Strambach, 2010). It would be extremely relevant to explore the influence of institutions and the need for institutional change to enable the development of new industries at the regional level. In that respect, it would also be worthwhile to explore further the extent to which related industries draw on (and require) similar sets of institutions, which could provide an additional explanation for the fact that related industries tend to benefit from each other's co-presence at the regional level. Moreover, we have to investigate more in detail whether this process of regional branching is more true for the entry of some industries than for other industries in regions, just like the economic effect of related variety at the regional level may differ between industries (Bishop and Gripaios, 2010) and the type of activities undertaken within industries (Robertson and Patel, 2007; Brachart et al., 2011).

This brings us to possible policy implications that can be drawn from our study. We will be very cautious here, because we did not investigate the impact of any government intervention on the development of comparative advantage in new products in regions. Nevertheless, our study suggests that some capabilities should be developed locally to raise the probability of developing new industries at the regional level. In that context, it would be wise to target policy intervention at the regional level, because it is at this level where the main assets to diversify successfully are present. This would bring us a step forward in the design of policy programs that are focused on regional diversification, despite all the unpredictability that is part and parcel of the development of new growth paths in regions (Boschma, 2011).

References

Bathelt, H. (2003) Geographies of production: growth regimes in spatial perspective 1 – innovation, institutions and social systems, *Progress in Human Geography* 27: 763-778.

Bathelt, H., A.K. Munro and B. Spigel (2011) Challenges of Transformation: Innovation, Re-bundling and Traditional Manufacturing in Canada's Technology Triangle, Papers in Evolutionary Economic Geography, no. 11.11, Urban and Regional research centre Utrecht (URU), University of Utrecht: Utrecht.

Bernard, A.B. and Jensen, J.B., (2004) Why some firms export, *The Review of Economics and Statistics* 86 (2): 561-569.

Bishop P, Gripaios P (2010) Spatial externalities, relatedness and sector employment growth in Great Britain, *Regional Studies*, 44 (4): 443-454.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics*, 87 (1): 115-143.

Boschma, R. (2011), Regional branching and regional innovation policy, in: K. Kourtit, P. Nijkamp and R. R. Stough (eds.) (2011), *Drivers of innovation, entrepreneurship and regional dynamics*, Springer Verlag, Berlin/Heidelberg, pp. 359-368.

Boschma, R. and K. Frenken (2011) Technological relatedness and regional branching, in: H. Bathelt, M.P. Feldman and D.F. Kogler (eds.), *Beyond Territory. Dynamic Geographies of Knowledge Creation, Diffusion and Innovation*, Routledge, London and New York, pp. 64-81.

Boschma, R.A. and S. Iammarino (2009), Related variety, trade linkages and regional growth, *Economic Geography* 85 (3): 289-311.

Boschma, R.A., A. Minondo and M. Navarro (2011), Related variety and regional growth in Spain. *Papers in Regional Science*, in print, doi: 10.1111/j.1435-5957.2011.00387.x

Boschma, R.A., and R. Wenting (2007) The spatial evolution of the British automobile industry. Does location matter? *Industrial and Corporate Change*, 16 (2): 213-238.

Brachart, M., A. Kubis and M. Titze (2011) Related variety, unrelated variety and regional functions. Identifying sources of regional employment growth in Germany from 2003 to 2008, IWH-Discussion Papers, Halle, Germany.

Buenstorf, G., M. Fritsch and L.F. Medrano (2010) Regional knowledge and the emergence of an industry. Laser systems production in West Germany, 1975-2005, Papers in Evolutionary Economic Geography, 10.16, Utrecht University, Utrecht.

Buerger, M. and U. Cantner (2011) The regional dimension of sectoral innovativeness. An empirical investigation of two specialized supplier and two science-based industries, *Papers in Regional Science*, 90 (2): 373-393.

Carlsson, B., Stankiewicz R. (1991) On the nature, function and composition of technological systems, *Journal of Evolutionary Economics,* 1: 93-118

Eriksson, R. (2011). Localized spillovers and knowledge flows: How does proximity influence the performance of plants? *Economic Geography*, 87(2): 127 –152

Farjoun, M. (1994) Beyond industry boundaries: Human expertise, diversification and resource-related industry groups, *Organization Science*, 5 (2): 185-199.

Feenstra, R.C., Lipsey, R.E., Deng, H. Ma, A.C. and Mo, H. (2005) World Trade Flows: 1962-2000, *NBER Working Paper 11040*, National Bureau of Economic Research, Cambridge, Massachusets.

Feldman, M.P., J. Francis and J. Bercovitz (2005), Creating a cluster while building a firm. Entrepreneurs and the formation of industrial clusters, *Regional Studies*, 39: 129-141.

Fornahl, D., S. Henn and M. Menzel (eds.) (2010), *Emerging clusters*, Cheltenham: Edward Elgar.

Frenken, K., Van Oort, F.G., Verburg, T. (2007) Related variety, unrelated variety and regional economic growth, *Regional Studies*, 41 (5): 685-697.

Gertler, M.S. (2003) Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there), *Journal of Economic Geography* 3: 75-99.

Glaeser, E.L. (2005) Reinventing Boston: 1630-2003, *Journal of Economic Geography* 5: 119-153.

Green, W.H. (2008). *Econometric Analysis. 6th Edition*, Pearson, New Jersey.

Hassink, R. (2005), How to unlock regional economies from path dependency? From learning region to learning cluster, *European Planning Studies*, 13 (4), 521-535.

Hausmann, R. and Hidalgo, C.A. (2010), "Country diversification, product ubiquity, and economic divergence", *CID Working Paper No. 201*.

Hausmann, R. and Klinger, B. (2007), “The Structure of the Product Space and the Evolution of Comparative Advantage”, *CID Working Paper No. 146*

Hidalgo, C.A. (2009), "The Dynamics of Economic Complexity and the Product Space over a 42 year period", *CID Working Paper No. 189*.

Hidalgo, C.A., Klinger, B., Barabási, A.L. and Hausmann, R. (2007), The product space conditions the development of nations", *Science*, 31**7**, 5837, 482-487.

Klepper, S. (2007), Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry. *Management Science*, 53 (4): 616 – 631.

Klepper, S. and K.L. Simon (2000), Dominance by birthright. Entry of prior radio producers and competitive ramifications in the US television receiver industry, *Strategic Management Journal*, 21: 997-1016.

Lawson, C. (1999), Towards a competence theory of the region, *Cambridge Journal of Economics* 23: 151-166.

Markusen, A. (1996), Sticky places in slippery space. A typology of industrial districts, *Economic Geography*, 72 (3): 293-313.

Maskell, P. (2001), Towards a knowledge-based theory of the geographical cluster, *Industrial and Corporate Change*, 10 (4), 921-943.

Maskell, P. and A. Malmberg (1999), The competitiveness of firms and regions. Ubiquitification and the importance of localized learning, *European Urban and Regional Studies* 6(1): 9-25.

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

Neffke, F. and M. Svensson Henning (2008), Revealed relatedness: Mapping industry space. Papers in Evolutionary Economic Geography, no. 8.19, Urban and Regional research centre Utrecht (URU), University of Utrecht: Utrecht.

Nooteboom B (2000) *Learning and innovation in organizations and economies*. Oxford University Press, Oxford.

Porter ME (2003), The economic performance of regions. *Regional Studies* 37 (6&7): 549-578.

Quatraro, F. (2010), Knowledge coherence, variety and economic growth. Manufacturing evidence from Italian regions. *Research Policy*, doi:10.1016/j.respol.2010.09.005

Robertson, P.L., Patel, P.R. (2007) New wine in old bottles: Technological diffusion in developed economies, *Research Policy* 36 (5): 708-721.

Rosenberg N, Frischtak R (1983), Long waves and economic growth. A critical appraisal, *American Economic Review* 73 (2): 146-151.

Simmie, J. and J. Carpenter (eds.) (2007), *Path dependence and the evolution of city regional development,* Working Paper Series No. 197, Oxford Brookes University, Oxford.

Storper, M. (1992), The limits to globalization: Technology districts and international trade, *Economic Geography* 68 (1) : 60-93.

Storper, M. (1995), The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies, *European Urban and Regional Studies*, 2 (3): 191 – 221.

Strambach, S. (2010) Path dependency and path plasticity. The co-evolution of institutions and innovation - the German customized business software industry, in: R.A. Boschma, R. Martin (eds.), *Handbook of Evolutionary Economic Geography*. Cheltenham: Edward Elgar, 406–431.

Tanner, A.N. (2011), The place of new industries. The case of fuell cell technology and its technological relatedness to regional knowledge bases, Papers in Evolutionary Economic Geography, no. 11.13, Urban and Regional research centre Utrecht (URU), University of Utrecht: Utrecht.

Teece, T.D., Rumelt, R., Dosi, G., Winter, S. (1994), Understanding corporate coherence. Theory and evidence, *Journal of Economic Behaviour and Organization,* 23: 1-30.

Table A1. Sensitivity analyses

Sensitivity Analysis	1	1	2	2	2	2	3	3
Dependent variable	RCA Index	RCA Index	RCA at 0.5	RCA at 0.5	RCA at 0.5	RCA at 0.5	Conditional Logit	Conditional Logit
Model	Linear probability	System-GMM	Linear probability	System-GMM	Linear probability	System-GMM	Conditional Logit	Conditional Logit
RCA index	0.632 (0.020)***	0.386 (0.022)***						
RCA Dichotomic (0,1)			0.604 (0.012)***	0.373 (0.010)***	0.580 (0.012)***	0.532 (0.008)***	4.119 (0.088)***	4.317 (0.062)***
Density at country level	0.060 (0.017)**	-0.004 (0.013)	0.021 (0.004)***	0.001 (0.003)			0.418 (0.034)***	
Density at province level	0.166 (0.029)***	0.236 (0.020)***	0.051 (0.004)***	0.061 (0.004)***			0.768 (0.086)***	
Density on new products. Country					0.020 (0.003)***	-0.034 (0.003)***		0.526 (0.038)***
Density on current products. Country					0.065 (0.011)***	0.061 (0.011)***		0.246 (0.050)***
Density on new products. Province					0.045 (0.003)***	0.092 (0.004)***		0.901 (0.078)***
Density on current products. Province					0.080 (0.017)***	0.150 (0.010)***		0.563 (0.113)***
Adj. R-squared	0.475		0.487				0.427	0.429
Hansen J statistic p-value		0.000		0.000		0.000		
AB test for AR(1) in first differences p-value		0.000		0.000		0.000		
AB test for AR(2) in first differences p-value		0.054		0.000		0.000		
Obs.	130,996	130,996	132,094	132,094	132,094	132,094	132,790	132,790

Note: In linear probability model estimations province clustered standard errors in parentheses. In System-GMM estimations and conditional (fixed-effects) logit estimations robust standard errors in parentheses. Linear probability regressions include province+period and product+period dummies. ***, **, * statistically significant at 1%, 5% and 10% respectively