C I R C L E
LUND UNIVERSITY

Papers in Innovation Studies

Paper no. 2015/35

Neighbor regions as the source of new industries

Ron Boschma (ron.boschma@circle.lu.se)
CIRCLE, Lund University; Utrecht University

Víctor Martín (victor.martin@urjc.es)
Universidad Rey Juan Carlos and ICEI

Asier Minondo (aminondo@deusto.es)
Deusto Business School and ICEI

This is a pre-print version of a paper that has been submitted for publication to a journal.

This version: September 2015

Centre for Innovation, Research and Competence in the Learning Economy (CIRCLE)
Lund University
P.O. Box 117, Sölvegatan 16, S-221 00 Lund, SWEDEN
http://www.circle.lu.se/publications

WP 2015/35

Neighbor regions as the source of new industries

Ron Boschma, Víctor Martín and Asier Minondo

Abstract

The development of new industries demands access to local capabilities. Little attention has yet been paid to the role of spillovers from neighbor regions for industrial diversification, nor has the role of network linkages between neighbor regions been investigated. As the spread of capabilities has a strong geographical bias, we expect regions to develop new industries in which their neighbor regions are specialized. To test this hypothesis, we analyze the development of new industries in US states during the period 2000-2012. We show that an US state has a higher probability of developing a comparative advantage in a new industry if a neighbor state is specialized in that industry. We also show that neighbor US states have more similar export structures. This export similarity seems to be explained by higher social connectivity between neighbor states, as embodied in their bilateral migration patterns.

JEL codes: R11, N94, O14

Keywords: new industries, regional branching, diversification, knowledge spillovers, US, regions, exports

Neighbor regions as the source of new industries

Ron Boschma (CIRCLE; Lund University; Utrecht University)

Víctor Martín (Universidad Rey Juan Carlos and ICEI)

Asier Minondo (Deusto Business School and ICEI)

Abstract

The development of new industries demands access to local capabilities. Little attention has yet been paid to the role of spillovers from neighbor regions for industrial diversification, nor has the role of network linkages between neighbor regions been investigated. As the spread of capabilities has a strong geographical bias, we expect regions to develop new industries in which their neighbor regions are specialized. To test this hypothesis, we analyze the development of new industries in US states during the period 2000-2012. We show that an US state has a higher probability of developing a comparative advantage in a new industry if a neighbor state is specialized in that industry. We also show that neighbor US states have more similar export structures. This export similarity seems to be explained by higher social connectivity between neighbor states, as embodied in their bilateral migration patterns.

Keywords: new industries, regional branching, diversification, knowledge spillovers, US, regions, exports

JEL Classification: R11, N94, O14

Acknowledgements: We thank the comments and suggestions by Jacob Holm and other participants at the 2nd Workshop on Industrial Relatedness and Regional Change (Umea University), and participants in Universidad Complutense de Madrid Research Seminar. Asier Minondo acknowledges financial support from the Spanish Ministry of Economy and Competitiveness (MINECO ECO2013-46980-P, co-financed with FEDER), and the Basque Government Department of Education, Language policy and Culture (IT629-13). Victor Martin acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, (HAR2013-40760_R). Ron Boschma acknowledges financial support from the European Union under the FP7 Cooperation Programme: Social Sciences and the Humanities. Grant number 320131.

1. Introduction

The spatial emergence of new industries is high on the scientific and political agenda. Especially in times of economic crisis, regions are searching for opportunities to diversify their industrial structure. An expanding literature claims that the emergence of industries is driven by the degree of relatedness with existing industries in regions, as new industries draw from and recombine local capabilities that are related to theirs (Boschma and Frenken 2011). Recent studies show indeed that new industries emerge systematically from related industries, and that the industrial structure of a regional economy has an impact on the diversification opportunities of regions (Neffke et al. 2011; Boschma et al., 2013; Muneepeerakul et al. 2013; Essletzbichler 2015).

However, a weakness of the related diversification literature is its almost complete focus on local capabilities, as if regions are self-contained entities. In reality, capabilities may spillover to other regions and trigger the diversification process there. At the same time, the spread of capabilities is heavily constrained by geographical distance: knowledge spillovers are more likely to occur between regions that are geographically close (Jaffe et al. 1993). In the related diversification literature, little attention has been paid to the role of spillovers from neighbor regions for diversification, nor has the role of network linkages between (neighbor) regions been investigated. In a recent paper, Bahar et al. (2014) found that a country had a higher probability to develop a comparative advantage in an industry if a neighbor country had a comparative advantage in that same industry, and that the export baskets of neighboring countries tend to look more similar. There exists no study that has systematically analyzed the effect of neighbors on the probability of regions to develop a comparative advantage in an industry.

The goal of this paper is to fill this gap. The paper has two objectives. The first objective is to assess the effect of neighboring regions on regional diversification. Following Bahar et al. (2014), we expect regions to develop new industries in which their neighbor regions are specialized. To test this hypothesis, we analyze the development of new industries in US states during the period 2000-2012. As this paper uses trade data to analyze regional diversification, following previous studies (Hidalgo et al., 2009; Boschma et al., 2013), we define new industries as those in which a US state has a low level of trade specialization at the beginning of the period and develops a strong trade specialization at the end of the period. We show that a US state has indeed a higher probability of developing a new industry if a neighbor state is specialized in that industry, and when the US state is well endowed with local capabilities that are related to that industry. The second objective is to assess whether neighbor regions have a higher similarity in their export structures, and whether social connectivity (as proxied by bilateral migration patterns) can explain the differences in the similarity of export structures across regions. Our study finds support for the claim that the export similarity between neighbor states is higher than the export similarity between non-neighbor states, and that social connectivity between US states is correlated positively with export similarity across states.

The structure of the paper is as follows. In Section 2, we describe the theoretical background of the study. Section 3 introduces the data and the methodology, and estimates the impact of specialization in neighbor regions on developing a comparative advantage in that industry in a region. Section 4 investigates whether neighbor regions have a more similar export structure, and if so, to what extent this result (i.e. export similarity) is determined by bilateral migration patterns between neighboring regions. Section 5 concludes.

2. Emerging industries: local capabilities, neighbor regions and networks

Local capabilities are perceived to be major assets for regions in a globalized world. This has led to a massive research effort to determine which local capabilities matter, and how these can be identified, as some capabilities are intangibles. Maskell and Malmberg (1999) associated local capabilities with a local knowledge base and an institutional setting that are tightly interwoven and the outcome of a long history. Storper (1995) referred to ‘untraded interdependencies’ such as local practices and conventions. These ‘localized capabilities’ have a high degree of tacitness that form a crucial asset for regions because they cannot be easily imitated by other regions (Gertler 2003). As a consequence, regions develop strong technological and industrial specializations that are hard to challenge, because they are deeply rooted in local capabilities.

Region-specific capabilities provide not only crucial assets on which existing specializations can thrive. There is increasing awareness that local capabilities also operate as key source of technological and industrial diversification, that is, they provide potentials for regions to diversify into new technologies and industries. At the same time, local capabilities also set limits to this diversification process: if a region does not possess the capabilities required for a new technology or new industry, it will be close to impossible to develop these.

Recently, attention has turned to local capabilities that provide opportunities to recombine pre-existing technologies or industries and that give birth to new activities. Jacobs (1969) was one of the first to claim that variety in regions conditions the scope for recombinant innovations: the more variety, the higher the potential to make new recombinations. Frenken et al. (2007) argued, however, that many technologies and industries cannot be meaningfully combined: variety must be related, that is, cognitively proximate, because this positively affects the scope for knowledge spillovers and learning (Nooteboom 2000). Therefore, recombinations are more likely to come from technologies or industries that share similar knowledge bases: the more variety of related technologies or sectors in a region, the more learning opportunities for local activities, and the higher the potential for local recombinations across technologies or industries.

Hidalgo et al. (2007) claimed that capabilities do not move easily between countries, and therefore are hard to acquire when missing. Therefore, capabilities at the country level determine which new industries are feasible to develop. Capabilities are captured by what they call the 'product space' which specifies the relatedness between products based on the frequency of co-occurrences of export products at the country level. Hausmann and Klinger (2007) demonstrated that countries expand their export activities by moving into products that are related in 'product space' to their current export products. Their studies also showed that countries with a wide range of related products (i.e. related variety) have more opportunities to diversify into new export products, as their capabilities can be redeployed in a larger number of new products.

Capabilities at the regional level (at the sub-national scale) might be as important for related diversification (Martin and Sunley 2006; Fornahl and Guenther 2010). Boschma and Frenken (2011) referred to 'regional branching' as a type of regional diversification in which new industries or technologies emerge from local recombinations of technologically related activities. They claim that related diversification tends to occur through channels of knowledge transfer that are often geographically bounded, such as entrepreneurial spinoffs and labor mobility. There is indeed substantial evidence that firms that originate from local related industries (either as diversifiers or new spinoff companies) are crucial for the development of new industries in a region (Klepper 2007). Labor mobility is regarded as another key mechanism through which knowledge and skills are transferred across (related) industries (Neffke and Henning 2013). As labor mobility occurs mainly within labour market regions (Eriksson 2011), labor flows between local related industries may initiate and contribute to new recombinations and, thus, act as a powerful potential source of regional branching.

While still little is known about these underlying mechanisms, there is substantial evidence of related diversification at the regional scale. Qualitative case studies show that new industries in regions are often deeply rooted in local related activities (Chapman 1992; Glaeser, 2005). Quantitative studies have provided evidence that regions diversify into industries that are closely related to their existing activities. Neffke et al. (2011) was the first study at the regional level to show that the entry probability of a new industry in a region is positively related with relatedness with existing industries in the region. Follow-up studies have confirmed that relatedness is indeed a driving force behind diversification of regions in new industries (Essletzbichler 2015), new technologies (Van der Wouden 2012; Kogler et al. 2013; Rigby 2013; Boschma et al. 2015; Feldman et al. 2015) and new eco-technologies (Tanner 2014, 2015; Van den Berge and Weterings 2014). Boschma et al. (2013) demonstrated that local capabilities are a more important driver of regional diversification than national capabilities.

Despite all this evidence, one could argue that the weakness of the related diversification literature is its almost complete reliance on local and national capabilities. Although there are good reasons to state that capabilities are locally sticky and hard to copy by other regions (Markusen 1996), it might not be excluded either. In reality, regions are not self-contained entities: they interact with other regions. Capabilities may spillover to other regions and trigger diversification through inter-regional trade (Boschma and Iammarino 2009) and labor mobility of star scientists, key engineers and top managers who embody scientific, technical and managerial competences (Otttaviano and Peri 2006; Saxenian 2006; Trippl 2013). There is increasing evidence that non-regional linkages are indeed key to avoid lock-in in regions (Asheim and Isaksen 2002;Moodysson 2008; Dahl Fitjar and Rodríguez-Pose 2011).

Having said that, we also expect the spread of capabilities to be heavily constrained by geographical distance. As shown first by Jaffe et al. (1993), and later confirmed by other studies (e.g. Anselin et al. 1997; Varga 2000), geography imposes severe barriers to the diffusion of knowledge. Therefore, knowledge spillovers occur more likely within regions, and between regions that are geographically close, than between regions that are geographically far. Bahar et al. (2014) have explored how this rapid geographical decay of (tacit) knowledge diffusion is reflected in patterns of comparative advantage of countries. Contrary to traditional accounts in trade theory that claim that a higher intensity of trade at shorter distances would lead neighboring countries to specialize in different rather than similar industries, Bahar et al. (2014) expect that neighboring countries develop similar specializations instead, because of significant obstacles to knowledge diffusion across large distances. Their study found that export portfolio's of neighboring countries look indeed more similar, even after controlling for similarity in other dimensions than geographical proximity, like factor endowments, cultural factors and demand structures. Moreover, their study showed that a country had a higher probability to develop a comparative advantage in an industry if a neighbor country had a comparative advantage in that same industry before.

The regional diversification literature has not yet paid attention to the possible effect of spillovers from neighbor regions on diversification. There exists no paper that has systematically analyzed the effect of neighbors on the probability of regions to develop a comparative advantage in an industry. Such a study on regional diversification at the sub-national level would also allow to control for country specific effects, such as language, currency or law. Moreover, Bahar et al. (2014) could not exclude the possibility that their findings were driven by factors other than knowledge diffusion, such as social interaction. Head et al. (2014) show that professional ties facilitate the diffusion of knowledge and highlight that these professional ties are geographically biased. In the context of trade, Millimet and Osang (2007) show that the level of bilateral migration is correlated with the amount of trade between US states. Combes et al. (2005) and Garmendia et al. (2012) found that both social and business connectivity facilitated trade between French and Spanish regions respectively. In this paper, we test whether social connectivity, as proxied by bilateral migration, is correlated with export similarity between regions.[1]

In sum, this paper focuses on two research questions. The first concerns the question whether a region has a higher probability of developing a new industry when a neighbor region is specialized in that industry, and when the region is well endowed with local capabilities that are related to that industry. We analyze the development of new industries in US states during the

[1] As explained by Rauch (2001), migrants facilitate bilateral trade and investment because they reduce the information barriers between the host and the home country.

period 2000-2012. New industries are defined as those in which the US state had a low trade specialization at the beginning of the period of analysis and develops a strong trade specialization at the end of the period of analysis. The second question addresses whether export structures between neighbor regions in the US show a higher degree of similarity, and whether the degree of social connectivity between regions is correlated with export similarity between regions.

3. Neighbor states and the development of new industries in US regions

To determine whether the specialization of a US state in an industry facilitates that a neighbor region also specializes in the same industry, following Bahar et al. (2014), we estimate the following regression equation:

$$N_{s,i,t+5} = \alpha + \beta_1 \ln\left(RCA_{ns,i,t}\right) + \beta_2 RCA_{s,i,t} + \beta_3 d_{s,i,t} + \mu_{i,t} + \mu_{s,ns,t} + \varepsilon_{s,i,t} \qquad (0)$$

where $N_{s,i,t+5}$ takes the value of 1 if US state s develops a new industry i between year t and year $t+5$ and zero otherwise. We consider that a new industry i is developed if US state s had a revealed comparative advantage (RCA) below 0.5 at the beginning of the period (t), and a RCA higher than 1 after 5 years ($t+5$).[2] Following Balassa (1965), RCA is determined dividing the share of an industry in a US state exports by the share of that industry in world exports. A RCA higher than 1 denotes that the US state is specialized in that industry. The industries in which state s had a RCA equal or above 1 at t are excluded from the sample.

Our variable of interest is $ln(RCA_{ns,i,t})$, the natural logarithm of the RCA of the neighbor state with the highest RCA in industry i.[3] The RCA of the neighbor state with the highest RCA enters in logarithms to attenuate the bias that might be generated by some extremely large RCA

[2] We performed the empirical analyses for three alternative thresholds for the beginning of the period RCA index: below 1, equal or below 0.2, and equal and below 0.1. As explained in the robustness section, the estimations are robust to the alternative thresholds.
[3] Neighbor states are defined as states sharing both a section of the border or a point of the border.

indexes.[4] We expect the coefficient β_1 to be positive. It is important to point out that the RCA of the neighbor captures the net effect that the neighbor has on the probability of developing a new industry. The main argument of our paper is that neighboring regions contribute to the development of new industries in which they are specialized through spillover effects. However, regions might also hinder the development of industries in which they are specialized in neighbor regions due to competition effects. $RCA_{s,i,t}$ is the comparative advantage of state *s* in industry *i* at the beginning of the period. As this variable is constrained between zero and 0.5, it does not need to be transformed into natural logarithms; besides, using the absolute RCA value allows to include the industries with a RCA equal to zero at *t* in the sample.

The variable $d_{s,i,t}$ denotes the density around industry *i* at the beginning of the period. Density measures to what extent a US state has the capabilities to develop the new industry *i*. A state will have a larger probability to possess those capabilities if the new industry is close to the industries in which the state is specialized; this closeness is measured by the proximity index developed by Hidalgo et al. (2007). Boschma et al. (2013) show that regions will have a larger probability to develop a new industry if their productive structure is close to this new industry. Algebraically, density is obtained through the following expression,

$$d_{S,i,t} = \frac{\sum_j \varphi_{i,j,t} x_{s,j,t}}{\sum_j \varphi_{i,j,t}} \tag{1}$$

where $x_{s,j,t}$ takes the value of 1 if state *s* has a comparative advantage in product *j* at time *t* and zero otherwise, and $\varphi_{i,j,t}$ is the proximity index between products *i* and *j* at time *t* calculated as,

$$\varphi_{i,j,t} = \min\left\{P(RCAx_{i,t} \mid RCAx_{j,t}), P(RCAx_{j,t} \mid RCAx_{i,t})\right\} \tag{1}$$

where $P(RCAx_{i,t} \mid RCAx_{j,t})$ is the conditional probability of having a comparative advantage in product *i*, given that the US state has a comparative advantage in product *j*, and $P(RCAx_{j,t}$

[4] We also transform the RCA variable into a Revealed Symmetric Comparative Advantage (RSCA) variable (Laursen, 2015). The RSCA is defined as (RCA-1)/(RCA+1) and has a [-1,1] range. Our results are robust to this transformation.

$|RCAx_{i,t})$ is the conditional probability of having a comparative advantage in product *j*, given that the US state has a comparative advantage in product i . If a US state has a comparative advantage in all goods related to product *i*, density will take the value of one. However if US state *s* does not have a comparative advantage in any of the products related to product *i*, density will take the value of zero. Finally, we control for year-specific industry fixed effects ($\mu_{i,t}$) and year-specific US state+neighbor US state fixed effects ($\mu_{s,ns,t}$); α is a constant and $\varepsilon_{s,i,t}$ is the random error term.

The model is estimated with a linear probability model. An advantage of this model is that it can handle the large number of fixed effects of our regression equation.[5] In particular, we use the reg2hdfe Stata command developed by Guimarães and Portugal (2010). However, the limitation of the linear probability model is that the effect of independent variables on the dependent variable is constant. In addition to that, the linear probability model can yield predicted probabilities below zero and above one. Moreover, the linear probability model is inherently heteroskedastic. In order to control for heteroskedasticity, we estimate the model with clustered standard errors at the state+neighbor state level. An alternative to the linear probability model is the system-GMM model, which addresses the endogeneity problems that might exist in our sample. However, as we only had two time periods, we cannot estimate this model.

To calculate US states RCA and density, we combine data on US state-level exports from the US Census Bureau Database and world exports from the Comtrade database. Our data uses the Harmonized System 4-digit disaggregation, which distinguishes 1,268 products (industries). We exclude from the sample the US states that do not have neighbor US states: Alaska and Hawaii. Figure 1 presents the histogram of the average number of new industries that emerge in a US state in a 5 years window. The histogram follows a normal distribution. Figure 2 shows the average number of new industries developed every 5 years in US states. A darker color implies a higher number of new industries. On average, a US state develops 36 new industries every 5 years, with a standard deviation of 11. The unconditional probability to develop a new industry is 3.8%. The states that add more industries per period are Colorado and Virginia, and the states that add a fewer number of industries are Louisiana and West Virginia. While there is not a clear

[5] Probit and logit models, due to the incidental parameters problem, can lead to biased and inconsistent estimates in the presence of a large number of fixed effects (Greene, 2008).

pattern of spatial concentration, the regional divisions that develop a higher number of new industries are New England (40) and South Atlantic (39), while the regional divisions which develop a lower number of new industries are the Pacific (25) and West South Central (29). Regarding new industries, most of them belong to chemicals (Section VI of the Harmonized Classification), machinery and electrical (Section XVI), and metals (Section XV). In particular, within chemicals, the industries that appear as new a higher number of times are inorganic chemicals (code 28), organic chemicals (code 29) and miscellaneous chemicals products (code 38); within machinery and electrical, nuclear reactors, boiler, machinery and chemical appliances (code 84), and electrical machinery and equipment (code 85); within metals, iron and steel (codes 72 and 73) and aluminum (code 76).

Table 1 presents the results of our baseline regressions. In columns (1) and (2) the dependent variable is whether US state *s* develops a new industry *i* between year *t* and year *t+5*. As expected, we find that the RCA index of the neighbor state with the highest RCA coefficient is positive and statistically significant. This result shows that the probability of developing a new industry in a US state is positively correlated with the specialization of a neighbor state in that industry. To measure the economic significance of this latter figure, we have to compare it with the unconditional probability of developing a new industry: 3.8%. A standard deviation increase in the (log) RCA of the neighbor leads to a 21% increase in the probability of developing a new industry ([2*0.004]/0.038). The initial RCA of a state in industry *i* is also positively correlated with the development of a new industry. The density coefficient is also positive and statistically significant, confirming that having a trade specialization in industries that are close to the new industry facilitates the development of this new industry. In fact, a standard deviation increase in density leads to a 59% increase in the probability of developing a new industry ([0.06*0.373]/0.038). Hence, the effect of density on the probability of developing a new industry is almost three times larger than the effect of the neighbor with the highest RCA index.

In column (1), we assume that there is a log-linear relationship between the neighbor state RCA index and the probability of developing a new industry. However, it might be the case that a state should have a minimum RCA level to exert an influence on neighbor states. It might be also the case that once a state reaches a RCA level, further increases in the RCA level do not increase knowledge spillovers. To capture these effects, we define five intervals for the neighbor state

RCA index: 0-0.5; 0.5-1; 1-2; 2-4; and more than 4. Except for the last, differences between intervals are constant in relative terms. To visualize the relationship between neighbor RCA index and the probability to develop a new industry, we draw a step function with the estimated coefficients for every interval.[6] If there was a linear relationship between the probability of developing a new industry and the neighbor state RCA index, the height of the steps should be the same. As shown in Figure 3, this is not the case. There is a similar increase in the probability of developing a new industry when the RCA of the neighbor increases from the 0-0.5 range to the 0.5-1 range, and when it increases from the 0.5-1 range to the 1-2 range. However, the probability of developing a new industry increases more than proportionally when the RCA index of the neighbor rises to the 2-4 range, and even more when the RCA index is larger than 4. This result points out that the probability to develop a new industry increases when the neighbor has achieved a high degree of specialization (more than four times the average trade specialization) in the industry. For example, moving from a neighbor in the 0-0.5 RCA index interval to a neighbor in the =4> RCA index interval would increase the probability of developing a new industry, over the unconditional probability, by 58% (0.022/0.038). Estimates for additional intervals, not reported in the figure, suggest that the influence of the neighbor does not increase further when RCA indexes are higher than 4.

In columns 3 and 4, we substitute the dependent variable in equation (1) with the annual average growth rate in industry *i* RCA index. The sample for this analysis is composed by all industries whose initial RCA index was below 1 at the beginning of the period. Our expectation is that the larger the neighbor state industry *i* RCA index the larger the growth in industry *i* RCA index. The neighbor RCA index coefficient in column 3 is positive and statistically significant, confirming the expectation. According to this coefficient, a standard deviation increase in the neighbor (ln) RCA index would lead to a 1.2 percentage point increase in the average annual RCA index growth. Note that now the (initial) RCA index is negative; this result is sensible, as percentage increases are easier to achieve if the initial RCA level is lower. The density coefficient remains positive and statistically significant, and its value increases substantially. In column (4), we perform estimations for different levels of neighbor state RCA indexes. As shown in Figure 4, the height between steps is similar. This points out that there is an almost log-linear

[6] Intervals enter the regression equation as dummies. The omitted category is the 0–0.5 interval.

relationship between increases in the neighbor state RCA index and the average annual growth rate in the RCA index.

It is interesting to compare our results with the country-level estimations in Bahar et al. (2014). They use a sample of 123 countries for the year 2000. Their neighbor RCA coefficient is equal to ours: 0.004. This result is surprising, because we expected a higher neighbor RCA coefficient in the regional sample than in the country sample, as barriers to knowledge flows ought to be lower between regions than between countries. We also find that the density coefficient in our estimation (0.373) is three times higher than in their estimation (0.130; Table 8 – Panel A, Specification 1). These differences are in line with Boschma et al. (2013), that concludes that the productive structure has a much larger influence on the development of new industries at the regional level than at the national level.

In order to test the robustness of our results, we perform some additional estimations. First, we only consider as jumps those industries that keep, at least, a comparative advantage one year after the jump.[7] As shown in Table 2 - column 1, the neighbor state RCA index coefficient drops to 0.003, but remains statistically significant. Second, we use an alternative definition for new industry. Now, a US state develops a new industry if at the beginning of the period the RCA index was equal or below 0.2 and at the end of the period it was higher than 1. The coefficient drops from 0.004 (Table 1 - column 1) to 0.003 (Table 2 - column 2), but remains statistically significant.[8] Third, we test whether the positive correlation between the neighbor RCA index and the development of new industries occurs by chance. To test this hypothesis, following Bahar et al. (2014), we pick each state's neighbors randomly. The sole condition is that the number of neighbors picked at random should be the same as the actual number of neighbors a US state has. We generate random neighbors 500 times, and each time we select the RCA index of the random neighbor with the highest RCA index. Then, we average the RCA indexes selected in each of the 500 iterations and introduce that value into the regression.[9] As shown in Table 2, columns (3) and (4), the neighbor RCA coefficients are negative and statistically significant. These results confirm that the positive correlation between neighbors RCA and the development of new industries is

[7] We use the year 2008 for the first interval and 2013 for the second interval.
[8] We also estimate the model with ten-fold jumps. The main conclusions are not altered.
[9] In these estimations the year-specific state+neighbor state fixed effects are also chosen at random.

not the result of a random event.[10] Fourth, we analyze whether results are robust to using a 10-year interval instead of a 5-year interval. As shown in columns (5) and (6), the neighbor RCA index coefficient remains positive and statistically significant.

In the baseline and robustness analyses, we only assess the contribution of the neighbor with the highest RCA to the development of new industries. We also explored whether the differences in the number of neighbors and their combined RCA also influences the probability of developing a new industry. First, we analyze whether the total RCA of the neighbors influences the probability of developing a new industry. To estimate this equation, we remove the year-specific US state+neighbor US fixed effects and substitute them by year-specific US state fixed effects. Table 3 – Column 1 shows that the coefficient for total neighbor RCA is positive and statistically significant, and its value is the same as in the baseline regression (Table 1 – column 1). In Column 2, we introduce the number of neighbors. To estimate this regression, we remove the year-specific state fixed effects because they are perfectly collinear with the number of neighbors. We find that the number of neighbors has no effect on the combined RCA coefficient, suggesting that it is the total RCA of neighbor regions, and not the number of neighbors, which determines the probability of developing a new industry. Note that when we remove the year-specific state level fixed effects, the density coefficient becomes not significant. This result suggests that it is the differences in density within the state, rather than the absolute levels of density, which drives the development of new industries. In Column (3), we introduce the total RCA of the neighbors that have, at least, an RCA equal or higher than one. Note that this condition reduces severely the number of observations in the sample. We find that the combined RCA coefficient, 0.006, is higher than in Column (1), confirming, as we saw in Figure 3, that the influence of neighbors will be higher if they have a comparative advantage in the product. Note that the density coefficient, although positive, is statistically not significant. This result might suggest that if the RCA of the neighbors in a product is high enough, it might overcome the limitations of the local industrial structure. However, we should be careful with this interpretation because the truncation of the database might lead to biased results, as we only keep those industries in which neighbor regions are specialized. In Column (4), we introduce the number of neighbors with a RCA equal or higher than 1. Note that in this estimation, we can keep the year-

[10] Alternatively, we also ran a different regression for each of the 500 random draw of neighbors. In none of these regressions the coefficient for the neighbor with the highest RCA was positive and statistically significant.

specific state fixed-effects, because the number of neighbors with RCA equal or higher than one varies across products. Remarkably, the combined RCA coefficient becomes statistically not significant and the number of neighbors is positive and statistically significant. According to this result, it is the number of neighbors with comparative advantage in the product, rather than the total RCA of the neighbors with a comparative advantage, which determines the probability of developing a new industry. This result might be explained by the fact that effect of the RCA seems to reach a plateau once the RCA is higher than four (see Figure 3); as 99% of US states have four or less neighbors with a comparative advantage in an industry, the probability of developing an industry might be more correlated with the number of neighbors with RCA equal or higher than 1 than the combined RCA of the neighbors with a comparative advantage.

Columns (4) to (8) re-estimate the regression equations using the growth rate of the RCA of the industries with an RCA below 1 as dependent variable. Similar to the jump analyses (Columns 1 to 4), the combined total RCA is positive and statistically significant in Columns (5), (6) and (7). The coefficient for the number of neighbors with a comparative advantage is also positive and statistically significant in Column (8). However, in the growth estimation, the total combined RCA is also positive and statistically significant, although by a small margin. It is interesting to see that in the growth estimations, density is always positive and statistically significant. This result points out that US states increase their specialization in industries that use their current capabilities.

The dynamic analyses conclude that there is a positive correlation between developing a new industry and the trade specialization of neighbor regions in the same industry. In the next section, we adopt a static view and analyze whether neighbor regions have also more similar export patterns.

4. Similarities in export structures between neighbor US states

If knowledge diffuses at short distances, adjacent regions should share more knowledge. Hence, we would expect adjacent regions to have a more similar export pattern than non-adjacent regions. To test this hypothesis, following Bahar et al. (2014), we calculate the following export similarity index:

$$S_{S,S'} = \frac{\sum_i \left(r_{S,i} - \overline{r}_S\right)\left(r_{S',i} - \overline{r}_{S'}\right)}{\sqrt{\sum_i \left(r_{S,i} - \overline{r}\right)^2 \sum_i \left(r_{S',i} - \overline{r}_{S'}\right)^2}} \quad (1)$$

This export similarity index is based on the Pearson correlation coefficient, where r_{si} is the log of the RCA of state *s* in industry *i* and $\overline{r_s}$ is the average of $r_{s,i}$ over all industries in state *s*.[11] This index will take positive values if US state *s* and US state *s´* are specialized in similar industries, and negative values if they are specialized in different industries.

To test whether US states that are geographically closer have more similar export structures, we estimate the following regression equation:

$$S_{s,s',t} = \alpha + \beta \ln dist_{s,s'} + \gamma n_{s,s'} + \delta' X + \mu_{s,t} + \mu_{s',t} + \varepsilon_{s,s',t} \quad (1)$$

where $S_{s,s',t}$ is the export similarity between US state *s* and US state *s´* at time *t*, α is a constant, $dist_{s,s'}$ is the distance between state *s* and state *s´*, $n_{s,s'}$ is a dummy variable that takes the value of 1 if state *s* is adjacent to state *s´* and zero otherwise, *X* is a vector of controls, $\mu_{s,t}$ is a year- specific *s* state fixed effect, $\mu_{s,t}$is a year-specific *s´* state fixed effect, and $\varepsilon_{s,s',t}$ is the random error. Controls vector *X* includes differences between US states in GDP per capita, physical capital per worker, human capital per worker and land per worker; it also includes the bilateral trade between US states and the accumulated 5-year bilateral migration between US states.As mentioned before, previous studies have shown that the diffusion of knowledge is deterred by geographical distance (Jaffe et al., 1993). Hence, we expect the similarity between export structures to be negatively correlated with distance. Boschma and Frenken (2011), and Neffke et al. (2011) also point out that diversification drivers, such as spin-offs, workers mobility or entrepreneurship, are geographically biased. According to the classical international trade theories, countries or regions specialize in products in which they are relatively productive, or in products that use its abundant factors intensively. Therefore, we expect differences in factors

[11] More specifically, r_{si} is the log of the RCA+0.1. The RCA enters in logs because very large RCA indexes might bias the covariance index. The 0.1 fraction is added to include in the analysis the industries whose RCA index is zero.

endowments and productivity, proxied by GDP per capita, to have a negative impact on the similarity of export structures across states. Finally, the literature has highlighted that networks also facilitate the diffusion of knowledge (Rauch, 2001). Hence, we expect more socially linked regions to have higher export similarity. Following previous studies, we proxy the degree of social connectivity with the level of bilateral migration between regions (Combes, 2005; Millimet and Osang, 2007).

Distance between US states is calculated as the driving distance between the two main cities of each US state. We calculate these distances using the Microsoft Mapoint 2012 and CDXZipStream software. Data for GDP per capita, relative factor endowments and bilateral trade come from the US Census Bureau. Data on bilateral migration is obtained from the U.S. Internal Revenue Service (http://www.irs.gov). Table A1 in the appendix presents summary statistics of these variables.

Figure 5 compares the density function of export similarities between non-neighbor US states with the density function of export similarities between neighbor states. Clearly, the density function for neighbor states is to the right of the density function of non-neighbor states, pointing out that neighbor states have, on average, a higher similarity in the export structure than non-neighbor states. To confirm this hypothesis, we estimate equation (3) pooling data for the years 2002 and 2007. Table 4 presents the results of the estimations. To facilitate the reading of the results, we normalize the export similarity index, with mean zero and unit standard deviation. In column (1), we estimate the regression only with distance and the fixed effects as independent variables. As expected, the distance coefficient is negative and statistically significant, confirming a negative correlation between distance and similarity in export structures.. In particular, a standard deviation increase in the (log) of distance leads to 0.84 standard deviations reduction in the export similarity index (1.1*0.763).

In column (1), we assume that there is a log-linear relationship between the export similarity index and distance. To test the validity of this hypothesis, we divide the distance into intervals. We define six intervals, starting from 0-100 km., 100-200, 200-400, 400-800, 800-1600, and more than 1600.[12] The distance doubles between an interval and the next, so the increase in distance is constant in relative terms. Figure 6 presents a step figure showing the

[12] The maximum distance in our database is 4,924 kilometers.

change of the distance coefficient for each distance interval (0-100 is the omitted interval). We can see that the coefficients for the 100-200 interval and for the 200-400 interval have the zero value within their $\pm$ 5% confidence interval. From the 400-800 interval onwards, the negative value of the distance coefficient increases at a relatively constant proportion. The figure points out that up to 400 km., distance seems to exert a mild negative effect on export similarity. After that interval, the negative effect rises proportionally with distance. Due to this non-linear relationship between distance and similarity, in the rest of estimations, distance is introduced in intervals in the regression.

In column (3), we introduce the neighbor state dummy variable in the regression. The coefficient is very large and statistically significant: the export similarity between neighbor states is around 0.4 standard deviations higher than the export similarity between non-neighbor states. This large coefficient is striking providing that the regression already controls for the lower distance between neighbor states. In Column (4), we introduce the number of commuting zones that are shared between neighbor states.[13] Our expectation is that the number of shared commuting zones is positively correlated with the similarity in the export structure between states. Shared commuting zones allow a higher interaction between professionals of neighbor states, which raises the probability that neighbor regions will have access to similar capabilities and, hence, be able to specialize in similar goods. As expected, the coefficient for the number of commuting zones is positive and statistically significant, confirming that the higher the number of shared commuting zones, the higher the similarity in export structures. We can see, as well, that there is a drop in the value of the neighbor coefficient, pointing out that this coefficient was capturing the effect of the number of shared commuting zones.

In column (5), we introduce additional controls that might explain why neighbor regions have a higher export similarity. These include controls on relative factor endowments: human capital, capital per worker and land per worker, differences in income per capita levels, and the amount of bilateral trade between US states.. The coefficients reported in Table 4-column 5 are in line with the theoretical expectations: differences in income per capita and in factor endowments are negatively correlated with export similarity. Classical theories of trade also predict that

[13] These data were obtained from United States Department of Agriculture Economic Research Service http://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas.aspx. In the year 2000 there were 3141 commuting zones in the US, of which only 3.6% where shared between states.

regions should trade more with more dissimilar regions in terms of export structure. In contrast to this expectation, we get a positive coefficient for the amount of bilateral trade. This result might point out that other factors, such as the existence of economies of scale and love for variety might characterize the pattern of trade between US states (Helpman and Krugman, 1985), or that production processes within an industry are fragmented across US states (Hillberry and Hummels, 2008). The introduction of the new control variables reduces substantially the value of the neighbor state coefficient, declining from 0.304 (column 4) to 0.194 (column 5) (36% reduction). In contrast, the coefficient for shared commuting zones rises from 0.139 to 0.199 (43% increase): once we control for differences in productivity and factor endowments, sharing a commuting zones has a larger positive effect on export similarity.

We also compared our results on the similarities between US states exports with those obtained by Bahar et al. (2014) on the similarities between country exports. When they introduce all the control variables, the distance coefficient is -0.316 and the adjacent-country coefficient is 0.650 (Table 3-Specification 3). Our equivalent (log) distance coefficient is -0.475, and our adjacent-state coefficient is 0.414. These coefficients are obtained running our specification (3), and using distances (in logs) instead of distance intervals. The differences in the distance coefficient might be explained by the fact that we use driving distances whereas Bahar et al. (2014) use great circle distances. In fact, if we use great circle distances between US states centroids, our distance coefficient drops to -0.199 and our neighbor coefficient rises to 0.662, a result very similar to that found by Bahar et al. (2014). This result points out that additional forces that might explain the geographical bias of knowledge transfer seem to have similar effects within countries and across countries.

As discussed in Section 2, this may be due to social connectivity (Millimet and Osang 2007; Garmendia et al. 2012). To test whether social connectivity is correlated with export similarity between regions, we introduce the level of bilateral migration between US states in the regression. As explained before, this data is obtained from the U.S. Internal Revenue Service (IRS). The IRS determines whether a tax payer has migrated to another state comparing her address in tax-year *t* and tax-year *t+1*. The level of bilateral migration is calculated as the accumulated migration flows in the five years previous to the analysis. So, for the year 2002, we use the accumulated figure for the period 1997-2001, and for the year 2007 analysis, we use the accumulated figure for the period 2002-2006. As shown in Table 4-Column 6, the coefficient for bilateral migration is positive and statistically significant, as expected. This result confirms that social connectivity is correlated positively with export similarity. The remarkable result is that when we control for bilateral migration, the neighbor coefficient and the number of shared commuting zones coefficient become statistically not significant. As bilateral migrations are larger between bordering US states, if we do not control for this variable, the border dummy and the number of shared commuting zones capture the positive effect that migration has on knowledge spillovers and on export similarity. Although it is not reported in the paper, all the distance interval coefficients become statistically not significant as well. This latter result points out that the negative effect of distance on export similarity is explained by the influence this variable has on bilateral migration.

To test the robustness of our result, in column (7), we use an alternative index to measure similarity between export structures, the Finger & Kreinin index (Finger and Kreinin, 1979). This index is defined as follows:

$$FKS_{s,s'} = \sum_i min\left(\left[x_{s,i} \Big/ \sum_i x_{s,i} \right], \left[x_{s',i} \Big/ \sum_i x_{s',i} \right]\right) \qquad (1)$$

The Finger and Kreinin similarity index (FKS) is the sum of the minimums of each industry *i* export share for a pair of US states (*s* ,*s´*). The index takes the maximum value of 1 when the distribution of exports across industries is the same in state *s* and *s´*, and takes the minimum value of zero if there is no overlap in the distribution of exports. As shown in column 7, the use of the Finger-Kreinin index does not alter the main results of previous estimations.

5. Conclusions

If knowledge spillovers decay with distance, we expect regions to develop new industries in activities in which their neighbor regions are specialized. We confirm this hypothesis using data for US states during the period 2002-2012. In particular, a US state with a neighbor highly specialized in an industry has a 58% higher probability to develop that industry than another US state with a neighbor poorly specialized in that industry. From a static perspective, we also show that sharing a border with another US states raises the similarity between export structures by 0.43 standard deviations. Our analyses suggest that the similarity in export structures between neighbor regions is positively correlated with higher social connectivity. These findings complement the regional diversification literature that has focused almost entirely on the importance of local capabilities in related industries. Our analyses also replicate this finding: density had a strong and positive effect on developing a new industry in a US state, and this density effect is stronger than the effect of neighbor regions that are specialized in that industry

As in any study, our study also generates new research challenges. First, we have referred to knowledge spillovers more in general to claim that regions are more likely to develop new industries when their neighbor regions are specialized. This finding is interesting, as well-known studies on the geography of knowledge spillovers in the US (e.g. Jaffe et al. 1993; Anselin et al. 1997; Varga 2000; Crescenzi et al. 2007) have shown that knowledge spillovers tend to occur mainly within US states, and hardly cross US state boundaries. These studies have focused largely on academic knowledge spillovers, based on patent and research collaboration data. As our study works with trade data, we could focus on all industries in the whole economy. In that sense, it would be interesting to analyze more in detail whether the effect of neighboring regions differs between industries, and between low and high-tech industries in particular.

Our study has not measured knowledge spillovers *per se*, and through which channels (such as observational learning, entrepreneurship, labour mobility, trade patterns, research collaborations, et cetera) this spatial diffusion process across neighboring regions actually takes place. In this study, we looked at the role of social connectivity between regions as a potential channel. The next step is to explore which channels on which geography imposes barriers on knowledge diffusion, can be held responsible for this neighboring effect, and how they drive regional branching. This would shed light on how network relations more in general shape the diffusion of capabilities across regions, and how these affect diversification opportunities of regions, a topic which has not yet been studied in a systematic way (Crespo et al. 2014). Although neighboring regions are likely to be more connected, it is a fact that regions are also connected over large geographical distances, especially large urban centers (see e.g. Ponds et al. 2007). While this role of non-regional linkages is covered in our analytical framework through fixed effects, more explicit attention on their role in the regional branching process is an interesting future research avenue.

Another research challenge is to include the role of institutions which has shown to have an impact on diversification at the national scale (Boschma and Capone 2015). Institutions could be included in our analytical framework both as a local capability variable (next to density, which captures local related capabilities) and as a similarity indicator, capturing the effect of institutional or cultural proximity between regions.

References

Anselin, L., A. Varga and Z. Acs (1997). Local geographic spillovers between university research and high technology innovations, *Journal of Urban Economics* 42: 422-448.

Asheim, B.T. and Isaksen, A. (2002). Regional innovation systems. The integration of local 'sticky' and global 'unbiquitous' knowledge, *Journal of Technology Transfer* 27: 77-86.Bahar, D., Hausmann, R. and Hidalgo, C.A. (2014). Neighbors and the evolution of the comparative advantage of nations: Evidence of International Knowledge Difussion?, *Journal of International Economics* 92 (1): 111-123.

Balassa, B. (1965). Trade Liberalization and Revealed Comparative Advantage, *Manchester School of Economics and Social Studies* 33: 99-123.

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

Boschma, R. and Frenken, K. (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.

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

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.

Chapman, K. (1992) Continuity and contingency in the spatial evolution of industries. The case of petro-chemicals, *Transactions of the Institute of British Geographers* 17 (1): 47-64.

Combes, P.P., Lafourcade, M. and Mayer, T. (2005) The trade-creating effects of business and social networks: evidence from France, *Journal of International Economics* 66 (1): 1-29.

Crescenzi, R., Rodriguez-Pose and M. Storper (2007) The territorial dynamics of innovation: a Europe-United States comparative analysis, *Journal of Economic Geography* 7 (6): 673-709.

Crespo, J., Suire R. and Vicente J. (2014) Lock-in or lock-out? How structural properties of knowledge networks affect regional resilience. *Journal of Economic Geography* 14(1), 199-219.

Dahl Fitjar R. and Rodríguez-Pose A. (2011) When local interaction does not suffice; sources of firm innovation in urban Norway. *Environment and Planning A* 43 (6): 1248–67.

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

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

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

Finger, J.M. and Kreinin, M.E. (1979). A measure of ‘export similarity’ and its possible uses, *Economic Journal* 89, 356, 905-912.

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

Garmendia, A., C. Llano, A. Minondo and F. Requena (2012). Networks and the disappearance of the intranational home bias, *Economic Letters* 116: 178-182.

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.

Guimarães, P. and Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. Stata Journal 10 (4): 628-649.

Head, K., Li Y.A. and Minondo, A. (2015). Geography, Ties and Knowledge Flows: Evidence from Citations in Mathematics, unpublished paper

Helpman, E. and Krugman, P. (1985). *Market Structure and Foreign Trade. Increasing Returns, Foreign Competition and the International Economy*, The MIT Press, Cambridge, Massachusetts.

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

Hillberry, R. and Hummels, D. (2008). Trade responses to geographic frictions: A decomposition using micro-data, *European Economic Review* 52 (3): 527-550.

Jacobs, J. (1969) *The Economy of Cities*. New York: Vintage Books.

Jaffe, A.B., Trajtenberg, M. and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations, *The Quarterly Journal of Economics* 108 (3): 577-598.

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.

Kogler, D. F., Rigby D. L. and Tucker, I. (2013) Mapping knowledge space and technological relatedness in US Cities. *European Planning Studies*, 21 (9):1374-1391

Laursen, K. (2015). Revealed comparative advantage and the alternatives as measures of international specialization, *Eurasian Business Review*, **5**, 1, 99-115.

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

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

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.

Millimet, D.L. and Osang, T. (2007). Do state borders matter for U.S. intranational trade? The role of history and internal migration, *Canadian Journal of Economics* 40 (1): 93-126.

Moodysson, J. (2008) Principles and practices of knowledge creation: On the organization of 'buzz' and 'pipelines' in life science communities. *Economic Geography* 84 (4): 449–69.

Muneepeerakul, R., J. Lobo, S.T. Shutters, A. Gomez-Lievano and M.R. Qubbaj (2013), Urban economies and occupation space: can they get “there” from “here”?, *PLoS ONE* 8(9): e73676. doi: 10.1371/journal.pone.0073676.

Neffke F., Henning M. (2013) Skill-relatedness and firm diversification, *Strategic Management Journal* 34(3): 297-316

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: 237-265.

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

Ottaviani, G. and G. Peri (2006). The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography* 6: 9-44.

Ponds, R., F.G. Van Oort and K. Frenken (2007) The geographical and institutional proximity of research collaboration, *Papers in Regional Science* 86: 423-443

Rauch, J.E. (2001) Business and Social Networks in International Trade, *Journal of Economic Literature***39**, 4: 1177-1203.

Rigby, D. (2013) Technological relatedness and knowledge space: Entry and exit of US cities from patent data, *Regional Studies*, forthcoming, DOI:10.1080/00343404.2013.854878

Saxenian, A.L. (2006) *The new Argonauts. Regional advantage in a global economy*, Cambridge (MA): Harvard University Press.

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.

Tanner, A.N. (2014) Regional branching reconsidered: Emergence of the fuel cell industry in European regions, *Economic Geography*, forthcoming, DOI: 10.1111/ecge.12055

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.

Trippl, M. (2013) Scientific mobility and knowledge transfer at the interregional and intraregional level. *Regional Studies* 47 (10): 1653-1667.

Van den Berge, M. and A. Weterings (2014) Relatedness in eco-technological development in European regions, Papers in Evolutionary Economic Geography, no. 14.13, Utrecht University, Utrecht.

Varga, A. (2000) Local academic knowledge spilovers and the concentration of economic activity, *Journal of Regional Science* 40: 289-309.

Wouden, F. van der (2012), The role of relatedness in economic variety on knowledge production of U.S. cities between 1975 and 2005, MA thesis, Department of Economic Geography, Utrecht University.

Figure 1. Number of new industries that emerge in a US state during a 5 year window

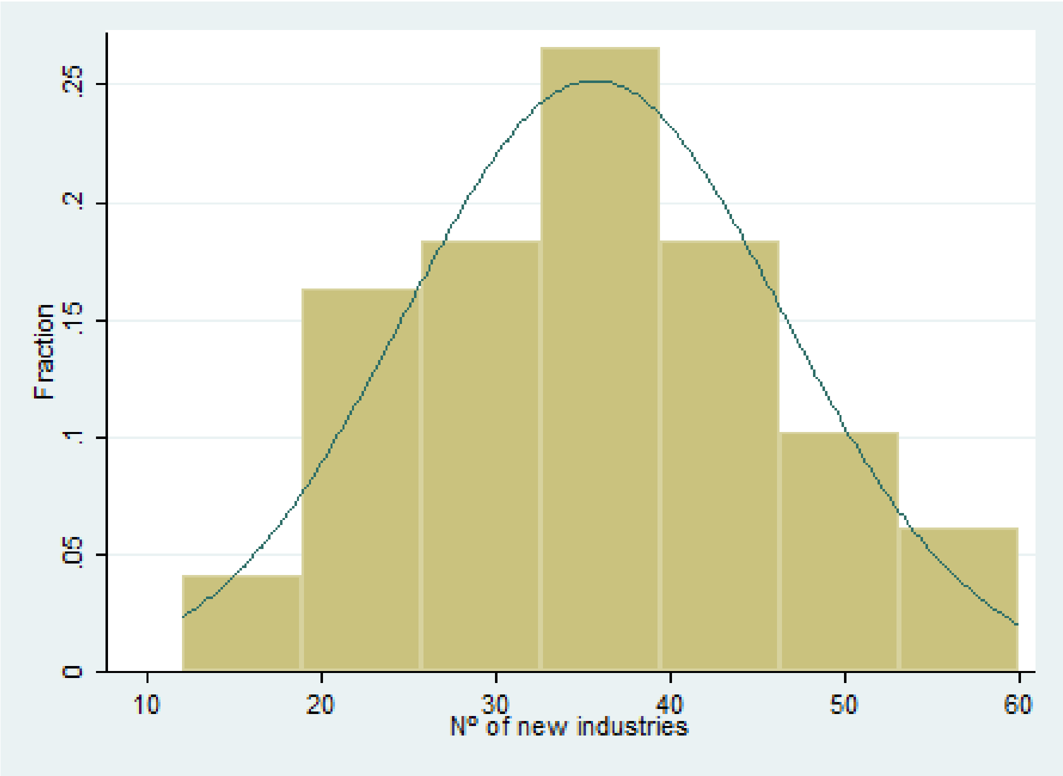

Figure 2. Average number of jumps in a 5-year interval, 2002-2012.

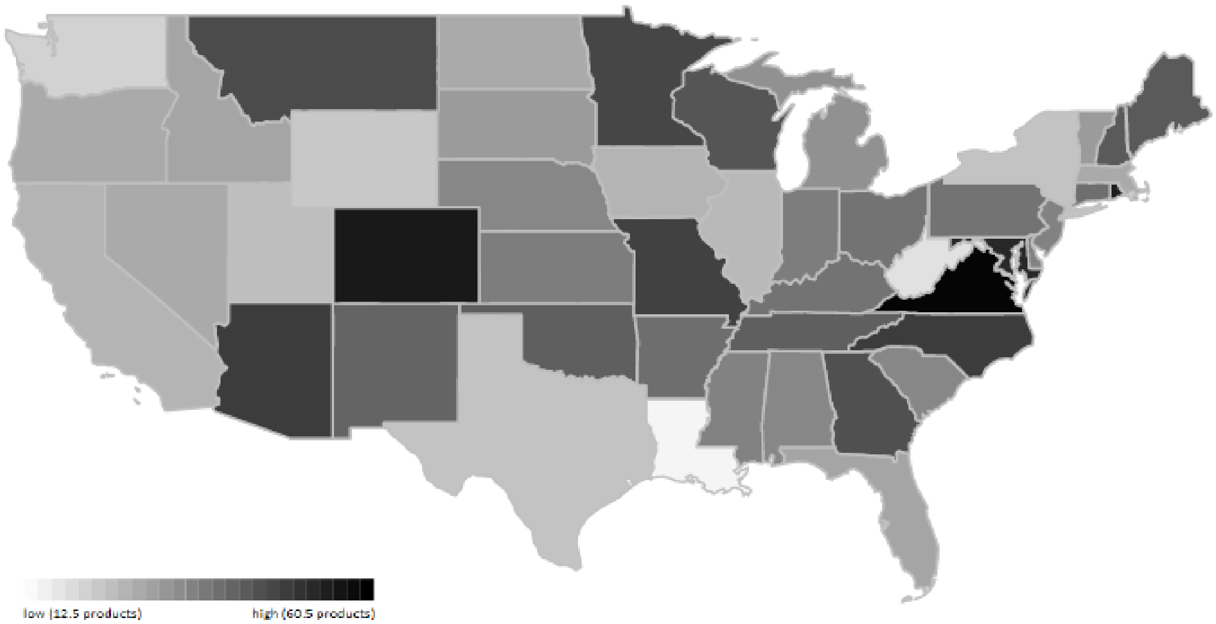

Figure 3. Step function for neighbor RCA index and the probability of developing a new industry

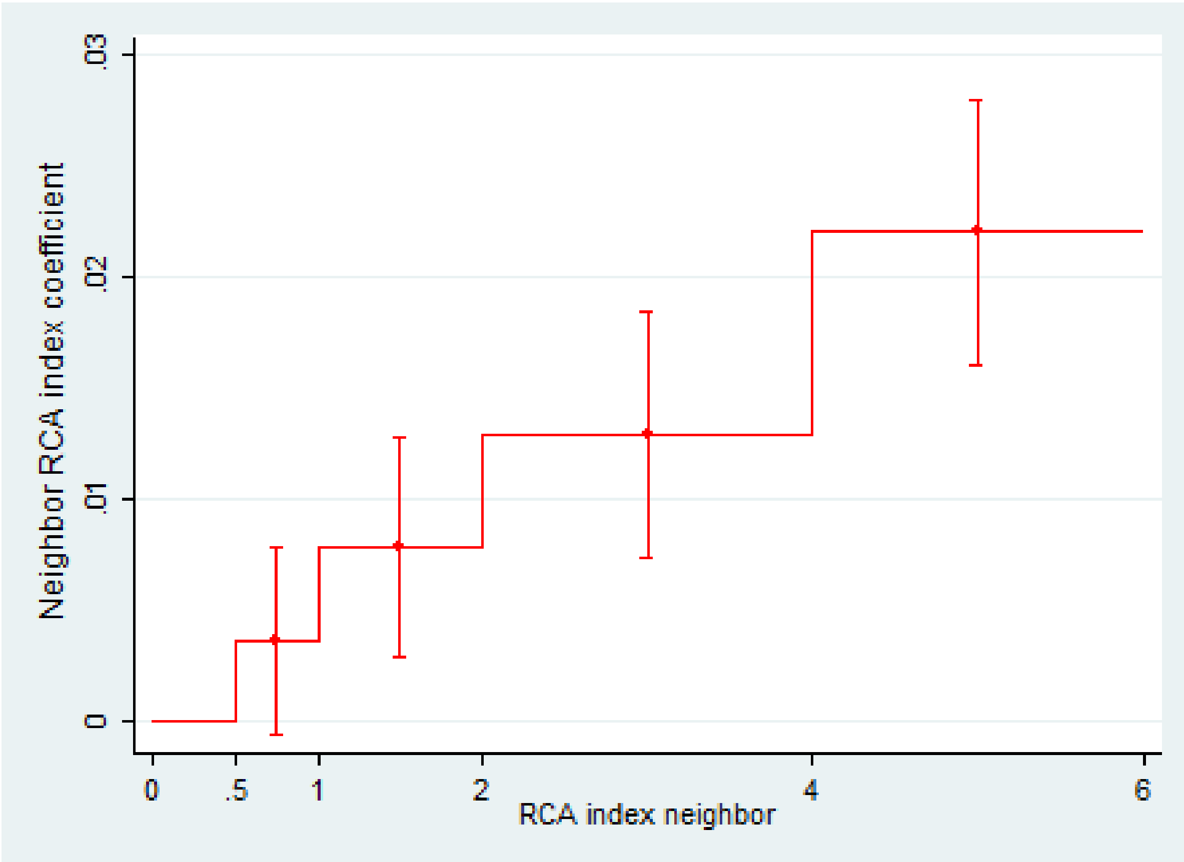

Figure 4. Step function for neighbor RCA index and the average annual growth rate in the RCA index

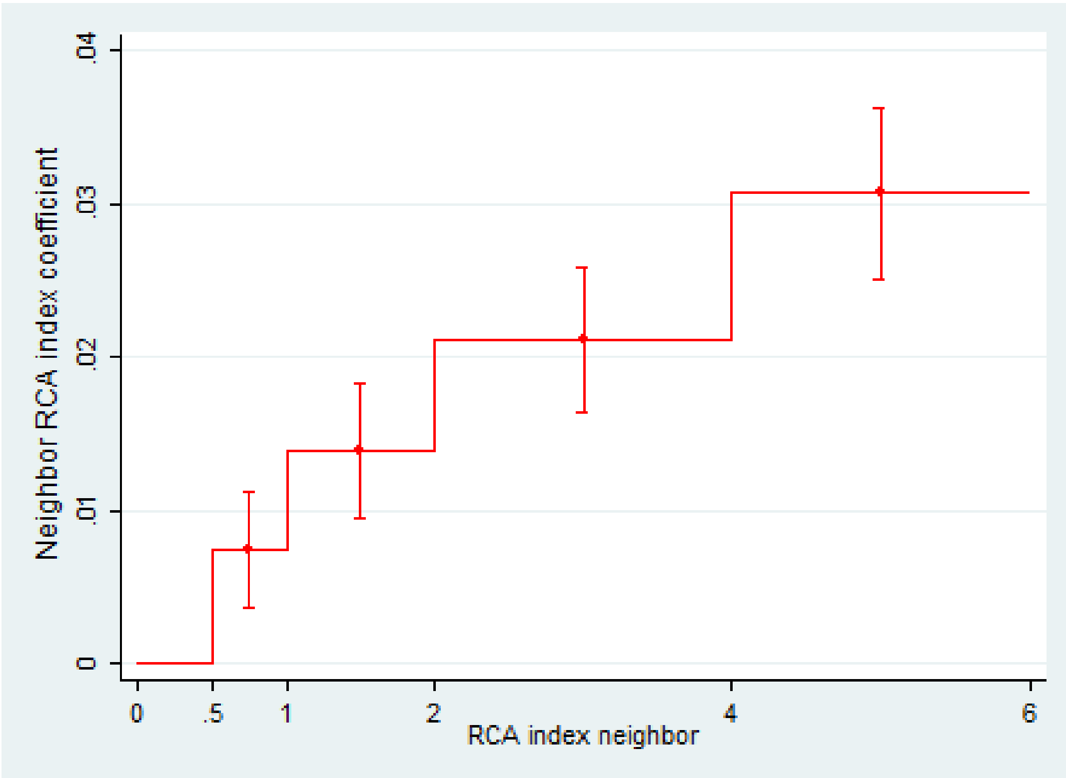

Figure 5. Similarities density functions: non-neighbors vs. neighbors

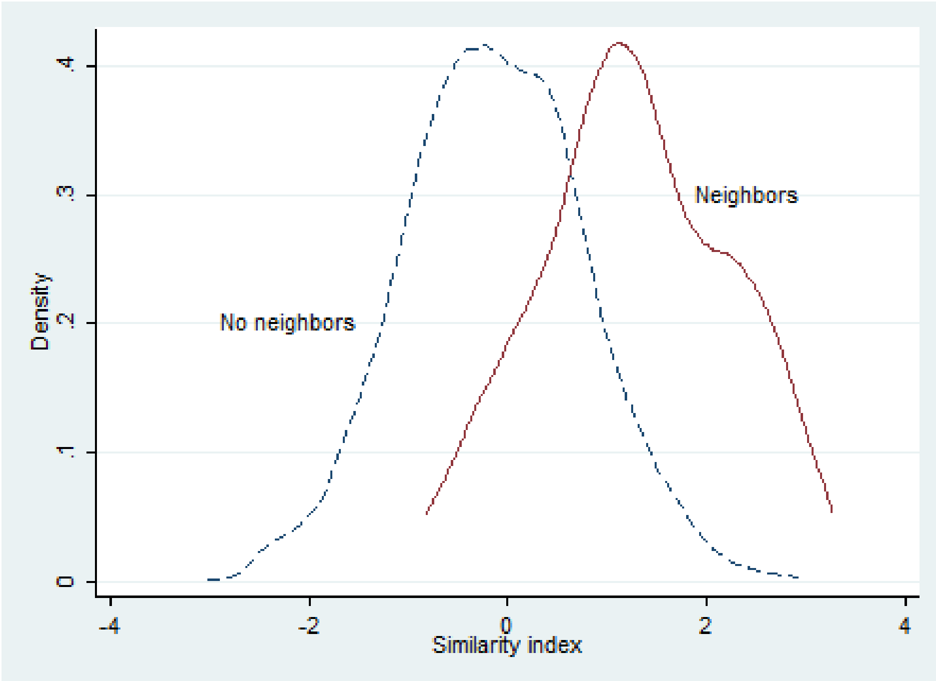

Figure 6. Step function for the distance coefficient in export similarity

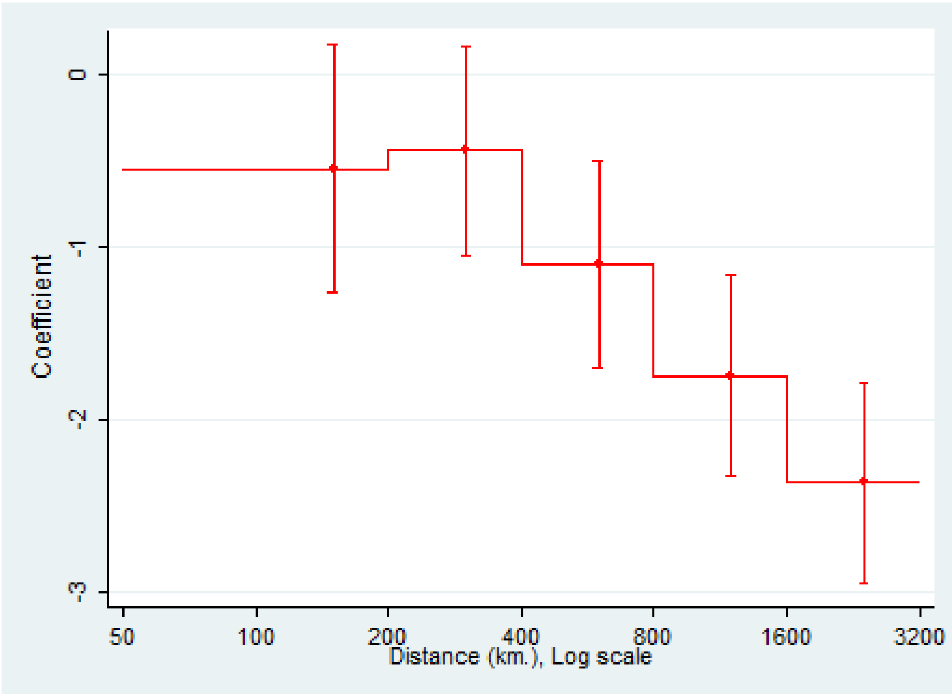

Table 1. Development of new industries. Baseline regressions

	(1)	(2)	(3)	(4)
Dependent variable	Jump	Jump	RCA growth	RCA growth
Initial RCA	<.5	<.5	<1	<1
RCA index neighbor(log)	0.004***	See Figure 3	0.006***	See Figure 4
	(0.000)		(0.001)	
RCA	0.252***	0.251***	-0.210***	-0.211***
	(0.010)	(0.010)	(0.005)	(0.005)
Density	0.373***	0.358***	1.153***	1.132***
	(0.131)	(0.133)	(0.086)	(0.088)
N	83598.000	83598.000	93867.000	93867.000
r2	0.091	0.091	0.130	0.130

Note: All independent variables are measured at the beginning of the period. The sample pools 5-year interval observations for the period 2002-2012. All regressions include year-specific state+neighbor state fixed effects and year-specific industry fixed effects. Standard errors clustered at the state+neighbor level in parentheses. ***, **, * statistically significant at 1%, 5% and 10% respectively.

Table 2. Development of new industries. Robustness analyses

	(1)	(2)	(3)	(4)	(5)	(6)
Analysis	Persistence	Alternative definition new industry	Random neighbors	Random neighbors	10-year interval	10-year interval
Dependent variable	Jump	Jump	Jump	RCA growth	Jump	RCA growth
RCA index neighbor(log)	0.003***	0.003***	-0.007***	-0.015***	0.005***	0.005***
	(0.000)	(0.000)	(0.003)	(0.003)	(0.001)	(0.001)
RCA	0.203***	0.260***	0.253***	-0.212***	0.268***	-0.283***
	(0.009)	(0.021)	(0.010)	(0.004)	(0.011)	(0.007)
Density	0.296***	0.225*	0.287***	1.094***	1.025***	2.265***
	(0.106)	(0.127)	(0.106)	(0.075)	(0.158)	(0.183)
N	81028	69294	90373	100776	42015	47190
r2	0.083	0.079	0.133	0.160	0.107	0.150

Note: All independent variables are measured at the beginning of the period. Except for columns (5) and (6), the sample pools 5-year interval observations for the period 2002-2012. All regressions include year-specific state+neighbor state fixed effects and year-specific industry fixed effects. Standard errors clustered at the state+neighbor level in parentheses. ***, **, * statistically significant at 1%, 5% and 10% respectively.

Table 3. Development of new industries. Total RCA and number of neighbors

	(1) Jump	(2) Jump	(3) Jump	(4) Jump	(5) Growth	(6) Growth	(7) Growth	(8) Growth
RCA sum neighbor (log)	0.004***	0.004***			0.007***	0.006***		
	(0.000)	(0.000)			(0.001)	(0.001)		
Nº of neighbors		0.000				-0.001		
		(0.001)				(0.001)		
RCA >1 sum neighbor (log)			0.006***	0.002			0.008***	0.003*
			(0.002)	(0.002)			(0.002)	(0.002)
Nº of neighbors with RCA>1				0.011***				0.011***
				(0.003)				(0.002)
RCA	0.252***	0.242***	0.312***	0.308***	-0.210***	-0.201***	-0.186***	-0.189***
	(0.013)	(0.015)	(0.021)	(0.021)	(0.006)	(0.007)	(0.007)	(0.007)
Density	0.351**	0.008	0.033	-0.007	1.150***	0.256***	0.850***	0.815***
	(0.154)	(0.026)	(0.220)	(0.222)	(0.114)	(0.030)	(0.158)	(0.159)
N	83598.000	83598.000	23154.000	23154.000	93867.000	93867.000	28592.000	28592.000
r2	0.087	0.081	0.143	0.143	0.126	0.117	0.167	0.168

Note: All independent variables are measured at the beginning of the period. The sample pools 5-year interval observations for the period 2002-2012. Regressions (1), (3), (4), (5), (7) and (8) include year specific state and product fixed effects. Regressions (2) and (7) include year-specific product fixed effects. Standard errors clustered at the state level in parentheses. ***, **, * statistically significant at 1%, 5% and 10% respectively

Table 4. Export similarity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Similarity indicator	Pearson	Pearson	Pearson	Pearson	Pearson	Pearson	Finger-Kreinin
Distance (log)	-0.763***	See Figure 6	Interval	Interval	Interval	Interval	Interval
	(0.036)						
Neighbor			0.434***	0.304***	0.194**	-0.093	0.064
			(0.099)	(0.109)	(0.097)	(0.101)	(0.089)
Nº of shared commuting zones				0.139**	0.199***	0.085	0.038
				(0.058)	(0.059)	(0.061)	(0.042)
GDPpc (diff)					-0.367**	-0.356**	-0.278*
					(0.164)	(0.158)	(0.151)
Physical Capital per worker(diff)					-0.391***	-0.402***	-0.189***
					(0.055)	(0.052)	(0.051)
Human capital per worker (diff)					-3.154***	-2.438***	0.157
					(0.613)	(0.588)	(0.463)
Land per worker (diff)					-0.175***	-0.085***	-0.073***
					(0.025)	(0.025)	(0.023)
COM					0.016**	-0.010	-0.016**
					(0.007)	(0.007)	(0.008)
Bilateral migration (log)						0.456***	0.191***
						(0.045)	(0.040)
N	2352.000	2352.000	2352.000	2352.000	2352.000	2352.000	2352.000
r2	0.612	0.626	0.634	0.636	0.681	0.709	0.736

Note: The sample pools year 2002 and year 2007 observations. All regressions include year-specific state fixed effects. Standard errors clustered at the state-pair level in parentheses. ***, **, * statistically significant at 1%, 5% and 10% respectively.

Table A1. Summary statistics for variables used in the similarity analysis (average years 2002 and 2007)

Variable	Mean	Standard deviation
Similarity index	0.198	0.098
Distance (km.)	1966.841	1137.531
Neighbor	0.093	0.290
Total bilateral trade (Ln)	6.590	3.151
Dif. GDP per capita (Ln)	0.233	0.286
Dif. Physical capital per worker (Ln)	1.409	1.241
Dif. Human capital per worker (Ln)	1.633	0.038
Dif. Land per worker (Ln)	0.052	1.339
Total bilateral migration	506907	98223

Source: see text.