Related variety and regional growth in Spain

Boschma, R.*, A. Minondo** and M. Navarro**/***

* Urban and Regional research center Utrecht (URU), Department of Economic Geography, Utrecht University, Utrecht

** Deusto University, Deusto Business School, San Sebastian

***Orkestra, The Basque Institute of Competitiveness, San Sebastian

Corresponding author:

Ron Boschma, Department of Economic Geography, Faculty of Geosciences, Utrecht University, Po Box 80.115, 3508 TC Utrecht, the Netherlands, email: r.boschma@geo.uu.nl

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

Abstract

This paper investigates whether related variety, among other types of spatial externalities, affected regional growth in Spain at the NUTS 3 level during the period 1995-2007. We found evidence that related variety matters for growth across regions, especially when using two new methods that measure revealed relatedness between industries. The first method is based on Porter's cluster classification while the second method uses the proximity index proposed by Hidalgo et al. Our analyses show that Spanish provinces with a wide range of related industries tend to show higher economic growth rates, once we control for other determinants of growth.

JEL Codes: D62, O18, R11

Keywords: related variety, Porter's cluster, product proximity, regional growth, Spain

1. Introduction

There is a growing body of literature stating that variety is beneficial for economic growth (Saviotti, 1996). Regional scientists have taken up this point and incorporated the virtues of variety in their regional growth models (Glaeser et al., 1992). Inspired by the seminal work of Jane Jacobs (1969), they argue that not necessarily cities *per se*, but cities with a diversified set of industries will be characterized by high economic growth, because local diversity will spark creativity, new ideas and innovations. However, some claim that this concept of Jacobs' externalities needs to be refined and specified more precisely (Porter, 2003), by making a distinction between related and unrelated variety (Frenken et al., 2007). In particular, they claim that an urban structure that consists of a wide set of technologically related industries is more conducive for urban growth. This is in line with an expanding literature that suggests that technological relatedness is a major asset for economic growth in regions, and for regional diversification in particular (Boschma and Frenken, 2011; Neffke et al., 2011).

Following Frenken et al. (2007), previous papers have used standard classifications of industries or products, such as the Standard Industrial Classification (SIC) or the Harmonized System (HS), to establish relatedness across products. There are two main shortcomings in using these classifications to establish relatedness across products or industries. On the one hand, these classifications use some priors to establish relatedness, such as similarities in product characteristics or in the production process, and they do not analyze whether those priors are relevant in practice. That is, they are ex-ante measures of relatedness. On the other hand, the use of some priors and not others does not enable to capture the whole range of possibilities by which products or industries can be related, like similarities in regulatory framework, complementarities in their use, the intensive use of a certain type of infrastructure, the use of advertisement to build trade marks, *etc*. To overcome these

shortcomings, the aim of this paper is to use more refined relatedness indicators to analyze the relationship between related variety and regional growth. The first indicator follows Porter's cluster classification and defines related industries on the basis of the geographical correlation of employment across traded industries (Porter, 2003). The second indicator rests on the products' proximity index developed by Hidalgo et al. (2007), which is based on the probability that a country develops comparative advantage in two products. In contrast to the conventional measures, these new indicators are ex-post indexes of relatedness, and are able to capture a larger range of factors affecting similarities across products and industries.

We test whether the new relatedness indicators are positively related with economic growth in 50 Spanish provinces during the period 1995-2007. The Spanish case is relevant for different reasons. First, the Spanish economy experienced an economic boom during the period of analysis, which came to a harsh stop with the burst of the housing bubble and the international financial crisis in the year 2008. If related variety also contributed to regional economic growth during the economic boom period, it can constitute an important strategy to resurrect value added and employment during the present recession in some Spanish regions. Moreover, the Spanish administrative structure, which deploys relevant decision capacity at the regional level, allows the fine-tuning of this strategy at the local level. Second, the availability of rich regional data in Spain (i.e. a large number of territorial units, a long time-period, and data to control for other determinants of growth) allows an accurate estimation of the related variety impact on regional growth.

The structure of the paper looks as follows. In Section 2, we explain the main theoretical ideas behind related variety, discuss a number of existing empirical studies on this topic, and present the advantages of the new relatedness indicators used in the paper. In Section 3 we

present the methodology followed in the empirical analysis and the dataset. Section 4 presents the main findings of the paper. Section 5 summarizes the main conclusions.

2. Spatial externalities, related variety and new relatedness indicators

2.1 spatial externalities and related variety

In the spatial externalities and regional growth literature, a key question is whether firms in cities learn principally from other local firms in the same sector, or from other local firms in a range of other sectors (Glaeser et al., 1992; Feldman and Audretsch, 1999). The former form of spatial externalities is known as localization economies, and dates back to the work of Marshall in the late nineteenth century (Marshall, 1890; Asheim and Gertler, 2005; Potter and Watts, 2011). According to Marshall, firms in specialized regions would benefit from local externalities due to the presence of specialized input suppliers, a local pool of specialized labor skills, and specialized knowledge concerning the secrets of the respective trade. The latter form of spatial externalities has been associated with Jacobs' externalities, and builds on the seminal work of Jane Jacobs developed in the 1960s (Jacobs, 1969; Lambooy, 1984; Becattini et al., 2003; Van Oort, 2004). A diversified economy would bring benefits to local firms because it would trigger and generate new thinking, new ideas and innovations.

Since Glaeser et al. (1992), many regional scientists have embarked on this type of research. Despite all their efforts, this literature has led to inconclusive results so far with respect to the question whether localization economies and Jacobs' externalities matter for urban and regional growth (Beaudry and Schiffauerova, 2009; De Groot et al., 2009). Among other reasons, this inconclusive finding may be caused by the potential misspecification of the

notion of Jacobs' externalities. For example, one can question whether knowledge will spill over across industries in diversified regions just because industries are each other's neighbors.

We believe that the literature on industry relatedness can shed light on this debate. In the 1980s and 1990s, there was focus on the degree of relatedness between technologies that are used in sectors, because this might affect the scope of knowledge spillovers and inter-industry learning (e.g. Rosenberg and Frischtak, 1983; Carlsson and Stankiewicz, 1991; Breshnahan and Trajtenberg 1995). As Nooteboom (2000) put it, knowledge is more likely to spill over across two industries when their cognitive distance is not too large, nor too small. That is, some degree of cognitive proximity between two sectors ensures effective communication and common understanding, and some degree of cognitive distance is needed to avoid cognitive lock-in.

When applying this concept of industry relatedness to the spatial externalities literature, one may expect that the extent to which the variety of technologies present in a region is related will 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) made the claim that the distinction between localization economies and Jacobs' externalities is therefore too simple, because it focuses too much on the industry itself. Instead, there is a need to emphasize the importance of externalities among related industries, which Porter linked to his concept of clusters as geographic concentrations of linked industries. As Porter (2003) put it, "clusters are important because of the externalities that connect the constituent industries, such as common technologies, skills, knowledge and purchased inputs" (p. 562). According to Porter, specialization in clusters of related industries, not in industries *per se*, should lead to better regional performance. And

next to having the benefits of clusters of related industries in a region, he argued that a range of overlapping clusters (caused by related industries that belong to more than one cluster) may be more beneficial for regional growth than having a diversity of clusters that are unrelated.

Frenken et al. (2007) incorporated this industry relatedness effect more explicitly in the spatial externalities and regional growth literature. They stated that the notion of Jacobs' externalities grasps two variety effects (i.e. related and unrelated variety) at the same time, and should therefore be disentangled. The related variety effect includes externalities that may come from a diversity of related industries in a region. The notion of regional related variety tries to capture a delicate balance between cognitive proximity and distance across sectors in a region that is needed for knowledge to spill over effectively between sectors. Thus, the more variety across related sectors in a region, the higher the number of technologically related sectors, and the more learning opportunities there are for local industries. This will result in more inter-sectoral knowledge spillovers, which enhance regional performance. This stands in contrast to localization economies in which regional specialization produces too much cognitive proximity between local firms (lock-in), while Jacobs' externalities *per se* may involve too much cognitive distance between local firms active in different industries.

The unrelated variety effect captures a portfolio-effect, which functions as a regional shock absorber (Essletzbichler, 2007). That is, when a region has a large number of unrelated industries, it may not be too vulnerable to sector-specific shocks. For instance, when an industry is affected by a sharp fall in demand, the workers that become redundant may find easily jobs in other local sectors that are unrelated and therefore will not be seriously damaged by this shock. Although both the related and unrelated variety effects are potential blessings for diversified regions, this latter effect is quite different from the related variety

effect, and therefore both effects normally associated with Jacobs' externalities should be empirically separated from each other.

Empirical studies have investigated the significance of related variety for regional growth in the Netherlands (Frenken et al., 2007), Great Britain (Bishop and Gripaios, 2010) and Italy (Boschma and Iammarino, 2009; Quatraro, 2010). These studies found quite strong empirical evidence for the importance of related variety for regional growth. This was, however, less true for the unrelated variety effect. Nevertheless, Frenken et al. (2007) found that Dutch regions with a high degree of unrelated variety performed better in terms of unemployment rates. As expected, unrelated variety dampened regional unemployment growth.

2.2 empirical measurement of industry relatedness

Following Frenken et al. (2007), previous studies have used standard classifications of industries or products, such as the Standard Industrial Classification (SIC) or the Harmonized System (HS), to establish relatedness across products. As explained in the introductory section, there are some limitations in using these classifications to establish relatedness across products. First, these classifications use some priors to establish relatedness, such as similarities in product characteristics or in the production process, and they do not analyze whether those priors are relevant in practice. Second, the use of some priors and not others does not enable to capture the whole range of possibilities by which products or industries can be related. To overcome these limitations, in this paper we establish relatedness across products based on ex-post indicators that are more agnostic on the sources of relatedness.

The first indicator follows Porter (1998) and uses a cluster classification to determine relatedness across products. Porter defines clusters as geographic concentrations of linked

industries that encompass producers, suppliers and providers of specialized services that generate (knowledge) externalities to local firms. In his US study, he used an ex-post indicator, the local correlation of employment across traded industries at the US state level, to define clusters of related industries. As Porter (2003) put it, “if computer hardware employment is nearly always associated geographically with software employment, this provides a strong indication of locational linkages” (p. 562). After applying this basic rule, Porter did basically two things to eliminate cases of spurious correlation: (1) he left out those cases in which no ‘logical’ externality was to be expected between two industries; (2) he excluded those cases that did not have any substantial input-output flows. Following this procedure, Porter identified 41 different clusters in the US, with an average of 29 industries each. The ex-post nature of Porter's relatedness indicator generates clusters that may include both manufacturing and service industries, a possibility which is ruled out by definition in the indicator that is based on standard classifications. As Porter (2003) describes himself, “clusters, then, represent a different way of dividing the economy that is embodied in conventional industrial classification systems that are based primarily on product type and similarities in production” (p. 563).

The second relatedness measure is based on the proximity indicator developed by Hidalgo et al. (2007). These authors argue that several dimensions may influence the degree of relatedness between two products: similarities in the combination of productive factors, the characteristics of the technology used in production, the use of a specific component, the features of the final customers, or the use of specific distribution channels. Due to the large number of factors that may determine relatedness between products, they also use an outcome, or ex-post, measure to calculate the degree of proximity between products. They argue that two products will be close to each other if countries tend to have revealed

comparative advantage in both products. Based on this idea, they calculate proximity (φ) between product *i* and product *j* at year *t* as:

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

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*.[1]

Based on this index and using network displaying techniques, Hidalgo et al. (2007) are able to draw a product space map, in which products are not evenly distributed: there are sections of the map with a high density of products, whereas other sections of the map are sparsely populated. Our argument is that these discontinuities in the product map are very important to determine learning opportunities. If a country specializes in products that are close to other products, learning opportunities will be larger. In contrast, if a country specializes in products that are far from each other, learning opportunities will be scant.

Hidalgo et al.'s proximity measure is different to Porter's measure in two respects. First, Hidalgo et al. are more agnostic than Porter on the sources of product relatedness and hence, they do not exclude any proximity, as in Porter's classification. Second, in Porter's classification an industry can belong to only one cluster. In contrast, with Hidalgo et al.'s proximity measure, a product can belong simultaneously to different relatedness-sets, a fact that will be explained more in detail in the next section.

[1] Neffke and Svensson-Henning (2008), following a similar idea to Hidalgo et al. (2007), developed a relatedness measure based on the frequency that two products are produced jointly at a plant level.

Due to their ex-post nature, Porter's cluster measure and Hidalgo et al.'s proximity indicator should better capture the degree of relatedness across products. Hence, we expect a stronger positive link between regional growth and related variety based on Porter's cluster and Hidalgo et al.'s proximity indicators, than between regional growth and related variety based on the conventional, ex-ante classification of relatedness. And due to the more comprehensive definition of relatedness and the possibility of overlapping relatedness sets, we expect related variety based on Hidalgo et al.'s proximity to show a stronger link with regional growth than related variety based on Porter's cluster measure.

3. Empirical framework

Section 3.1 explains the ways we identified the degree of relatedness between industries. We present three indicators of industry relatedness, which will be used as an input to assess the effect of related variety on regional growth. Section 3.2 explains how the variety measures have been calculated, while section 3.3 gives the main features of the model we estimated, and what kind of data sources we used.

3.1. Identification of related variety sets

As explained before, we use two novel indicators to measure relatedness across industries. To contrast the merits of these new indicators, we also calculate relatedness indicators using standardized product classifications. The first step to calculate a relatedness indicator is to determine with what other industries (products) an industry (product) is related to.

Using the Standard Industrial Classification (SIC), Frenken et al. (2007) assume that 5-digit industries are technologically related when they share the same 2-digit class. These industries

are perceived to show some degree of cognitive proximity, because these 5-digit sectors (e.g. sub-branches in chemicals) will share some technology and product characteristics in the same 2-digit class (e.g. chemicals). At the same time, these industries are considered to show some degree of cognitive distance, because these sectors differ at the 5-digit level. Then, the more variety there is at the 5-digit level within each 2-digit industry in a region, the more related variety and thus real learning opportunities are available in a region, and the more a region might benefit from externalities from such a wide set of different but related industries. The Frenken et al. study measures unrelated variety as the degree of variety of industries at the 1-digit level in a region. This is because at the 1-digit level, industries are unlikely to have much in common with respect to technology and product characteristics. Consequently, this indicator grasps the portfolio effect of variety explained earlier.

In contrast to Frenken et al. (2007), our database uses the Harmonized System 6-digit classification which obliges us to introduce some minor changes in the definition of related and unrelated varieties. In particular, we assume that 6-digit export sectors that share the same 2-digit class have some but not too much cognitive proximity and, hence, consider them related varieties. We also assume that 1-digit export sectors are not close in technology or other product characteristics and consider them as unrelated varieties.

To construct our second related variety measure for Spanish regions, we make use of Porter's cluster classification, as outlined in his 2003 study in the Table on page 563. Based on the correspondence table provided by the Institute for Strategy and Competitiveness of Harvard University, we established a link between our HS 6-digit industry classification and the 36 different Porter clusters (Harvard University).[2] In our Spanish case, each cluster has, on

[2] Table A1 in the appendix presents the list of clusters included in our analysis.

average, 345 different HS-6 digit products. An additional advantage of using this Porter classification based on US data rather than on Spanish data is that it may overcome the risk of endogeneity when identifying clusters in our study.

For relatedness measures based on Hidalgo et al. (2007) indicator, we calculated the proximity measures across products[3] using a sample of 102 countries for the years 2004 and 2005 from the UN Comtrade database, for the 1,244 products that compose the HS 4-digit 2002 Classification.[4] Figure 1 presents the histogram of the proximity measures calculated with our sample.[5] As shown in the figure, the proximity index follows a bi-modal distribution: there are high frequencies at zero and 0.15. The difficulty with this index is to determine what level of proximity is needed to consider two products as related. We have taken a conservative position and have considered that two products are related if their proximity is equal or above 0.25. There are 107,275 product-pairs (14% of all product-pair combinations) that meet this criterion. As mentioned in the previous section, in contrast to other indicators, when we use proximity, a product may belong to more than one relatedness set. This overlapping possibility makes the proximity-based indicator richer than the conventional classification or the cluster-based index of relatedness.

- Figure 1 here -

[3] To calculate conditional probabilities, first, we determine whether countries have revealed comparative advantage (a la Balassa) in product *i*. Second, we calculate the probability of having comparative advantage in product *i*, 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* given a country has comparative advantage in product *j*, dividing the joint probability of having comparative advantage in product *i* and product *j* by the probability of having comparative advantage in product *j*. Finally, we selected the minimum of the pairs of conditional probabilities.

[4] The Comtrade database allows the calculation of proximity indexes at the 6-digit HS level. However, there have been some changes in the HS classification at this disaggregated level. As we had to match these data with the Spanish regional data, we calculated the proximity measures at the 4-digit level, to minimize the loss of data due to changes in classification.

[5] The proximity matrix encompasses 773,143 indexes: (1,244 products x 1,243 products)/2.

3.2. Variety indexes

We made use of entropy measures to calculate the different variety indexes. To calculate the related variety index, firstly, we grouped HS products into related variety sets: S_r. As explained above, in the conventional measure, a related variety set is composed of those 6-digit HS products that belong to the same 2-digit HS products' class. In the Porter measure, a related variety set is composed of those 6-digit HS products that belong to the same cluster. Finally, in the proximity indicator, we define a related variety set for each product, which is composed of the rest of HS 4-digit products that have, at least, a proximity equal to 0.25. We calculated the share of HS *i* product (6-digit or 4-digit) in total regional exports (p_i) and the share of each related variety set in total regional exports (P_r). With these shares we calculate the entropy within the related variety set (H_r) as follows:

$$H_r = \sum_{i \in Sr} \frac{p_i}{P_r} \log_2 \left(\frac{1}{p_i / P_r} \right). \qquad (2)$$

Related variety is calculated as the exports-weighted entropy in each related variety set.

$$RELATED \quad VARIETY \quad = \sum_{r=1}^{R} P_r H_r \qquad (3)$$

Unrelated variety is calculated in a similar way. Now, we have to define the unrelated variety sets for each measure. For the conventional and the Porter measure, there is only one unrelated variety set. In the first case, the unrelated variety set is composed of each 1-digit HS industry, while in the second case, it is composed of each cluster. For these two measures, the unrelated variety index is calculated as follows:

$$UNRELATED\ VARIETY = \sum_{j=1}^{N} P_j \log_2 \left(\frac{1}{P_j} \right) \qquad (4)$$

where P_j denotes the share of each 1-digit sector or cluster in total exports.

For the proximity measure, there is one unrelated variety set for each product, which is composed of the rest of products whose proximity to the analyzed product is below 0.25. As in the related variety set, we calculate entropy using equation (2), and then calculate unrelated variety as the weighted sum of entropy at each unrelated variety set using equation (3).

To analyze whether disentangled variety measures better identify learning opportunities than conventional variety measures, we also calculated a Jacobs' externalities or variety measure. The variety index is calculated as follows:

$$VARIETY = \sum_{i=1}^{N} p_i \log_2 \left(\frac{1}{p_i} \right) \qquad (5)$$

where p_i stands for the share of 6-digit HS product i in total regional exports.

3.3. Empirical model and data

To analyze the relationship between related variety and growth we estimate the following equation:

$$va_{growth_i} = \beta_0 + \beta_1 RV_i + \beta_2 UV_i + \beta_3 Urb_i + \beta' X_i' \qquad (6)$$

where va_growth_i denotes the value added growth in region i, RV_i is related variety and UV_i unrelated variety. Related and unrelated variety are measured at the beginning of the period. Our expectation is that regions with a productive structure characterized by related industries will have higher value-added growth rates than other regions. In addition to related and unrelated variety, we also control for the effect of urbanization economies, measured by population density, on growth. Finally, X' is a vector that includes other factors that may influence regional growth, such as human capital and labor-productivity.

The geographic unit used in our analysis are Spanish provinces, classified as NUTS-3 in Eurostat's regional classification. As shown in Map 1, Spain is divided in 50 provinces and two autonomous community cities.[6] The period of analysis is 1995-2007.[7] We divide the period of analysis in four-year intervals.[8] Growth is measured as the average annual value-added growth in the 4-year interval.

- Map 1 here -

Data on value added and population of Spanish provinces come from the Spanish Statistical Institute's (INE) Regional Economic Accounts database. Human capital data are obtained from the Instituto Valenciano de Investigaciones Económicas (Ivie) database. Human capital is proxied by the percentage of occupied population that has upper-secondary or tertiary

[6] Due to the availability of fewer and less reliable data, we have excluded the two autonomous cities from the sample (Ceuta and Melilla). With respect to the remaining fifty provinces, seven of them are uni-provincial autonomous communities.
[7] Although the Spanish Statistical Institute (INE) provides data before 1995, these data are less reliable.
[8] The use of alternative intervals, such as 3-year or 2-year intervals, did not alter the results of the econometric analyses.

studies. To calculate relatedness indicators, we use Spanish provinces exports and imports at the Harmonized System 6-digit level. These were obtained from the Spanish Dirección General de Aduanas - Agencia Tributaria database.

The use of international trade data to construct the variety indicators has some limitations (see also Boschma and Iammarino, 2009). Obviously, not all industries are export sectors, so the export profile of a region may not fully reflect the industrial composition of a region. In our measurements, there will be some bias toward manufacturing activities, due to, among others, the relatively low tradability of most service industries. Having said that, in manufacturing industries, knowledge complementarities between sectors can be approximated by export structures of regions, since industries that are most open to international competition are also those that contribute most to new knowledge, innovation and economic growth (e.g., Dosi, 1988; Fagerberg, 1988). As exporting occurs in almost all manufacturing industries, and export industries are among the strongest in a region, we expect the effects of related and unrelated variety to matter most among export sectors.

4. Empirical findings

Table 1 presents the results of estimating equation (6). First, we estimate the model with Jacobs' externalities (variety) as independent variable. Then, this diversity index is divided into a related variety and unrelated variety variable. We estimate the model for our three relatedness indicators, that is, the conventional measure, Porter's cluster classification, and Hidalgo et al.'s proximity indicator.[9] The independent variables are measured at the beginning of each four-year-interval. We pool all the observations and estimate an OLS model.

[9] Table A2 in the appendix presents a correlation matrix for the independent variables used in the empirical analyses.

- Table 1 here -

As shown in Table 1, both population density and variety have a positive effect on regional growth in Spain. In other words, more urbanized and more diversified regions in Spain perform better. With respect to the control variables, we find that provinces with an initial high level of labor productivity show higher value-added growth rates, while the opposite is true for initial levels of human capital.[10] Furthermore, we can observe that coefficients are not biased due to spatial autocorrelation (Moran I's is never statistically significant).[11]

The findings presented in Table 1 confirm that it is essential to split the effects of Jacob's externalities into two components, that is, related and unrelated variety. As expected, we find that related variety has a positive and statistically significant effect on regional growth for all our three relatedness indicators. This result is very robust and in line with previous studies, such as Frenken et al. (2007) and Boschma and Iammarino (2009), that found a positive relationship between related variety and growth in value-added and employment at the regional level. What we also observe in Table 1 is that the value of the related variety coefficient rises when we use the new relatedness indicators. When additional control factors are not introduced in the model, the related-variety coefficient is 0.00271 for the conventional measure, 0.00423 for the clusters-based measure and 0.00550 for the proximity-based measure. When additional controls are introduced in the model, the coefficients become 0.00376 for conventional, 0.00590 for clusters and 0.00526 for proximity. These results confirm our expectation that ex-post relatedness indicators better capture the economic effects of relatedness across industries, as witnessed by a stronger relationship between related variety and regional growth. We also expected the proximity-based measure to perform better

[10] We also introduced an additional variable developed in Minondo (2010) to control for the sophistication level of provinces' production. This variable did not alter the results.
[11] We also estimated a fixed-effects model and found that province-level effects were not statistically significant.

than the cluster-based measure. This expectation is confirmed for the regressions without control variables, but it is not confirmed for the regressions that include control variables.

The findings in Table 1 show that, overall, unrelated variety has no positive effect on regional growth, in contrast to related variety. With respect to the conventional indicator, unrelated variety has a positive coefficient, but it is not statistically significant. When clusters are used to establish relatedness across industries, unrelated variety has a negative coefficient, and becomes statistically significant when we include control variables. Finally, when we use proximity to build the relatedness indicator, unrelated variety is negative, but it is not statistically significant. In any case, we should treat these results with care, as a four-year period might be too short to allow portfolio effects to smoothen industry-specific shocks.

Boschma and Iammarino (2009) argue that regions can also learn from the varieties imported from other countries. Following these authors, we constructed variety, related variety and unrelated variety measures using data on Spanish provinces imports from other countries, and found that these variables had no effect on value-added growth.[12]

To test the robustness of our results, following Frenken et al. (2007) and Boschma and Iammarino (2009), we also estimated a model using regional employment growth, instead of regional value-added growth, as dependent variable. As an additional control variable, we included the level of employment in the region at the beginning of the period. As shown in the Table A3 in the Appendix, the coefficient for related variety becomes stronger and statistically significant when we move from the conventional measure of relatedness to the cluster-based and proximity measures. Moreover, the related variety coefficient based on

[12] Results are not reported, but can be obtained upon request.

proximity is statistically significant in both the regression with controls and in the regression without controls, whereas the related variety coefficient based on clusters is only statistically significant when controls are introduced in the regression.[13]

In sum, the use of alternative measures to determine relatedness between industries lead to a stronger effect of related variety on regional growth. These results tend to confirm that clusters and product proximity better identify the relatedness across industries than the conventional measures based on standard industry or product classification.

5. Conclusions

This study has investigated the importance of related variety for regional growth in Spain using new relatedness indexes. In the conventional manner, one makes use of standard industrial or product classifications, and industries are defined as related when they share the same digit class at a more aggregated level. However, this measure of relatedness between industries is not unproblematic. One reason is that conventional measures establish relatedness ex-ante, and do not test whether their criteria is relevant in practice (Neffke and Svensson-Henning, 2008). Another reason is that relatedness measures based on standard classifications do not capture the whole range of factors that may influence relatedness among products. This problem seems especially acute with respect to technology relatedness (Breschi et al., 2003).

In our study, first, we used an alternative measure that was proposed by Porter, in which industries are classified by means of their geographic correlation. We also calculated an additional related variety measure based on the product proximity concept developed recently

[13] Following Boschma and Iammarino (2009), we also estimated the model with labor-productivity growth as dependent variable. In this case, related variety was never statistically significant.

by Hidalgo et al. (2007). In the econometric estimations, we find strong and robust evidence that related variety is positively related with regional value-added growth and regional employment growth in Spain. And, as expected, the effect of related variety on value-added growth and employment growth at the regional level becomes stronger when we used the related variety measures based on cluster and proximity indicators.

Our results show that Spanish regions characterized by a productive structure of related industries are able to grow faster than other regions. This result is very relevant in the present context, in which Spanish regions need to resurrect value-added and employment based on productive models that are less dependent on construction.

These outcomes also call for further research. Among others, there is a need to open the black box of relatedness between industries as a source of agglomeration externalities by exploring the channels through which spillovers occur. Labor mobility is one such channel, because it is a key mechanism through which knowledge diffuses within regions. There is indeed some evidence that labor flows between related industries are of particular importance (Boschma et al., 2009), but more systematic evidence is needed. In addition, in order to measure more accurately the importance of relatedness on regional development, one should investigate its effects on the growth of industries and firms at the regional level. This may be more relevant for some industries, as compared to other industries (and depending on the stages of their life cycles), and more relevant for some firms than for other firms (e.g. small versus large firms). This is a promising research area that still needs to be explored.

References

Asheim B, Gertler M (2005) The geography of innovation. Regional innovation systems. In: Fagerberg J, Mowery D, Nelson R (eds), *The Oxford Handbook of Innovation*. Oxford University Press, Oxford, pp. 291-317.

Beaudry C, Schiffauerova A (2009) Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy* 38 (2): 318-337.

Becattini G, Bellandi M, Ottati G dei, Sforzi F (2003) *From industrial districts to local development. An itinerary of research*. Edward Elgar, Cheltenham.

Bishop P, Gripaios P (2010) Spatial Externalities, Relatedness and Sector Employment Growth in Great Britain. *Regional Studies*, 44 (4): 443-454.

Boschma R, Frenken K (2011) Technological relatedness and regional branching. In: Bathelt H, Feldman MP, Kogler DF (eds), *Dynamic Geographies of Knowledge Creation and Innovation*. Routledge, Taylor and Francis (forthcoming).

Boschma R, Iammarino S (2009) Related Variety, Trade Linkages, and Regional Growth in Italy. *Economic Geography* 85 (3): 289-311.

Boschma R, Eriksson R, Lindgren U (2009) How does labor mobility affect the performance of plants? The importance of relatedness and geographical proximity. *Journal of Economic Geography* 9 (2): 169-190.

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

Breshnahan TF, Trajtenberg M (1995) General purpose technologies. Engines of growth? *Journal of Econometrics* 65 (1): 83-109.

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

De Groot HLF, Poot J, Smit MJ (2009) Agglomeration externalities, innovation and regional growth: theoretical perspectives and meta-analysis. In: Capello R, Nijkamp P (eds) *Handbook of Regional Growth and Development Theories.* Edward Elgar, Northampton MA, pp. 256-281.

Dosi G (1988), Technical change and international trade. In: Dosi G, Freeman C, Nelson R, Silverberg G, Soete L (eds) *Technical change and economic theory*. Pinter, London.

Essletzbichler J (2007) Diversity, Stability and Regional Growth in the United States, 1975-2002. In: Frenken K (ed) *Applied Evolutionary Economics and Economic Geography*. Edward Edgar, Cheltenham.

Fagerberg J (1988) International competitiveness. *Economic Journal* 98: 355–74.

Feldman MP, Audretsch DB (1999) Innovation in cities. Science-based diversity, specialization and localized competition. *European Economic Review* 43: 409-429.

Frenken K, van Oort FG, Verburg T (2007) Related variety, unrelated variety and regional economic growth. *Regional Studies* 41 (5): 685–97.

Glaeser EL, Kallal HD, Schinkmann JA, Shleifer A (1992) Growth in cities. *Journal of Political Economy* 100:1126–52.

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

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

Lambooy JG (1984) The regional ecology of technological change. In: Lambooy JG (ed) *New Spatial Dynamics and Economic Crisis*. Finnpublishers, Tampere, pp. 63-77.

Marshall A (1890) *Principles of Economics*. Macmillan, London.

Minondo A. (2010) Exports' productivity and growth across Spanish regions. *Regional Studies* 44 (5): 568-577.

Neffke, F.M.H. and Svensson-Henning, M. (2008) Revealed relatedness. Mapping industry space, Department of Economic Geography, Utrecht.

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

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

Porter ME (1998) Clusters and the new economics of competition. *Harvard Business Review* 76 (6): 77-90.

Porter ME (2003) The Economic Performance of Regions. *Regional Studies* 37 (6&7): 549-578.

Potter A, Watts DG (2010) Evolutionary agglomeration theory. Increasing returns, diminishing returns, and the industry life cycle. *Journal of Economic Geography*, 11(3): 417-455.

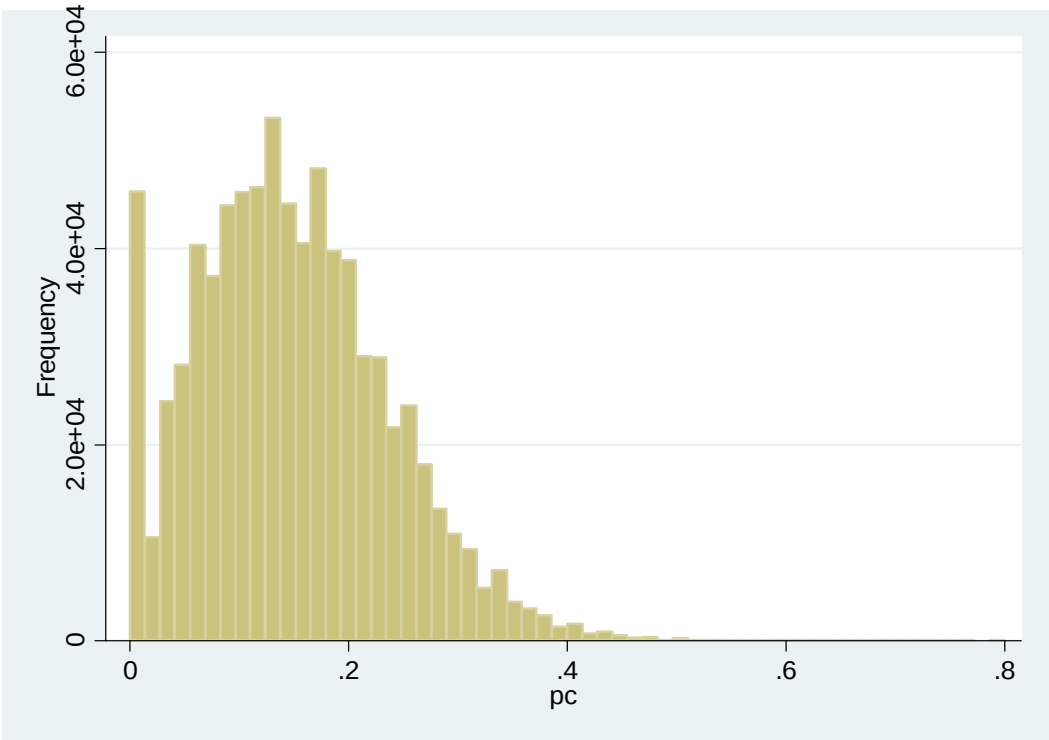
Quatraro, F. (2010) Knowledge coherence, variety and economic growth. Manufacturing evidence from Italian regions, *Research Policy*, doi:10.1016/j.respol.2010.09.005

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

Saviotti PP (1996) *Technological evolution, variety and the economy*. Edward Elgar, Cheltenham UK, Brookfield Vt.

Van Oort FG (2004) *Urban Growth and Innovation. Spatially Bounded Externalities in the Netherlands*. Ashgate, Aldershort.

Figure 1. Histogram of Hidalgo et al.'s product proximity indicator



Source: authors' calculations based on Comtrade database.

Map 1. The Spanish provinces

Table 1. Related variety and value-added growth across Spanish provinces, 1995-2007 (4-year intervals)

	Reg (1.1)	Reg. (1.2)	Reg. (2.1a)	Reg. (2.2a)	Reg. (2.1b)	Reg. (2.2b)	Reg. (2.1c)	Reg. (2.2c)
Dependent variable	value-added growth	value-added growth	value-added growth	value-added growth	value-added growth	value-added growth	value-added growth	value-added growth
Relatedness indicator			conventional	conventional	clusters	clusters	proximity	proximity
Population density (Log)	0.00204**	0.00192*	0.00152	0.00125	0.00253**	0.00224**	0.00183**	0.00178*
	(0.000824)	(0.000956)	(0.00104)	(0.00110)	(0.00101)	(0.00109)	(0.000825)	(0.000931)
Variety (Jacobs)	0.00127*	0.00160**						
	(0.000683)	(0.000674)						
Related variety			0.00271*	0.00376**	0.00423**	0.00590***	0.00550**	0.00526**
			(0.00147)	(0.00185)	(0.00173)	(0.00183)	(0.00215)	(0.00219)
Unrelated variety			0.000136	9.80e-06	-1.71e-05	-2.22e-05*	-0.00150	-0.000923
			(0.00229)	(0.00233)	(1.32e-05)	(1.28e-05)	(0.00155)	(0.00160)
Labor-productivity (Log)		0.0203		0.0211		0.0326**		0.0176
		(0.0133)		(0.0135)		(0.0137)		(0.0126)
Human capital (Log)		-0.0131**		-0.0159*		-0.0191**		-0.0114*
		(0.00650)		(0.00808)		(0.00797)		(0.00666)
Constant	0.0160***	-0.150	0.0196**	-0.144	0.0133***	-0.260**	0.0195***	-0.125
	(0.00474)	(0.128)	(0.00740)	(0.129)	(0.00454)	(0.126)	(0.00595)	(0.121)
Observations	150	150	150	150	150	150	150	150
R-squared	0.105	0.126	0.110	0.186	0.133	0.176	0.146	0.161
Model	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS
Moran's I p-value	0.597	0.532	0.582	0.527	0.575	0.528	0.628	0.560

Note: Province-clustered standard errors in parentheses. All regressions include period dummies. *** p<0.01, ** p<0.05, * p<0.1

Table A1. Porter's clusters

Cluster code	Cluster Name
1	Aerospace engines
2	Aerospace vehicles and defense
3	Agricultural products
4	Analytical instruments
5	Apparel
6	Automotive
7	Biopharmaceuticals
8	Building fixtures, equipment and services
9	Chemical products
10	Communications equipment
11	Construction materials
12	Reproduction equipment
13	Fishing and fishing products
14	Footwear
15	Forest products
16	Furniture
17	Heavy machinery
18	Information technology
19	Jewellery and precious metals
20	Leather products
21	Lighting and electrical equipment
22	Marine equipment
23	Medical devices
24	Metal manufacturing
25	Motor driven products
26	Oil and gas products and services
27	Plastics
28	Power generation and transmission
29	Prefabricated enclosures
30	Processed food
31	Production technology
32	Publishing and printing
33	Sporting, recreational and children’s goods
34	Textiles
35	Tobacco
36	Coal

Table A2. Correlation matrix for independent variables

	Population density	Variety	Related variety conventional.	Related variety clusters	Related variety proximity.	Unrelated variety conventional.	Unrelated variety clusters	Unrelated variety proximity.	Labor productivity	Human capital
Population density	1									
Variety	0.5561	1								
Related variety conventional	0.6667	0.84	1							
Related variety clusters	0.5785	0.8832	0.9145	1						
Related variety proximity	0.4751	0.8181	0.6836	0.7686	1					
Unrelated variety conventional	0.0685	0.6435	0.1596	0.334	0.5099	1				
Unrelated variety clusters	0.7068	0.8408	0.8271	0.7754	0.7233	0.3834	1			
Unrelated variety proximity	0.4402	0.8876	0.8329	0.8374	0.7351	0.4442	0.82	1		
Labor-productivity	0.3911	0.2602	0.3492	0.1844	0.2483	-0.0866	0.3443	0.2296	1	
Human capital	0.3496	0.3973	0.4683	0.3916	0.3235	0.0011	0.4291	0.4062	0.5805	1

Table A3. Related variety and employment growth across Spanish provinces, 1995-2007 (4-year intervals)

	Reg. (1.1)	Reg. (1.2)	Reg. (2.1a)	Reg. (2.2a)	Reg. (2.1b)	Reg. (2.2b)	Reg. (2.1c)	Reg. (2.2c)
Dependent variable	employment growth	employment growth	employment growth	employment growth	employment growth	employment growth	employment growth	employment growth
Relatedness indicator			conventional	conventional	clusters	clusters	proximity	proximity
Population density (Log)	0.00373**	-0.000909	0.00321	-0.00129	0.00464**	-0.000556	0.00339**	-0.00185
	(0.00174)	(0.00351)	(0.00219)	(0.00351)	(0.00221)	(0.00332)	(0.00156)	(0.00392)
Variety (Jacobs)	0.00128	0.00168					0.00854**	0.00723**
	(0.00138)	(0.00130)					(0.00425)	(0.00356)
Related variety			0.00280	0.00337	0.00362	0.00931***	0.00854**	0.00723**
			(0.00287)	(0.00326)	(0.00318)	(0.00337)	(0.00425)	(0.00356)
Unrelated variety			-0.000869	0.000440	-1.81e-05	-5e-05**	-0.00349	-0.00247
			(0.00293)	(0.00360)	(1.87e-05)	(1.94e-05)	(0.00245)	(0.00259)
Labor-productivity (Log)		0.132***		0.133***		0.155***		0.129***
		(0.0313)		(0.0317)		(0.0321)		(0.0298)
Employment (Log.)		0.00292		0.00281		0.00425		0.00404
		(0.00414)		(0.00436)		(0.00419)		(0.00455)
Human capital (Log)		-0.0409***		-0.0428***		-0.0508***		-0.0372***
		(0.0122)		(0.0148)		(0.0145)		(0.0134)
Constant	0.0118	-1.216***	0.0196**	-1.209***	0.00891	-1.427***	0.0190**	-1.189***
	(0.00880)	(0.307)	(0.00740)	(0.309)	(0.0104)	(0.304)	(0.00928)	(0.290)
Observations	150	150	150	150	150	150	150	150
R-squared	0.074	0.303	0.110	0.304	0.079	0.359	0.111	0.326
Model	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS
Moran's I p-value	0.411	0.410	0.582	0.402	0.380	0.419	0.420	0.435

Note: Province-clustered standard errors in parentheses. All regressions include period dummies. *** p<0.01, ** p<0.05, * p<0.1.