EXTENDED ABSTRACT

BREAKTHROUGH PATENTS AND TECHNOLOGICAL RELATEDNESS

Ron Boschma

Department of Economic Geography, Utrecht University & UiS Business School, University of Stavanger

ron.boschma@circle.lu.se

Ernest Miguélez

GREThA, University of Bordeaux & AQR-IREA, University of Barcelona

ernest.miguelez@u-bordeaux.fr

Rosina Moreno

AQR-IREA, University of Barcelona

rmoreno@ub.edu (corresponding author)

Diego Ocampo

AQR-IREA, University of Barcelona

diegob_ocampo@ub.edu

Keywords

Relatedness, Patents, Co-ocurrence

BREAKTHROUGH PATENTS AND TECHNOLOGICAL RELATEDNESS

Main objectives and hypotheses

In this paper we focus on breakthrough technological knowledge in European regions and analyze whether the generation of breakthrough knowledge is shaped by the existing knowledge base of the regions. The main hypothesis in the paper is that super patents in a region (those receiving more citations) combine technological classes that are unrelated and available in the region. A region's characteristics drive the process of generation of breakthrough patents but also the overall set of unrelated technologies that are present in the region. In this sense, our first objective tries to study whether super patents in a region actually combine technology classes that are unrelated (defined through co-occurrence analysis) and present in the region concerned. Secondly, we pursue to analyse whether super patents in a region actually connect to technology classes that are unrelated and present in the region through the patents they cite.

Empirical model

Our endogenous variable proxies for the existence of super patents and is regressed as a function of an index of unrelatedness, while taking control of regional and technological characteristics. All the specifications are estimated at the region-technology level, using the following specification:

$$SUPERPAT_{r,i,t} = \beta_0 + \beta_1 UNRELATEDNESS_{r,i,t-1} + \beta_2 GDP_{r,t-1} + \beta_3 DENS_{i,t-1} + \beta_4 DENS^2_{i,t-1} + \emptyset_r + \delta_i + \gamma_t + \varepsilon_{r,i,t}$$

where r refers to region, i refers to technology and t to time period and a regression residual is included. Our panel consists of 265 NUTS2 regions and 634 technologies over the period 1981-2010. We average the data over non-overlapping five-year periods, denoted by *t*. To dampen potential endogeneity issues, all the independent variables are lagged by one period, denoted by *t-1*.

Database and variables

The database we use is the REGPAT September 2015 edition. In order to construct a measure of unrelatedness we must first compute the degree of relatedness between technologies. For this purpose we use the co-occurrence of IPC classification codes in patent documents. Specifically, we take the second level of disaggregation of the IPC (four digits), getting a total number of 634 technological classes for the first five year period (1981-1985) and 634 in the most recent time interval (2006-2010). This way we can have a measure of knowledge relatedness for each pair of technologies (Breschi et al., 2003). The underlying idea of this indicator is that when two different technological classes appear together in a patent document, it implies that they are knowledge-related. Since the co-occurrence we are interested in is the technological co-location, as a measure of co-occurrence we employ the count of times any two technologies appear together in any patent. To control for the fact that this co-occurrence can be caused by the number of times any two technologies appear at the same time in a patent randomly, we normalized this measure using the association measure presented in Eck et al. 2009.

To translate the measure of co-occurrance to the regional European NUTS2 level we first construct the revealed comparative advantage index of each region, RCA (Hidalgo et al. 2007, Boschma et al. 2013). We combine it with the co-occurrence matrix to know the degree of relatedness that every technology of every region has with the technologies in which such a region presents RCA through a density index in the way of Boschma et al. (2014).

To capture the number of super patents in each technology and region (our endogenous variable) we define them as those with more forward citations, under the assumption that if a patent received many citations, such patent should contain influential knowledge for the creation of new ideas. In exact terms we define a super patent as the one that, within the distribution of number of forward citations, is located in the upper 95% of the distribution. To control for the fact that in some sectors patenting and citing could be more dynamic, we measure thresholds for each of five technological sectors as defined by Schmoch (2008). At the same time, to take into account that according to the year of release of a patent, it is subject to different period lengths to be cited, so that older patents have higher chance to be cited. We also classified the patents according to their priority year to calculate the 95% thresholds that allow us to assign the super patents.

For the construction of the dependent variables we constructed time windows of five years starting in 1981 until 2010. Inside each time period we calculated the number of super patents for the region-technology pairs.

Very preliminary results

As preliminary results we performed a pooled OLS estimation including regional, period and technology fixed effects. As dependent variables we use two alternative measures for the creation of super patents. The first one is the simple count of super patents in period t. The second measure is a binary variable that is equal to 1 when in a certain region and in a certain technology, at least one super patent appears; in other words, if in period t-1 there were no super patents, in period t at least one super patent appeared. In this case, the model is a linear probability model. The key variable is our measure of unrelatedeness lagged one period. As controls we introduce one period lagged values of the regional GDP per capita, the population density and its square term. To include the controls at the regional level, we took the five year average for each period.

In models 1 to 4, where the dependent variable is the number of super patents in a region, in a certain technology, we can see that the level of unrelatedness between the technologies in which the region has revealed comparative advantage has a positive and statistically significant coefficient. This result maintains even when controlling for the regional, time and technology fixed effects. With respect to the controls, as expected, the GDP per capita has a positive and significant coefficient, whereas for the population density and its square term that proxy for agglomeration externalities, we observe a significant and positive effect but with diminishing returns.

In models 5 to 8, where the dependent variable is a binary one, we have a linear probability model. In this case, the level of unrelatedness increases the probability that a certain region, in a particular technology, will have at least one super patent in the next period. Similarly to the previous estimations, the GDP per capita has positive and significant impact on the probability of having a super patent in a certain technology field. Also, the population density and its square term present the same behavior as before.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number of super patents (upper 5%)				Apperance of supertents=1			
lag unrelatedness	0.0108***	0.00514***	0.00299***	0.00299***	0.00319***	0.00241***	0.00235***	0.00231***
lag gdp per capita		8.341***	8.328***	8.328***		1.048***	1.263***	1.184***
lag pop. Density		0.116***	0.914***	0.901***		0.0210***	0.124***	0.102***
lag pop. Density (square)		-0.0233***	-0.0675***	-0.0667***		-0.0035***	-0.0089***	-0.0075***
Constant	-0.0545***	-0.159***	-0.110***	-0.190***	-0.0145***	-0.0286***	-0.0206***	-0.0365***
Period fixed effects	NO	NO	YES	YES	NO	NO	YES	YES
Region fixed effects	NO	NO	YES	YES	NO	NO	YES	YES
Technology fixed effects	NO	NO	NO	YES	NO	NO	NO	YES
Observations	1,023,226	854,597	854,597	854,597	1,023,275	854,632	854,632	854,597
R-squared	0.007	0.009	0.023	0.089	0.032	0.032	0.049	0.143

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

REFERENCES

Boschma, R., A. Minondo & M. Navarro (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.

Boschma, R., Heimeriks, G., & Balland, P. A. (2014). Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy*, *43*(1), 107-114.

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research Policy*, *32*(1), 69-87.

Eck, N. J. V., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the Association for Information Science and Technology*, *60*(8), 1635-1651.

Hidalgo, C. A., B. Klinger, A. -L. Barabasi & R. Hausmann (2007). The product space conditions the development of nations, Science, 317, 482–487.

Schmoch, U. (2008). Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Organization, WIPO. Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany.