Explaining the Patenting Propensity: A Regional Analysis using EPO-OECD Data

Cozza, Claudio and Schettino, Francesco

University of Piemonte Orientale, Second University of Naples

1 March 2013

Explaining the Patenting Propensity: A Regional Analysis using EPO-OECD Data

C. Cozza[a], F. Schettino[b,*]

[a] *Department of Economics and Business (DiSEI), University of Piemonte Orientale "Amedeo Avogadro", Novara, Italy*

[b] *Department of Law, Second University of Naples, Via Mazzocchi 5, 81055 S. Maria Capua Vetere, Italy*

1. Introduction

Over the last decades, several scholars have investigated various aspects of the patenting process as a means to depict the multifaceted nature of innovation. Their studies can be divided into two groups: those inquiring into qualitative features of patent and into its value, others investigating patent application determinants. It is well known that Griliches (1990) – remarking that patents are not a perfect measure of R&D output, but still a relevant proxy for technological dynamics – encourages further research, both empirical and theoretical, on patents.

Especially after the diffusion of worldwide patent datasets by USPTO and OECD-EPO, several studies have started to investigate the features of patent documentation, focusing on (backward) citations in particular. By means of that, the relevance of both geographical and industrial knowledge spillovers has been empirically stressed[1]. On the other side, studies on patent quality have been carried out in order to depict the heterogeneity – principally in terms of market value – of granted patents[2]. These studies also analyze patent characteristics such as forward citations, claims and the propensity to apply to the three main patent offices: EPO, JPO and USPTO. However, mainly for what concerns quality determinants, the outcomes are not convergent and homogeneous, probably because the majority of these articles are based on estimations made using different surveys data (see Van Zeebroeck et al. (2011)).

From a purely quantitative point of view, the analyses have been made using patent count and/or distribution (by year, by country or by geographical unit) and their determinants. The pioneering results obtained by Hausman, Hall and Griliches (1984, 1986) seriously influenced subsequent literature: using firm-level data, they investigate the R&D-patents relationship, concluding that the number of patent applications is a function of R&D

*Corresponding author

Email addresses: claudio.cozza@unipmn.it (C. Cozza), francesco.schettino@unina2.it (F. Schettino)

[1] Among others, Caballero and Jaffe (1993), Henderson et al. (2005), Jaffe et al. (2005), Schettino (2007).

[2] See Haroff et al. (1999), Lanjouw et al. (2004) and Hall et al. (2007).

expenditure. Their approach has been followed by other scholars such as Montalvo (1993), Cincera (1997), Hall and Ziedonis (2001) and more recently Cincera (2005) and Yuen (2009). Although they explore this relationship applying more sophisticated econometric tools to different datasets, their key results substantially confirm a relevant influence of R&D expenditure (present or lagged) on patent applications, as in Hausmann et al. (1984, 1986).

Following the latter approach, this paper aims to preliminary assess the patenting propensity from a regional point of view: the novelty of our investigation consists in the fact that we estimate the R&D-patents relationship on the totality of applications filed to the EPO in 2000-2010 by the European region[3] both of inventors' and of applicants' residence. In order to set our panel, we extract data from the OECD-EPO regional patent dataset (REGPAT[4]) counting the yearly number of applications[5] by region[6]. Furthermore, selecting explanatory variables from EUROSTAT and OECD databases[7], we proceed highlighting the differences in terms of patents distribution by inventor and by applicant region of residence; indeed, we estimate the relative impact of R&D expenditure (at business enterprises and governmental level), of investments in high qualification of human capital, of employment level in high tech industries and of enterprise size on patenting propensity.

2. Data and Statistics

As explained in Cozza et al. (2012), EUROSTAT R&D data at the regional level reflect the *execution* of R&D in a specific region, no matter who is the *owner* of the investment. This means that a firm undertaking R&D in many different regions of the same country (and even more a multinational firm), will contribute to EUROSTAT regional figures proportionally to the real involvement in that region. In other words, the location of the headquarters in a different region does not play any role in this kind of data. The choice of variables for the estimations follows the idea that the link between R&D efforts and patent propensity has to be analysed looking at the territory where such efforts have been undertaken.

The same logic is reflected by our use of patent applications by region of inventor. This way, although same caveats apply[8], we catch the connection to the territory where the invention has been made. The residence of the inventor is the best proxy for this, while the use of the region of the applicant would imply a bias towards the location of firm

[3] EU27 + EFTA countries, NUTS 2 level.

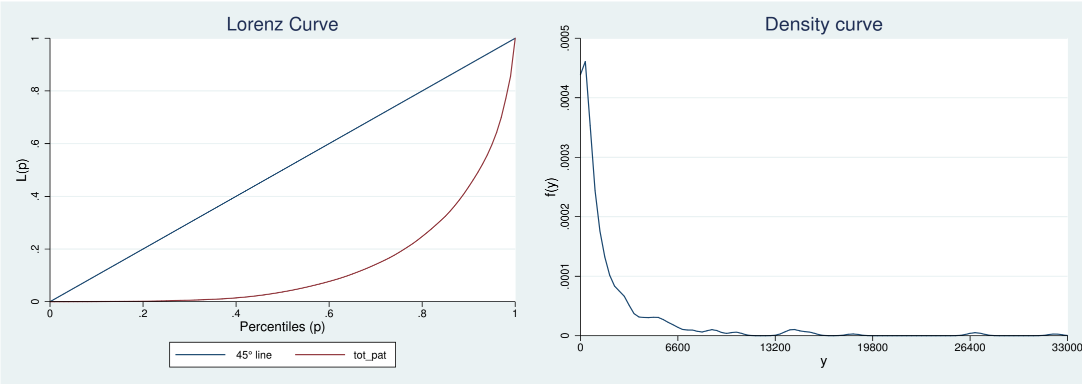
[4]OECD REGPAT database, June 2012 - covers patent applications to the EPO (derived from PATSTAT, April 2012) and PCT patents at international phase (derived from the OECD patent database, including patents published up to May 2012). Note that the regional breakdown refers to the latest revision of the NUTS. The dataset covers regional information for most OECD and EU27 countries, plus the BRICS countries. We thank Helene Dernis for providing data.

[5] Fractioned as in Narin and Breizman (1995) approach.

[6]The choice of the inventor instead of the applicant has been discussed in Section 2.

[7]Selection have been made at NUTS 2 level.

[8]We refer to the case of an inventor commuting everyday from the region of residence to the one of workplace. This way, we would lose the connection to the territory of invention (the workplace region), overestimating the region of pure residence.



Source: Our estimation using OECD-EPO REGPAT data

Figure 1: Distribution of total patent applications by region, years 2000-2010

headquarters. This choice would then be discrepant with that made for R&D data and produce a distortion in the analysis.

We can provide an example. A firm is headquartered in region X, where it performs R&D for 100 and where 5 of its researchers/inventors are resident. This firm has also R&D activities for 50 in region Y and for 30 in region Z, where 3 and 2 of its researchers/inventors are resident. Each researcher is inventor of a single patent. In our approach, the three regions account for 100, 50 and 30 R&D investment, and for 5, 3 and 2 patent applications respectively. Using the R&D *owned* and the region of the applicant, instead, region X would account for 180 R&D investment and 10 patent applications, while regions Y and Z would account for 0. In our opinion, this is strongly misleading in the assessment of the real contribution of each territory (region) to technological change.

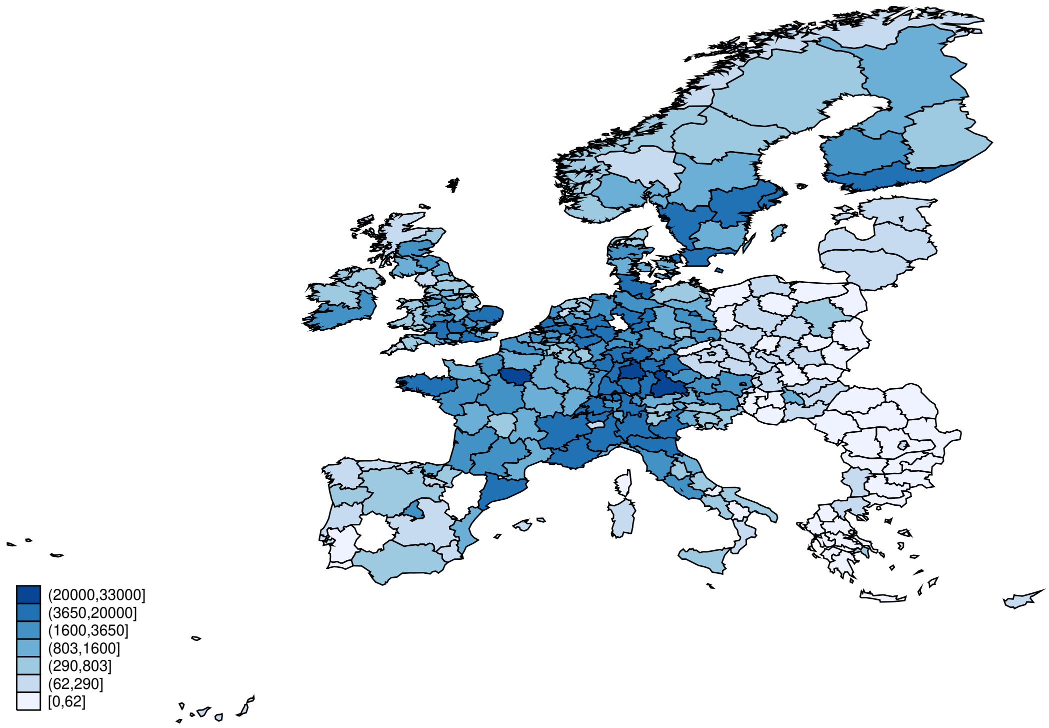
To the same extent, finally, also the other variables have been chosen: the average size of firms by region has been calculated using the number of employees divided by local units in each region; the HRST variable is compiled directly addressing individuals (graduates and workers) and their region of residence; and also the regional GDP is a variable built at the regional level.

Both the REGPAT and the OECD regional R&D databases help in shaping a clear overview of the technological efforts in Europe over a long time-span. We use such data here for a quick insight on the top performing regions and on the level of concentration of both patent applications and R&D expenditure.

Figure 1 highlights the uneven distribution of total applications (2000-2010) by region. Both density and Lorenz curves show a high degree of concentration in top regions, as confirmed also by Figure 4.

Moreover, Figure 2 shows that patent distribution follows, in general, local GDP levels: in richest regions (and countries) the patenting propensity is higher, indirectly confirming the twofold linkage existing between growth and innovation.

From this point of view, Figure 2 depicts significant differences between European countries: the number of applications by inventor resident in eastern regions is strongly lower than in western ones. In the latter group, the regions of the so called *PIIGS* countries (Por-



Source: Our estimation using OECD-EPO REGPAT data

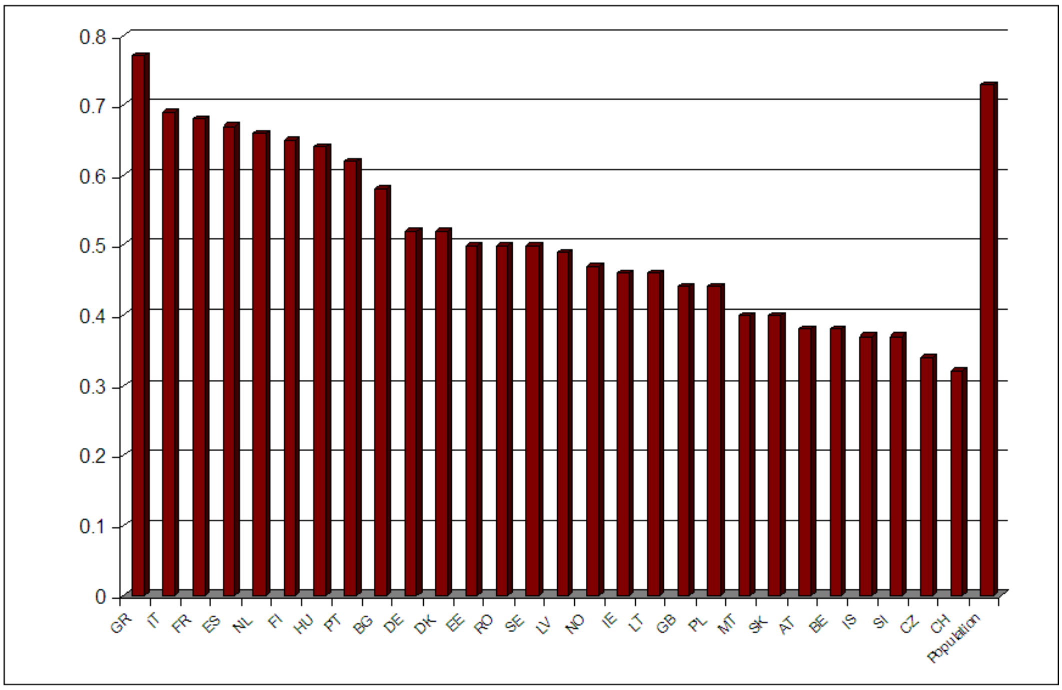
Figure 2: Total Patent Applications by region, years 2000-2010

tugal, Italy, Ireland, Greece and Spain) show a patenting propensity (with the exception of Italian region Lombardia) lower than central and northern European ones (especially those in Germany). It is important to notice that the uneven feature of application distribution has grown after the 2007 crisis explosion. Since the adoption of new technologies is crucial for income growth and catch-up, it emerges that the absence of strong political measures by local policymakers would increase the distance among these groups of European regions.

In Figure 3, we depict patent concentration by country, performing Gini index on the same time-span. Generally, confirming what highlighted in Figures 1 and 2, the total patent distribution is highly uneven, since Gini outcome for the whole population is equal to 0.72, value that is stable in the whole period[9]. For what concerns the concentration by country (see Figure 3), France and Italy show higher values: this is mainly due to the relative relevance of Ile-de-France and Lombardia (and Lazio) regions, whose inventors filed a significant share of total applications in the period (see below for a deeper explanation). In contrast, the number of applications in three top performing countries, such as Germany, Great Britain or Switzerland, is not concentrated in a limited number of regions: this suggests that, in these countries, a high-level innovation culture is well distributed on the territory as a whole.

For what concerns top performing regions, it is interesting to notice that the same ones

[9]The test performed on Gini2000 and Gini2010 excludes a statistically significant difference between the indexes.



Source: Our estimation using OECD-EPO REGPAT data

Figure 3: Gini index by country – Total Patent Applications, years 2000-2010

appear both in the overall period (1978-2011) and in the shorter period used for the econometric estimations (2000-2010). The majority of inventors are resident in the regions of the most important industrial European cities (e.g. Paris, Munich, Stuttgart, Eindhoven, Milan) and German regions are definitely predominant (six regions out of the top ten). Also the relative weight of such ten regions over the total remains stable, comparing the two periods. They account for about one third of total patent applications (see Table 4).

The same regions show up also when considering the gross investment in R&D. In year 2009, for instance, eight of the above mentioned regions are also the top performers in R&D, with minor exceptions. As shown in Figure 5, in any case, the composition of such gross investment varies across regions, with some of them strongly relying on business expenditure (e.g. Stuttgart), others on higher education expenditure (e.g. Inner London) and some others on government expenditure (e.g. Madrid or Koln).

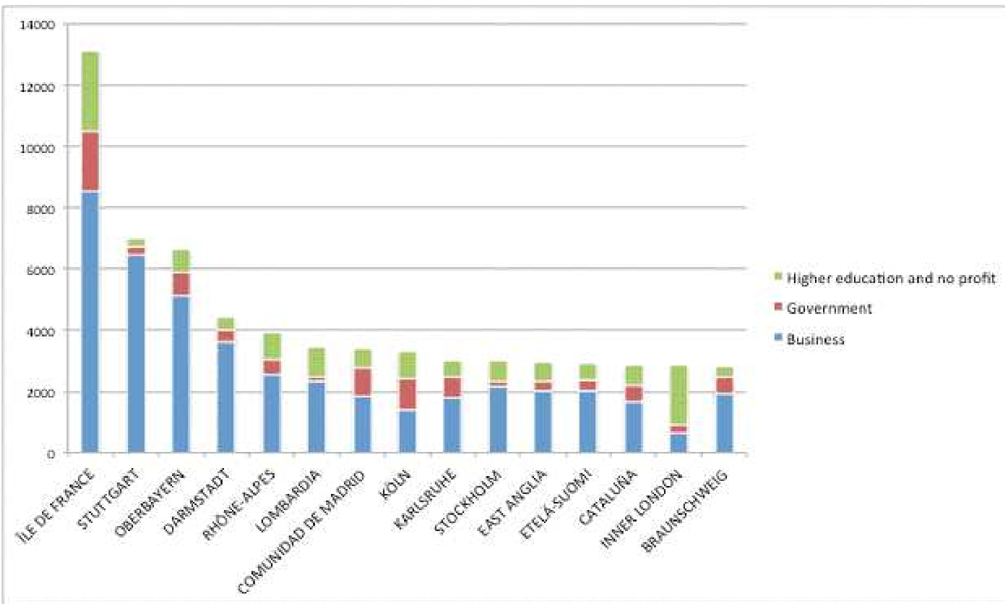
On the basis of the same reasons sketched above, also the R&D expenditure concentration by region appears similar to the patent applications one. Figures 6 and 7 depict a Europe characterized by high levels of R&D expenditure (business and total) in central-western regions, i.e. in richer countries, while the eastern and *PIIGS* regions do not present satisfactory data.

Main countries also show a high degree of internal variability. As shown in Figures 3 and 8, top patent application countries in the 2000-2010 period (e.g. France, Germany and the Netherlands) exhibit a high distance between their respective top and least regions. For instance, values for Germany range from 26,998 of Stuttgart region to 532 of Trier region.

Region	Patent applications 1978-2011	Cumulated percentage of patent applications 1978-2011	Ranking 1978-2011	Patent applications 2000-2010	Cumulated percentage of patent applications 2000-2010	Ranking 2000-2010
ÎLE DE FRANCE	71.56	6%	1	32.152	5%	1
OBERBAYERN	53.514	11%	2	26.517	10%	3
STUTTGART	48.38	15%	3	26.998	14%	2
DARMSTADT	37.671	18%	4	15.606	17%	5
DÜSSELDORF	35.333	21%	5	15.338	19%	6
KÖLN	31.878	24%	6	14.271	22%	9
NOORD-BRABANT	30.303	26%	7	18.446	25%	4
RHÔNE-ALPES	28.721	29%	8	14.527	27%	8
LOMBARDIA	28.532	31%	9	14.544	29%	7
KARLSRUHE	26.547	33%	10	13.943	32%	10
Top regions total	392.439			192.342		
Overall total	**1.176.346**			**607.06**		

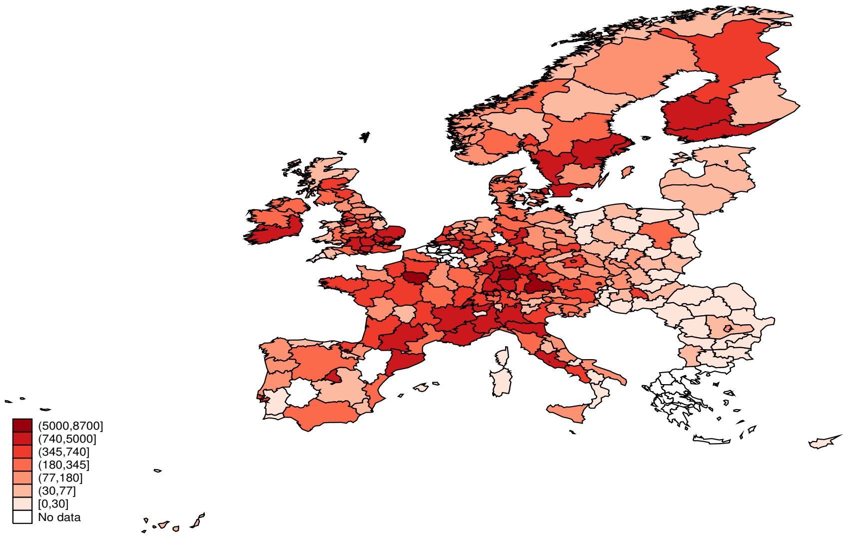
Source: Our estimation using EPO-OECD REGPAT dataset

Figure 4: Total Patent Applications in Top Regions, years 1978-2011



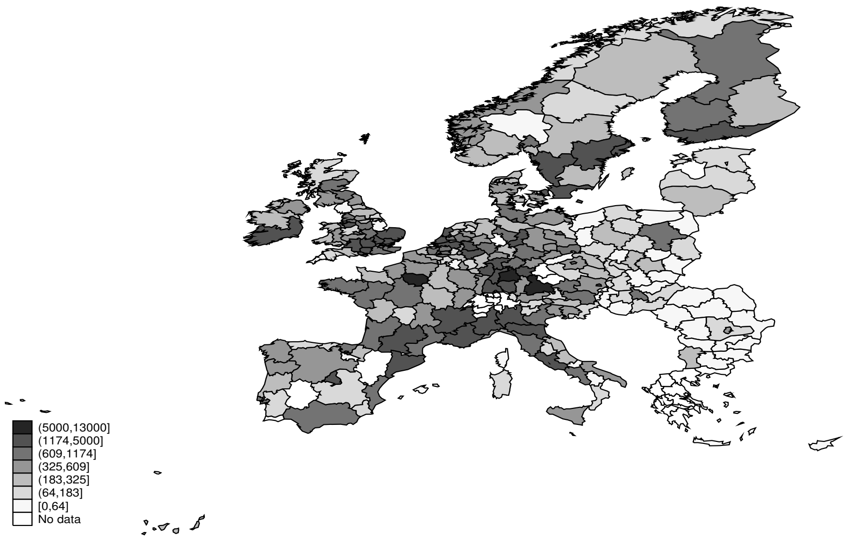
Source: Our estimation using EUROSTAT/OECD data

Figure 5: Gross Investment in R&D in top performing regions, in 2000pps, year 2009, by sector of performance



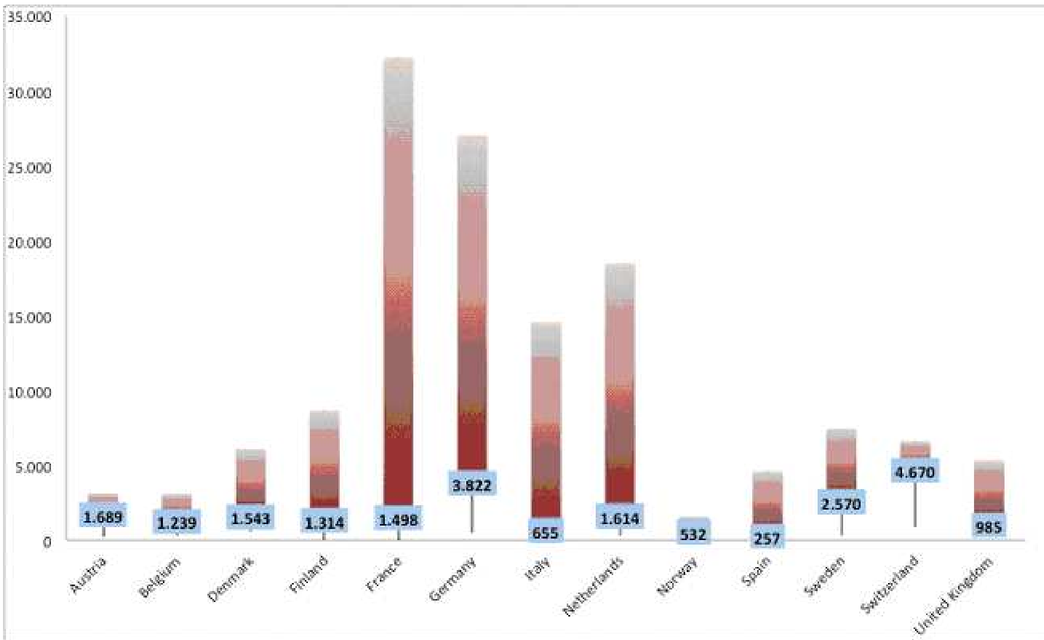
Source: Our estimation using EUROSTAT/OECD data

Figure 6: Average Business Enterprises R&D expenditure (BERD) by region, years 2000-2009



Source: Our estimation using EUROSTAT/OECD data

Figure 7: Average gross R&D expenditure (GERD) by region, years 2000-2009



Source: Our estimation using EPO-OECD REGPAT dataset

Figure 8: Total patent applications in top and least region and median in each country, years 2000-2010

Of course, such values incorporate the differences in regional extension and populations. Therefore, in Figure 8, we add the median value, in order to distinguish between two groups of countries: those whose regions show a high number of patents and are all close to the country median (e.g. Switzerland and Sweden); and those whose median is far lower than the top value. Within this second group, the case of Italy is particularly significant: it is the fourth country for the top region value (14,544 by Lombardia region), but it also shows the lowest regional applications (17 for Molise region) and one of the lowest median values.

Previously we have described the principal features of our whole dataset: as remarked, the lack of data, for a significant number of region in the first two decades, leads us to restrict the empirical analysis to the period 2000-2010. This choice is justified from a theoretical point of view by the fact that many European countries strongly changed their IPR structures after USSR collapse; this long process began about in 1990 and stabilized not before than ten years. From an empirical point of view, in the years before 2000, for many independent variables we found the so-called "excess zero" problem (see also in Yueh (2009)); as a consequence, in order to limit the correlated regression bias, we produce our estimations on the 2000s decade data. Thus, our panel data (unbalanced) is composed of 3,209 raws, that is less than 300 observations per year[10].

It is well known that the discrete non-negative nature of patent data (count data) makes the usual linear regression models inappropriate: at the same time, Hausmann et al.(1986) remark that a Poisson process could be certainly useful in order to describe events that

[10] The observations number fluctuates year by year (282-300) because the number of region per country is generally constant in time but for the eastern ones, whose regional structure changed in last decade.

happen independently and randomly in time, as the patent production does. Thus, in line with Cincera (1997 and 2005) and Yueh (2009) we adopt two distinct models – Negative Binomial and Poisson – in order to estimate the regional patent propensity function relying on the following determinants:

1. Private R&D expenditure (*Berd*[11]);
2. Government R&D expenditure (as a fraction of local Gdp) (*Goverd_gdp*);
3. Share of human resources employed in high tech industries (*Htec_emp*);
4. The number of highly qualified and/or workers in science and technology occupations (*Hrst*);
5. The average enterprise size (*Ent_size*).

The use of R&D without lagged time, besides a similar use in literature (see Hall and Ziedonis, 2001), is justified by the persistent nature of R&D. It is, in fact, highly improbable that a firm undertaking a certain amount of R&D in year *t-2* (or *t-4* in other sectors, usually science-based, needing longer time for getting results from R&D) and applying for patents in year *t*, presents in the same year *t* a considerably lower level of R&D.

We use the *Goverd* variable as a share of GPD for two main reasons: 1. in order to remove hypothetical collinearity with other variables (first of all, the *Berd*); 2. in order to eliminate a possible related "regional bias". In the case of France or Italy, for instance, the highest share of R&D spending by public research institutions such as CNRS or CNR is located in the region of the capital city. This happens although both institutions perform their R&D all over the national territory. Therefore, the use of the variable as a share of GDP reduces the distance of these two regions towards the rest of the respective countries. For other federal countries (e.g. Germany or Switzerland) this issue does not create the same bias.

The variable *Ent_size* is computed as the number of employees in the region divided by the number of local units in the region. This reflects the average size of the local units (laboratories, plants, offices etc.) of a region, not of the firms as a whole. In other words, this "size" is not excessively influenced by the presence of large firms, especially those widespread on the national territory or abroad.

It is well known that, in panel analysis, a first problem is represented by the choice of random/fixed effect estimation: in literature, however, it is stated that, when firm- or region-specific effects are correlated with some explanatory variables, the random effects model is no longer consistent. In particular, as Cincera (1997) remarks, in R&D-patents relationship, the correlation between dependent and independent variables is significant and, indeed, random effect specification could give an upward-biased estimate. This intuition has been confirmed by the Hausman test that we performed on each model: as we can see in Table 2, fixed effect is considered as the best specification.

Table 3 shows results substantially consistent both with Cincera (2005) and with Yueh (2009). The coefficients of the variables related to human capital are significant in both

[11]Both BERD and GOVERD have been calculated in PPS, 2000 prices

	Negative Binomial	**Poisson**
Test: Ho: difference in coefficients not systematic		
chi^2 = (b-B)'[(V_b-V_B)^(-1)](b-B)	18.61***	119.19***
Prob>chi^2	0.0023	0

, **,* = <0.1, <0.05, <0.01 confidence level*

Source: Our estimations on REGPAT and OECD datasets

Table 1: Hausman test output

	Negative Binomial (NBM)	**Poisson (PM)**
Dependent Variable		
Patent Application by inventor (2000-2010)	Coefficients	Coefficients
Determinants		
Hrst	0.0006***	0.0004***
Htec_emp	0.014***	0.011***
Ent_size	-0.00007	-0.0001**
Berd	0.0001***	0.00004***
Goverd_gdp	0.26***	0.033
constant	3.44***	
Number of obs	1257	1257
Number of groups	209	209
Wald chi^2	248.68	400.01
Prob > chi^2	0	0

, **,* = <0.1, <0.05, <0.01 confidence level*

Source: Our estimations on REGPAT and OECD datasets

Table 2: Regressions results

models: it means that investments in high qualified human capital (*Hrst*) and the level of employment in high tech industries (*Htec_emp*), increase the EPO patenting propensity. For what concerns R&D expenditure, the *Berd* has a positive and strong impact on the dependent variable: it means (for NBM) that 1000 euros of business investment in R&D leads to, on average, one patent application. Governmental expenditure on R&D has a significant impact in the NBM, while this is not significant in the Poisson Model. Anyway, the positive outcome in NBM suggests that a higher share of regional GDP spent in public R&D increases not only the patent propensity of local public researchers but also, by means of regional knowledge spillovers, that of private researchers[12].

Finally, the enterprise size variable (*Ent_size*) is not significant in NBM, while it has a weak negative impact in PM. Such a result in the estimations for "size" can be explained by the way this variable has been constructed, as sketched above. For instance, a large firm with 1.000 employees and 10 similar local units in 10 regions, accounts in each region as a medium firm with 100 employees in its single location. Given that, at the regional level the R&D spending of this medium firm has the same "weight" as that of 1 of the 10 local units of the large firm. Therefore, a region with a low "size" can be either a region with only small firms resident in it or a region with small firms and local units of large multi-located firms. This explains the counterintuitive behaviour of the size variable in our models.

3. Preliminary Conclusions

This paper has studied the patenting propensity at the European regional level, using the REGPAT dataset for patents and EUROSTAT-OECD databases for explanatory variables. Descriptive and econometric results confirm two main issues:

1. both R&D performance and patenting are strongly concentrated in some European regions, mainly within central and northern countries, notwithstanding the stability of distribution form in time-span considered;
2. R&D spending, in particular the private one, has a positive and strong impact on patenting propensity.

Such an uneven distribution suggests that local policymakers, especially in eastern and southern countries, should take strong political measures to reduce the existing distance with the top European ones. Among the possible tools, direct and indirect support to innovation can be recalled:

1. increase in the overall governmental expenditure in R&D;
2. incentives (e.g. tax credits) to increase private expenditure in R&D;
3. support to high-tech departments in public research centres and universities;
4. promotion of science and technology employment careers.

[12] See Hall et al. (2003).

A second value added of the paper is the exclusive reference to variables representing the direct linkage to the territory. That is: the regions where the R&D is in effect *executed*, no matter if by a local firm or by the subsidiary of multi-located and multinational firms; and the regions of residence of the inventors, not that of the applicants. In fact, while the correlation of R&D *owned* and patent applications in regions where large companies have their headquarters would be obvious, the results of our analysis highlight an actual relation between R&D spending (as well as investments in high qualified human capital) and patenting propensity on the territory.

The reasons for such a relation can be twofold:

1. Either many different innovation players (large firms, innovative SMEs, multinational corporations, public research centres) belong to the identified top regions, classifiable as regional innovation systems (see, among others, Cooke et al. (1997));
2. Or the *execution* of R&D and patenting are in effect very centralized in top regions, where this *execution* and its *ownership* still coincide, kept well close to the headquarters of top global players (as in Patel and Pavitt (1991)).

In our model, this twofold explanation is confirmed by the counterintuitive behaviour of the enterprise size variable, playing a not significant or weak negative impact. For increasing the patenting propensity, it is not important for a region to be residence of large companies, as the main driver for it appears to be a distributed innovation on the territory. It is however unclear if this distributed innovation is that of many independent SMEs or of subsidiaries and branches *owned* by large firms resident close or elsewhere. The fact that main regions in our descriptive statistics are those where usually large firms and multinationals have their headquarters suggest that *ownership* might still play a crucial role in addressing innovative activities (both R&D performing and patenting). Further research is, in any case, needed to assess this point.

Keywords: Patents, Intellectual Property Rights, Innovation, EPO, R&D
JEL classification: O34 K29 O4 O53 K19

Bibliography

Caballero and Jaffe (1993),How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth, in Olivier Jean Blanchard and Stanley Fischer, editors, *NBER Macroeconomics Annual 1993, Volume 8* MIT Press (1993)

Cincera (1997), Patents, R&D, And Technological Spillovers At The Firm Level: Some Evidence From Econometric Count Models For Panel Data, *Journal Of Applied Econometrics*, 12: 265-280.

Cincera (2005), The link between firms' R&D by type of activity and source of funding and the decision to patent, *Working paper DULBEA*, Research series, N°05-10.RS, April

Cooke, Uranga, and Etxebarria (1997). Regional Innovation Systems: Institutional and Organizational Dimensions, Research Policy, 26(4-5), 475-491.

Cozza, Paci and Perani (2012), Ownership and Execution of Regional Corporate R&D – Methodological Hints from the Italian Case, *Eurasian Business Review*, Vol. 2, No. 1

Hall and Ziedonis (2001), The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995, *RAND Journal of Economics*, 32 (2001), 1 (Spring), 101-28.

Hall, Link and Scott (2003), Universities as Research Partners, *The Review of Economics and Statistics*, vol. 85, issue 2, pages 485-491

Hall, Toma and Torrisi (2007) The Market Value of Patents and R&D: Evidence from European Firms, *NBER Working Paper 13426.*

Harhoff, Narin, Scherer and Vopel (1999) Citation Frequency and the Value of Patented Inventions, *The Review of Economics and Statistics*, 81: 511-515.

Hausman, Hall and Griliches (1984), Econometric models for count data with an application to the patents-R&D relationship, *Econometrica*, 52: 909-938.

Hausman, Hall and Griliches (1986), Patents and R&D: is there a lag?, *International Economic Review*, 27: 265-283.

Henderson, Jaffe and Trajtenberg (2005), Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment, *American Economic Review*, vol. 95, issue 1, pages 461-464

Jaffe, Trajtenberg and Hall (2005), Market Value and Patent Citations: A First Look, No E00-277, *Economics Working Papers from University of California at Berkeley*

Lanjouw and Schankerman (2004) Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators, *Economic Journal*, 114:441-465.

Montalvo (1993), Patents and R&D at the firm level: a new look, *Revista Espanola de Economia*, 67, 81.

Narin and Breizman (1995), Inventive productivity, *Research Policy*, 24 (1995): 507-519

Patel and Pavitt (1991), Large Firms in the Production of the Worlds Technology - an Important Case of Non-Globalization, *Journal of International Business Studies*, 22, 1-21.

Schettino (2007), Us Patent Citations Data and Industrial Knowledge Spillovers, *Economics of Innovation and New Technology,* vol. 16, issue 8, pages 595-633

Yuen (2009), Patent laws and innovation in China, International, *Review of Law and Economics* 29 (2009) 304–313.

Van Zeebroeck and van Pottelsberghe de la Potterie (2011), The vulnerability of patent value determinants, *Economics of Innovation and New Technology*, 20:3, 283-308

Verspagen, Criscuolo and Narula (2005), Role of home and host country innovation systems in R&D internationalisation: a patent citation analysis, *Economics of innovation and new technology,* 14, 417-433.