Do Inventors Patent More in Urban Areas? Evidence From Employer-employee Data*

Sabrina Di Addario[†] and Andrea Lamorgese[‡]

February 19, 2019

Abstract

The paper assesses empirically whether working in an urban environment increases workers' propensity to patent with respect to their counterparts in non-urban areas. We use administrative data from an employer-employee panel of the Italian manufacturing sector for the years 1987-2006, provided by the Italian Social Security Administration (INPS) and containing information on employees' earnings, work status, and demographic characteristics. To single out inventors we match these data to the records from the European Patent Office (EPO), and we obtain the employment history of more than $13,000$ inventors. Since the dataset provides information on the municipality of residence of both inventors and patenting firms, we can verify empirically whether knowledge spillovers are most relevant across geographically proximate actors (be they individuals or organizations). Results indicate that the urban innovation premium is partly driven by composition effects of inventors' observable and unobservable characteristics, while where inventors currently work is more important than where they acquired experience.

Keywords: Patents, Inventors, Cities, Urban Areas, Agglomeration.
JEL Classification: .

---

*The views expressed herein are those of the authors and not necessarily those of the Bank of Italy.

[†]Bank of Italy, Rome Research branch.

[‡]Bank of Italy, Research Department.

1 Introduction

The literature on agglomeration shows that large cities benefit from higher productivity and thus higher wages than non-urban areas. Besides static advantages, enabling large-city workers to be more productive because of agglomeration economies (see Duranton and Puga, 2004, 2014, for a literature review), productivity increases with population also because the best workers sort in the largest cities and because of the existence of dynamic advantages (De La Roca and Puga, 2016). Advantages may be dynamic, both because large-city workers may be better able to accumulate more valuable experience over time than their counterparts in non-urban areas, and because urban areas facilitate learning.

In this paper we assess empirically whether city size has a positive impact on the propensity to invent both statically and dynamically. In particular we study whether inventors working in an urban environment exhibit a higher propensity to patent than their non-urban area counterparts.

We are by no means the first ones to explore the existence of urban premia at the individual level: the literature on the subject is extremely rich (see, among others, Combes et al. (2008), Mion and Naticchioni (2009), D'Costa and Overman (2014), Rosenthal and Strange (2008), and Duranton and Puga (2004, 2014)). However, while most papers proxy individual productivity with wages, in this paper we are able to measure productivity directly, with the number of patent applications that inventors submit to the European Patent Office (EPO) during their lives.

Focussing on inventors may be important for various reasons. First, studying patents may give specific insights on the learning channel through which agglomeration economies operate (Henderson, 2007). Indeed, inventors are a specific type of knowledge-workers, who may benefit more from quicker learning in large cities than low-skilled workers. To this aim, we adopt De La Roca and Puga's (2017) approach and estimate the dynamic benefits from working in a large city through experience acquired in cities of different sizes.

Second, the majority of the papers examining knowledge spillovers use data at the city or at the zip code level, but not all the employment at the zip or city level necessarily produces knowledge spillovers or innovation. In contrast to most of the previous literature on the subject, we are able to use information at the individual level and thus to focus on the workers who are more likely to create and/or benefit from knowledge spillovers. Jaffe, Trajtemberg and Henderson (1993) find indeed that knowledge spillovers are highly geographically localized. The authors examine patent citations and obtain that the probability of citing a patent is inversely related to the physical distance between the cited and citing inventor. Ten years later, Carlino, Chatterjee and Hunt (2006) show that patent intensity (i.e., the city rate of patenting per capita) increase with both SMA's employment density and the degree of local market structure competitiveness.

We use administrative data from an employer-employee panel of the Italian manufacturing sector for the years 1987-2006, provided by the Italian Social Security Administration (INPS) and containing information on employees' earnings, work status, and demographic characteristics. To single out inventors we match these data to the records from the European Patent Office (EPO) and we obtain the employment history of more than $13,000$ inventors. Since the dataset provides information on the municipality of residence of both inventors and patenting firms, we can verify empirically whether knowledge spillovers are most relevant across geographically proximate actors (be they individuals or organizations).

We find evidence of static advantages of cities, selection of best inventors, and lower learning in small cities. Results indicate that the urban innovation premium is partly driven by composition effects of inventors' observable and unobservable characteristics, but, in contrast to (De La Roca and Puga, 2016), who find that experience matters, it seems that where inventors use experience matters more than where they acquire it. Indeed, even though acquiring experience in a small city slightly lowers the propensity to patent, if the inventor moves from a small city to Rome and Milan the probability increases by 3 percent.

The remaining of the paper is organized as follows: Section 2 and 3 describe the dataset we use for the analysis. Section 4 presents the basic methodology and the results, while Section 5 concludes.

2 The data

In this paper we use two datasets: the European Patent Office (EPO) Worldwide Patent Statistical Database (Patstat) and the employer-employee matched data from the Italian Social Security Institute (Istituto Nazionale di Previdenza Sociale, INPS).

2.1 INPS

INPS provides administrative data on the private-sector employees' work history in the period 1987-2009. In particular, it includes information on workers' age, gender, location (both municipality of residence and birth) and job characteristics, like work status (blue collar; white collar; manager; other), type of contract (full-time versus part-time) and gross yearly earnings. In contrast to most datasets on wages, earnings are not top-coded.

INPS data also includes some variables at the firm level: sector, date of plant opening and closure, and plant location (at the municipality level). Thus, we are able to track inventors over space and compute years experience acquired in the city of work. We are also able to compute years of experience in the labor market and tenure in current firm.

2.2 Patstat

Patstat comprehends the universe of patent applications and grants presented at the EPO. The dataset is very rich, providing information on the characteristics of patent submissions (title, abstract, technological field, dates of application filing, publication, grant obtainment, number of citations, etc.), name and address of both inventors and applicants (i.e., the applicant is the firm or individual submitting the patent application and retaining the relative property rights).

We select all the applications presented by any firm resident in Italy. In April 2009 the stock of Italian firms' submissions at the EPO amounted to 58, 465.[1]

Although Italian firms started applying at the EPO at the end of the 1970s,[2] we exclude the 4, 924 patents submitted before 1987, because this is the first year INPS data were available. We also drop the 4, 738 patent applications filed after 2006, because the EPO takes about three years to update its records and the three most recent years are thus incomplete. Finally, we exclude the applications filed by Universities or Public Research Centers not registered with INPS.[3] Clearly, before dropping these observations we retain the necessary information (e.g. the inventor's stock of patent submission in each period is based on the universe of his/her patents).

We thus are left with 48, 194 patent submissions.

Note that providing figures on the number of inventors and firms is difficult, because the original EPO applicant and inventor identifiers are not reliable. Thus, we standardized both firm and inventors' names and created new identifiers. According to the EPO identifiers, for instance, our dataset would comprise 67.985 inventors, but after standardizing the names of different individuals are just 33.667. Similarly, 14.681 applicants are identified by EPO, but only 10.807 have a different standardized identifier.

2.3 The Patstat-INPS matching

We asked INPS to match our inventors to their employees on the basis of the name, surname, employer, and year of submission. To this aim we gave INPS the list of EPO standardized inventors' and firms' names. We matched Patstat to INPS data in four steps.

In the first step we asked INPS to match the first name and surname of

[1]The EPO releases a new version of Patstat twice a year. In this paper we use the April 2009 version. We exclude all the applications which do not contain the information on inventors or applicants (we were able to retrieve the information on missing firm in 490 cases) and those presented only by individuals.

[2]Precisely, 1 patent was submitted in 1936, 8 in 1978 and 204 in 1979.

[3]Besides Universities, Politecnici, consorzi or union of Universities or Politecnicos, Alma Mater Studiorum, Azienda Ospedaliera Universitaria, Istituto di elettronica ed ingegneria, Ministero dell'Universit e della Ricerca, CNR, ENEA, etc.

the 33,047 Patstat's inventors names (after having standardized the names and surnames) to the employees in their archives in the year of application. INPS was able to find 23,145 individuals having the same name of an EPO inventor and employed in an INPS firm in the year of application (70 per cent of the total). The reason why INPS could not find the remaining 30 percent of the names in its archives may be due to names mis-spelling or to the fact that the inventor was not formally employed in an INPS firm (e.g. consultants, self-employed or non-formally employed, or employed in an institution not registered with INPS).[4] The 23,145 names correspond to 79,299 different individuals because of homonymy cases. To determine which of these were correct matches, we verified whether there was also a correspondence between Patstat's and INPS' firms. To this aim we attributed VAT codes to EPO firms on the basis of the company name and location, since they lack of a firm identifier (like the US Patent Office dataset).

Thus, in the second step we asked INPS to match the 10.807 applicant firms from Patstat to the companies in its archives. INPS was able to find 6,452 companies (corresponding to 14,830 plants) with the same name, address or VAT code of EPO firms.

Finally, we considered a matched to be valid if there was a correspondence between INPS and Patstat' name and surname of inventors, name of inventor's employer, in the year of patent submission. We also checked the correspondence between the two datasets with respect to firm location.

Thus, our employer-employee matched dataset comprises 14,829 matched inventors (64 per cent of the names present in INPS' archives) working in 3,792 establishments. The database includes the full work-history of the private-sector employees working in any of the patenting firms that INPS was able to match, even if they moved from / to a non-patenting firm, in the years 1987-2006. The unbalanced panel includes 226,570 observations (on average, almost 12,000 inventors per year).

2.4 Italian urban areas

The definition of city in the empirical urban literature is heterogeneous. Urban areas do not necessarily overlap with the administrative borders of single municipalities nor with the NUTS3 regions (i.e., provinces in Italy). In this paper the territorial unit of analysis is the Local Labor Market (LLM), which is conventionally deemed as a good representation of a spatial agglomeration (they correspond to the French zones d'emploi and the British Travel-to-Work areas). LLMs partition the Italian territory and are self-contained labor markets: there is a high overlap between workplace and

[4] Most public sector firms were associated to INPDAP, the other main social security institution, besides INPS.

place of residence.[5]

In this paper we establish whether a LLM is urban or non-urban on the basis of the OECD–Eurostat methodology, according to which an urban area is a homogeneous set of territories whose population density exceeds a certain threshold.[6] To define an urban area Eurostat performs a three-step procedure. First, it partitions the territory of the European Union in a grid of 1 square km cells, selecting the cells with a population density of at least $1,500$ inhabitants per square km, while clustering together —in what it calls an urban center— all the neighboring dense cells reaching a population of $50,000$ inhabitants or more. Second, it considers the administrative borders and aggregates to the urban center all the (LAU2) municipalities with at least half of the population resident in the urban center. The so formed unit of agglomeration is defined urban if: i) there is an administrative link; ii) half of the population of the urban area lives within one of the urban centers therein; iii) at least 75% of the population of the urban center therein lives in the urban agglomeration (Figure1).

The urban agglomeration thus defined is the metropolitan *core*. Third, OECD-Eurostat defines an urban area as the union of the metropolitan core and its commuting area. Such urban area is called a Larger Urban Zone (LUZ).[7]

In this work we adopt the first two steps of the OECD-Eurostat definition, while in the third step we consider Italian LLMs, rather than OECD-Eurostat LUZ. Thus, we define an urban area as a LLM containing a urban center as defined by OECD-Eurostat. This methodology singles out 73 urban areas (or urban LLMs) in 611 LLMs in 2011. Non-urban areas are the

[5]At the end of 2014 Istat issued the fourth LLM classification, based on the 2011 Census' commuting flows (see `http://www.istat.it/it/strumenti/territorio-e-cartografia/sistemi-locali-del-lavoro`). The three classifications before were based on the 1981, 1991, and 2001 Census' commuting flows. The 2011 definition has changed so as to be consistent with the European LLM definition.

[6]See `http://ec.europa.eu/eurostat/statistics-explained/index.php/European_cities_%E2%80%93_the_EU-OECD_functional_urban_area_definition`. This definition is consistent with: i) the traditional view that urban agglomerations are the places where production and knowledge spillovers take place, because density creates thick markets and favors the matching between demand and supply; ii) the theory on the Marshallian sources of agglomeration (*labor pooling*, *cost sharing* and *knowledge spillovers*), which are at the core of the birth of the industrial cities in XIX century; iii) the international standards set by the Urban Audit program, thus allowing international comparisons.

[7]The commuting area is similar to the LLM, with slightly different thresholds. Namely, OECD–Eurostat's commuting zones is constituted by all municipalities with at least 15% of residents who work in a neighboring municipality, such that the LUZ is continuous and self contained. A LLM is characterized by the fact that: i) people commute for work reasons; ii) LLMs are self contained (at least 75% of resident work within the LLM; 25% at most outside it); iii) municipalities within a LLM are contiguous (commuting takes place between contiguous municipalities, non contiguous ones are excluded); iv) the *core* of the LLM is the municipality toward which commuting flows are maximum.

remaining LLMs. While the 73 urban areas have grown both in population and size, having absorbed an increasing number of municipalities, between 1981 and 2011 the number of non-urban areas has diminished over time (from 880, to 710, to 612 to 538 over the 4 Censuses).

In the next Section we will further split the urban LLMs into small-, medium- and large-sized cities, respectively, if they are smaller than 250.000 inhabitants, between 250.000 and 3 millions, or above the 3 million threshold (in this category fall only Rome and Milan).

3 Descriptive statistics

As it is well known, the intensity of patents is very skewed both across regions and across cities (see, for instance, Carlino et al. (2007)). In Italy, most applications are concentrated in the North: just the Lombardy region accounts for more than 40 percent of total submissions (see Table 1 and Figure 2). Moreover, patenting activity is mainly concentrated in cities: between 1987 and 2006 urban areas show a significantly larger number of patent submissions, inventors and applicants per 1,000 inhabitants than non-urban areas (Table 2). Indeed, the LLM average probability of patenting in the 1987-2006 period is positively correlated to the log of population (Figure 3).

The distribution of patents per inventor is also very skewed. In our dataset almost two-fifths of the inventors contributes to just one patent submission in their life, less than one-fifth to two, and just 8 percent of the inventors applies at least 5 times (Figure 4, upper panel). The distribution of granted patents is similar: 30 percent of the inventors do not obtain any grant within our observational period, and almost one-third is granted just one (Figure 4's lower panel).

As expected, patents are mostly concentrated in the industrial sector, especially in terms of inventors (98 percent of the total). The retail sector accounts for 1.8 percent of submissions; artisan businesses, which tend to be smaller and more traditional than the others, hardly apply.

Table 3 exhibits the descriptive statistics. Inventors' earn slightly more than $40,000$ euros per year in the 1987-2006 period average. The great majority of inventors are full-time workers and males; the prevailing work-status is white-collar (60 percent of the total). The average employee of our sample is almost 40 years old.

4 Uncovering urban patent premia

4.1 Static urban premia

Table 1's preliminary evidence points toward the existence of an urban average premium in innovation. In this Section we examine whether inventors[8] benefit dis-proportionally from agglomeration externalities, other observable characteristics being the same. To do this we adopt De La Roca and Puga's (2017) two-step procedure for the static model. In the first step we estimate:

$$\mathrm{y}_{ijct} = \alpha_c + \delta X_{it} + \varepsilon_{ijct}. \tag{1}$$

where i indexes the worker, j the firm, c the city (LLM), and t the time; y_{ijct} is the innovation outcome, measured with the probability of submitting a patent at t.[9] X_{it} is a set of time-varying worker observable characteristics (including years of experience in the labor market and its square, tenure in the current job and its square, gender, nationality, work status and type of contract), and α_c are LLM fixed effects. Moreover, we always control for the firm's sector, since workers could be more creative just because they work in industries that are structurally more likely to innovate (i.e., the pharmaceutical sector applies for patents more often than services). Finally, in all regressions we also control for time dummies.

In the second step, we estimate:

$$\alpha_c = \beta I_{urban} + \varepsilon_{ct}. \tag{2}$$

where I_{urban} is alternatively: (a) a dummy variable equal to 1 if the LLM is an urban area; (b) three dummy variables for small, medium and large urban LLMs (versus non-urban areas) - to test whether there are non-linearities in the effect of city size; (c) a continuous variable measuring LLM population size. Thus in the absence of sorting or endogeneity, a positive β, which is our parameter of interest, implies that inventors' productivity grows with city size.

Results are reported in Table 5. The first column shows the outcome of the first-step OLS estimation. Experience and tenure are both significant and exhibit the expected sign, having a full-time contract (rather than a short-term job) increases the probability of submitting a patent, while being male is not significantly different from being female. Surprisingly, the productivity of the inventors who were born abroad is higher than that of

[8]We define as inventor any INPS worker who contributed to a patent application between 1978 and 2009. We thus keep memory also of the patent applications presented before 1987 and after 2006.

[9]We also tested alternative specifications where we substituted the dependent variable with the number of patent applications by inventor i at time t, although we did not obtain different results (available upon request).

the counterparts born in Italy, probably because foreign high-skilled workers are positively self-selected.

In the second stage (reported in columns (5.2)-(5.4), we regress the LLM fixed effects estimated in equation (1) on the various measures of city size described above. We obtain a 0.6 percent-large static urban premium for urban areas, statistically significant at the 1 percent level (specification (5.2)). When we split the urban dummy into the three LLM different sizes we find that the premium is larger in the two biggest Italian cities of Rome and Milan (4.2 percent against 0.4-0.5 percent in the smallest cities; specification (5.3)). When we measure agglomeration with LLM population size, we find that the probability of submitting a patent increases by 0.7 percent for each 1 percent increase in the number of inhabitants (column (5.4)).

4.2 Sorting

The static urban premium estimated in the previous specifications is a very crude measure of the inventor advantage in terms of patent activity, since it could be the result of unobservable characteristics of urban workers correlated with city size: more able workers may settle more easily in the largest urban areas. One of the possible explanations of why this could be the case is that by earning more, more able workers are better able of paying the higher rents due in the more congested cities. If the innovative advantage of large-city inventors were due to a simple composition effect, the urban environment would be no special place in terms of better matching possibilities or better learning opportunities for its dwellers. Although, even in this case, it would interesting *per se* to know whether high-skilled workers like inventors were assortatively matched.

Thus, in Table's 5 specifications (5.5)-(5.8) we repeat the same exercise of the first four columns, while adding individual fixed effects (μ_i) to take into account the impact of composition in term of individuals' unobservable characteristics (i.e., sorting of better workers in urban areas):

$$y_{ijct} = \alpha_c + \mu_i + \delta X_{it} + \varepsilon_{ijct}. \tag{3}$$

When we include these variables the premium slightly increases: from 0.6 to 0.8 percent in terms of the urban dummy, and from 0.7 to 0.12 percent in terms of LLM population (columns (5.6) and (5.8)). Specification (5.7) confirms previous results: the premium is higher in Rome and Milan (where inventors are 3.5 percent more likely to fill a patent with respect to those working in non-urban areas) than in small- and medium- sized cities (where the likelihood is just 0.7 percent higher).

However, we find that, if anything, the OLS urban premium estimation is downward biased, in contrast to other findings of the literature based on wages. Thus, on average, the most able inventors tend to locate in the non-urban areas, possibly because in Italy the industrially denser areas do not necessarily coincide with the biggest cities (e.g. the industrial districts, for instance, are generally located in non-urban areas). When we compare the OLS and the FE estimates of the three urban dummies, however, we find that the large-city premium estimate is slightly upward biased. Thus, the best inventors tend to locate either in Rome and Milan or in non-urban LLMs, showing some sort of polarization.

Thus, while on average the OLS estimation of the urban patent premium underestimates the true effect, the inventors' productivity premium is larger than OLS would predict in small and medium sized towns and it is lower in Rome and Milan.

Note that once we add individual fixed effects, the identification of the city-specific effect, α_c, is estimated only on the subsample of movers, that is on the workers who change municipality from one year to the other. Table 6 and figure 5 show the quantity of switchers year by year. On average, 6.8 percent of the sample's workers switches LLM in the 1987-2006 period.

Table 7 shows that about half of inventors who relocate move across cities at least once (although a quarter just once).

4.3 Dynamic urban effects

In the previous section we dealt with the possible sorting of workers across cities of different size. However, if inventors sorted across urban and non-urban areas on the basis of time-variant individual characteristics the previous estimates would be biased. In particular, if inventors acquired experience dis-proportionally faster (slower) in larger cities than elsewhere, previous' OLS estimate would be upward (downward) biased. Thus, in this section we take possible dynamic effects into account.

In equation (4) we add the effect of learning, measured by the experience acquired in city c:

$$\mathrm{y}_{ijct} = \alpha_c + \mu_i + \sum_j \gamma_{jc} e_{ijt} + \delta X_{it} + \varepsilon_{ijct}. \tag{4}$$

where e_{ijt} is experience acquired in city c.

Clearly, if experience in city increased with population the size of the average bias would depend on the amount of migrations from small to big areas relatively to that from large to small LLMs. If these movements compensated each other the net bias would be small. Table 8 shows that inventors tend to move mainly across LLMs of similar size. Moreover, migration flows from large to small areas are only slightly higher than those from small to large cities.

The size of the bias depends also on the extent to which experience accumulated in a large city is portable once the workers move to a smaller area (see (De La Roca and Puga, 2016)). We will test empirically whether this is the case by estimating equation (4).

Results are reported in Table 9. The first column displays the first stage, where we regress the probability of submitting a patent to the EPO against the same variables as before, plus the interactions between experience and LLM size. In particular, the control variables include years of experience acquired in Rome and Milan, in medium-sized LLMs, and in small-sized cities, in addition to experience interacted with currently working in Rome and Milan, or in a medium-sized city, or in a small-sized LLM. We find that, other things being equal, the years of experience acquired in a small city slightly lowers the propensity to patent, but if the inventor moved from a small LLM to the two biggest cities the probability increases by 3 percent.

Thus, in contrast to (De La Roca and Puga, 2016) it seems that in the inventors' case where workers use experience matters more than where experience was acquired.

Nevertheless, the urban premium remains even in the dynamic model. LLM population size increases the probability of submitting a patent application to the EPO by 1.3 percent. The impact is non-linear, greater for the biggest cities. Working in Rome and Milan raises the likelihood of applying by 2.8 percent.

5 Concluding remarks

(to be continued)

References

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

D'Costa, S. and Overman, H. G. (2014). The urban wage growth premium: Sorting or learning? *Regional Science and Urban Economics*, 48(C):168–179.

De La Roca, J. and Puga, D. (2016). Learning by working in big cities. *Review of Economic Studies*, Forthcoming.

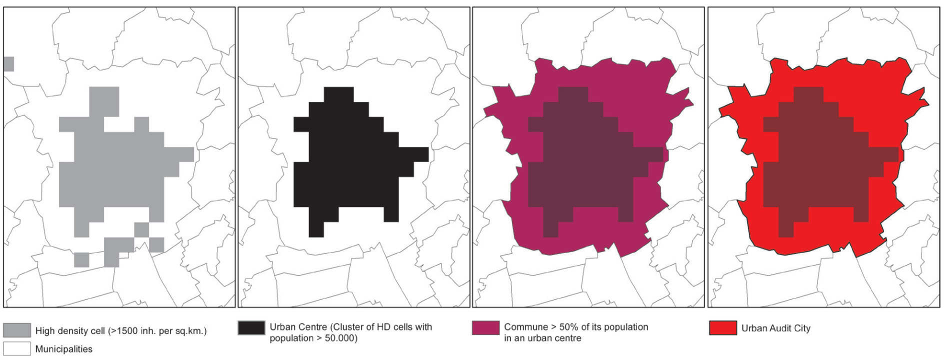
Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Henderson, J. V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4 of *Handbook of Regional and Urban Economics*, chapter 48, pages 2063–2117. Elsevier.

Duranton, G. and Puga, D. (2014). The Growth of Cities. In *Handbook of Economic Growth*, volume 2 of *Handbook of Economic Growth*, chapter 5, pages 781–853. Elsevier.

Mion, G. and Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics*, 42(1):28–55.

6 Tables and figures

Figure 1 Urban centers and urban cities

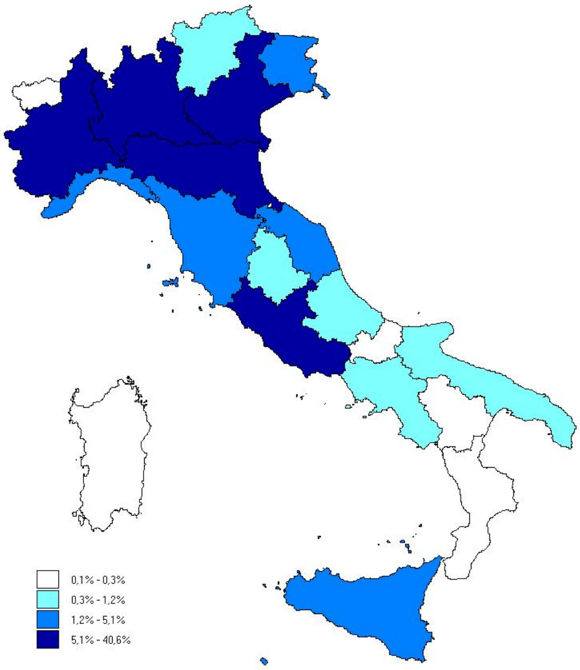


Source: Eurostat.
Note: Municipalities are white-colored; cells of the density grid that are denser than 1500 inhabitants per square km are gray-colored; urban centers (i.e. *cluster* of dense cells with more that 50,000 inhabitants) are black-colored; the sets of municipalities that contain the urban center are crimson-colored; urban areas are red-colored.

Table 1 Distribution of patent applications by region (percent)

	Applications	Inventors	Applicants
Abruzzo	1.2	1.1	0.9
Aosta Valley	0.2	0.1	0.2
Apulia	0.4	0.4	0.8
Basilicata	0.3	0.2	0.2
Calabria	0.1	0.0	0.1
Campania	0.7	0.9	1.4
Emilia-Romagna	11.8	12.4	16.0
Friuli-Venezia Giulia	5.1	4.1	3.3
Lazio	4.7	6.2	4.6
Liguria	2.1	2.7	2.2
Lombardy	41.1	38.6	35.1
Marche	2.0	1.9	2.7
Molise	0.2	0.1	0.1
Piedmont	15.3	17.0	13.0
Sardinia	0.2	0.0	0.1
Sicily	2.0	1.5	0.4
Trentino-Alto Adige / Sudtirol	0.9	0.7	1.3
Tuscany	4.7	4.9	4.5
Umbria	0.5	0.3	0.7
Veneto	6.7	7.2	12.6

Figure 2 Distribution of patent applications by region



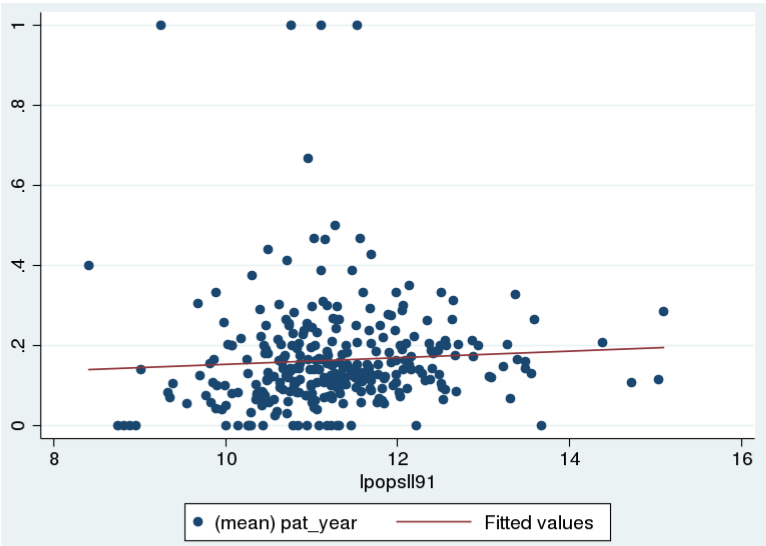
Source: Authors' computation on Patstat data (1987-2006 average).

Table 2 Urban innovation premium: preliminary evidence

	Areas	
	urban	non urban
No. patent submissions/1,000 inhabitants	0.59	0.22
No. inventors/1,000 inhabitants	0.22	0.09
No. applicants/1,000 inhabitants	3.18	2.10
Observations	73	538

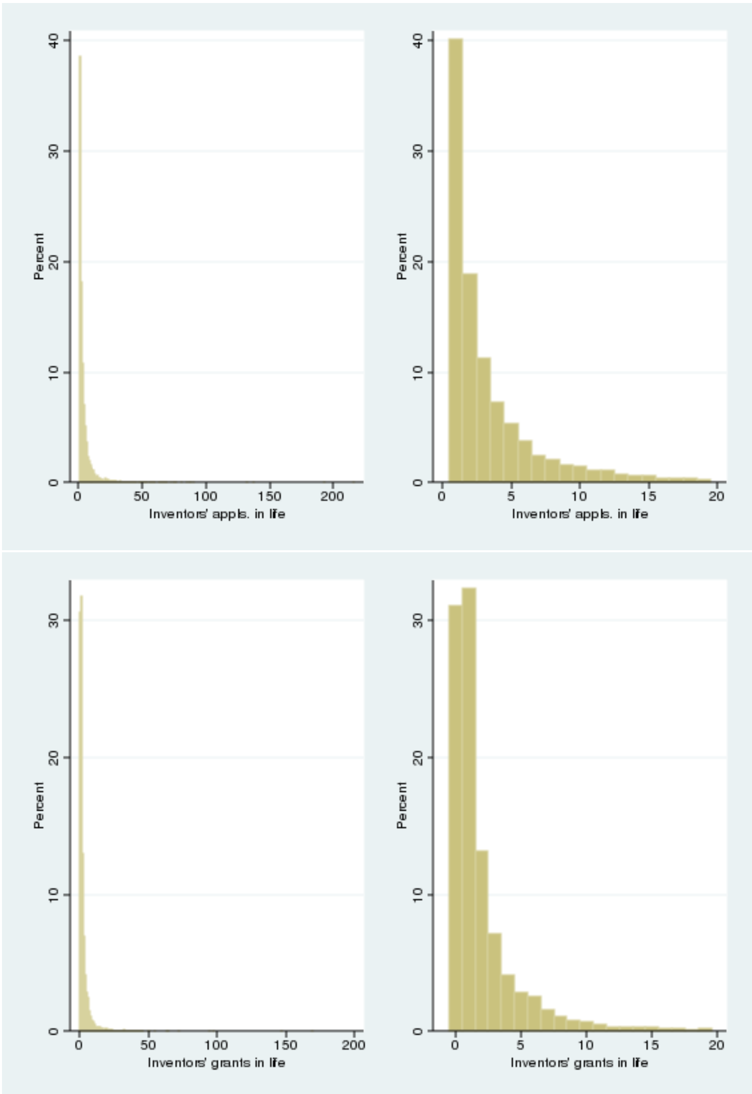
Source: Authors' computation on data from INPS and Patstat, 1987-2006.

Figure 3 Average probability of patenting and population



Source: Authors' computation on Patstat data (1987-2006 average).

Figure 4 Patent applications per inventor (upper panel) and patent grants per inventor (lower panel)



Source: Authors' computation on Patstat data (1987-2006 average).

Table 3 Descriptive statistics

Variable	Mean	S.D.
Yearly wage	40924.9	34009.0
Female	0.1	0.3
Age	39.8	9.2
Blue collar	0.0	0.2
White collar	0.6	0.5
Manager	0.0	0.1
Other work status	0.3	0.5
Full-time	1.0	0.1
No. plants per firm	1.6	2.6
No. obs.	160,217	

Table 4 Descriptive statistics on the regressors

	mean	sd	min	max	median
	Non-urban LLMs				
Citizenship (Italian nationals=1, other=0)	.97	.16	0	1	1
Fulltime workers	.99	.09	0	1	1
Gender (Males=1, Females=0)	.94	.24	0	1	1
Years of experience in any job	8.4	5.5	0	19	8
Years of experience in the plant	4.7	4.6	0	19	3
Years of experience in small urban areas (up to 250k)	.26	1.2	0	18	0
Years of experience in medium urban areas (250k to 3M)	.59	2	0	19	0
Years of experience in large urban areas (Rome and Milan)	.45	1.8	0	19	0
Years of experience in the LLM	6.4	5.3	0	19	5
Observations	53485				
	Urban LLMs				
Citizenship (Italian nationals=1, other=0)	.98	.16	0	1	1
Fulltime workers	.99	.077	0	1	1
Gender (Males=1, Females=0)	.92	.27	0	1	1
Years of experience in any job	8	5.5	0	19	7
Years of experience in the plant	4.4	4.4	0	19	3
Years of experience in small urban areas (up to 250k)	.89	2.8	0	20	0
Years of experience in medium urban areas (250k to 3M)	3.9	5.3	0	20	1
Years of experience in large urban areas (Rome and Milan)	3.6	5.1	0	20	0
Years of experience in the LLM	6.8	5.4	0	19	6
Observations	173120				
	All LLMs				
Citizenship (Italian nationals=1, other=0)	.97	.16	0	1	1
Fulltime workers	.99	.08	0	1	1
Gender (Males=1, Females=0)	.93	.26	0	1	1
Years of experience in any job	8.1	5.5	0	19	8
Years of experience in the plant	4.5	4.5	0	19	3
Years of experience in small urban areas (up to 250k)	.74	2.6	0	20	0
Years of experience in medium urban areas (250k to 3M)	3.1	4.9	0	20	0
Years of experience in large urban areas (Rome and Milan)	2.9	4.8	0	20	0
Years of experience in the LLM	6.7	5.4	0	19	6
Observations	226605				

Source: Inps and Patstat.
Note:

Table 5 Probability of applying for a patent

Dependent variable	Proba of applying for a patent	SLL component			Proba of applying for a patent	SLL component		
LLM's FE	yes				yes			
Workers' FE	no				yes			
year's FE	yes	yes	yes	yes	yes	yes	yes	yes
work status FE	yes				yes			
experience	0.007***				-1.654			
	(0.001)				(24156.116)			
experience^2	-0.001***				-0.001***			
	(0.000)				(0.000)			
tenure	0.010***				0.010***			
	(0.001)				(0.001)			
tenure^2	-0.000***				-0.000***			
	(0.000)				(0.000)			
native	-0.015***				0.000			
	(0.005)				(.)			
fulltime	0.059***				0.067***			
	(0.007)				(0.009)			
male	0.001				0.000			
	(0.003)				(.)			
urban LLMs		0.006***				0.008***		
		(0.001)				(0.002)		
urban $<$ 250K			0.005***				0.007***	
			(0.002)				(0.003)	
250K $\leq$ urban $<$ 3M			0.004**				0.007**	
			(0.002)				(0.003)	
Rome & Milan			0.042***				0.036***	
			(0.004)				(0.002)	
Population				0.007***				0.012***
				(0.001)				(0.001)
R-sq	0.032	0.003	0.006	0.016	0.176	0.001	0.002	0.015
Obs.	226570	5891	5891	5891	226493	5890	5890	5890

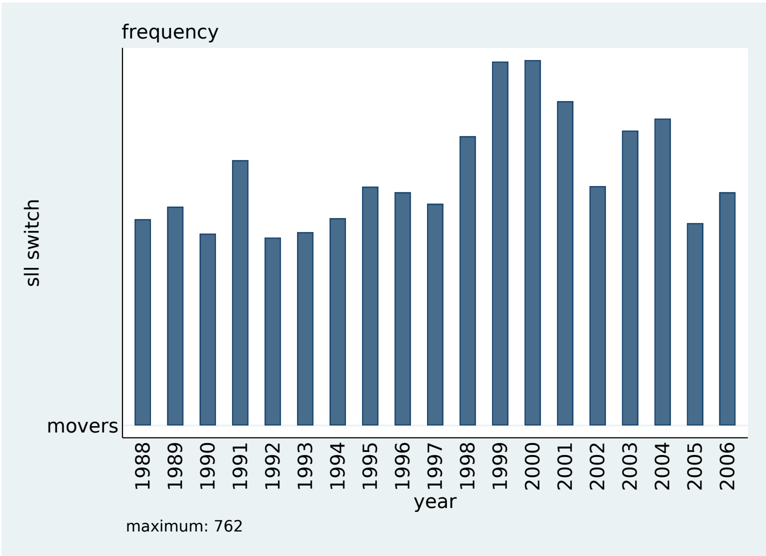
Source: Inps and Patstat.

Table 6 City switchers

	stayers	movers	movers[1]	missing	Total
1987	0	0		7,955	7,955
1988	7,003	618	8.1	656	8,277
1989	7,347	653	8.2	757	8,757
1990	7,758	727	8.6	664	9,149
1991	8,175	655	7.4	550	9,380
1992	8,555	521	5.7	502	9,578
1993	8,657	533	5.8	498	9,688
1994	8,766	629	6.7	556	9,951
1995	9,013	582	6.1	589	10,184
1996	9,254	639	6.5	520	10,413
1997	9,385	689	6.8	761	10,835
1998	9,630	874	8.3	538	11,042
1999	9,770	921	8.6	537	11,228
2000	9,899	1,025	9.4	607	11,531
2001	10,350	813	7.3	454	11,617
2002	10,626	632	5.6	430	11,688
2003	10,635	628	5.6	225	11,488
2004	10,389	738	6.6	334	11,461
2005	10,627	461	4.2	177	11,265
2006	10,365	521	4.8	135	11,021
Total	176,204	12,859	6.8	17,445	206,508

Source: Authors' computation on Patstat data (1987-2006).
Note: [1] Movers as a share of "total" net of "missing".

Figure 5 City switchers



Source: Authors' computation on Patstat data (1987-2006).

Table 7 Re-localizations by inventor

re-localizations	Freq.	Percent	Cumulate
0	7,227	50.27	50.27
1	3,726	25.92	76.19
2	1,993	13.86	90.05
3	882	6.14	96.19
4	357	2.48	98.67
5 - 10	191	1.34	100
Total	14,376	100.00	

Source: Authors' computation on Patstat data (1987-2006).

Table 8 Type of re-localization

Re-localizations	Freq.	Percent	Cumulate
from Rome and Milan to non-urban areas	286	2.22	2.22
from Rome and Milan to small urban areas	142	1.10	3.33
from Rome and Milan to medium-sized urban areas	659	5.12	8.45
from medium-sized to non-urban areas	325	2.53	10.98
from medium-sized to small urban areas	144	1.12	12.10
from small to non-urban areas	128	1.00	13.10
between LLM of similar size	2,936	22.83	35.93
from non-urban to small urban areas	160	1.24	37.17
from small to medium-sized urban areas	85	0.66	37.83
from non-urban to medium-sized urban areas	285	2.22	40.05
from medium-sized urban areas to Rome and Milan	593	4.61	44.66
from small urban areas to Rome and Milan	157	1.22	45.88
from non-urban areas to Rome and Milan	169	1.31	47.20
Missing	6,790	52.8	100
Total	12,859	100	

Source: Authors' computation on Patstat data (1987-2006).

Table 9 Probability of applying for a patent

Dependent variable	Prob. of applying for a patent	SLL component			SLL component + experience in SLL		
LLM FE	yes						
Workers' FE	yes						
year dummies	yes	yes	yes	yes	yes	yes	yes
work status dummies	yes						
tenure	0.009***						
	(0.001)						
tenure^2	-0.000***						
	(0.000)						
experience in the LLM	0.002***						
	(0.000)						
experience^2	-0.000***						
	(0.000)						
exp. in small cities	-0.004**						
	(0.002)						
exp. x exp. in small cities	0.000						
	(0.000)						
exp. in medium cities	-0.001						
	(0.001)						
exp. x exp. in medium cities	0.000						
	(0.000)						
exp. in Rome and Milan	-0.000						
	(0.003)						
exp. x exp. in Rome and Milan	-0.000**						
	(0.000)						
exp. in small cities	0.030***						
x now in Rome and Milan	(0.006)						
exp. in medium cities	-0.001						
x now in Rome and Milan	(0.003)						
exp. in Rome and Milan	-0.000						
x now in Rome and Milan	(0.003)						
exp. in small cities x exp.	-0.002***						
x now in Rome and Milan	(0.000)						
exp. in medium cities x exp.	0.000						
x now in Rome and Milan	(0.000)						
exp. in Rome and Milan x exp.	0.000						
x now in Rome and Milan	(0.000)						
urban LLMs		0.014***			0.014***		
		(0.002)			(0.002)		
urban < 250K			0.017***			0.018***	
			(0.003)			(0.003)	
250K ≤ urban < 3M			0.008***			0.008***	
			(0.003)			(0.003)	
Rome & Milan			0.027***			0.028***	
			(0.002)			(0.002)	
Population				0.013***			0.013***
				(0.001)			(0.001)
R-sq	0.179	0.004	0.004	0.018	0.010	0.010	0.024
Obs	226493	5890	5890	5890	5890	5890	5890

Source: Inps and Patstat. *Notes:* native dummy and fulltime worker are always controlled for.

Table 10 Probability for applying for a patent (second step): male vs. female inventors

Dependent variable: dummy variable for applying						
	(1)	(2)	(3)	(4)	(5)	(6)
Males						
urban LLMs	0.004***			0.007***		
	(0.001)			(0.002)		
urban $<$ 250K		0.001			0.005*	
		(0.002)			(0.003)	
250K $\leq$ urban $<$ 3M		0.005***			0.007**	
		(0.002)			(0.003)	
Rome & Milan		0.040***			0.035***	
		(0.005)			(0.002)	
Population			0.007***			0.011***
			(0.001)			(0.001)
	(0.001)	(0.001)	(0.010)	(0.002)	(0.002)	(0.017)
R-sq	0.002	0.005	0.016	0.001	0.002	0.013
Obs.	5846	5846	5846	5845	5845	5845
Females						
urban LLMs	0.002			0.006		
	(0.002)			(0.004)		
urban $<$ 250K		0.004			0.010**	
		(0.003)			(0.005)	
250K $\leq$ urban $<$ 3M		-0.001			0.002	
		(0.002)			(0.004)	
Rome & Milan		0.021***			0.016***	
		(0.005)			(0.004)	
Population			0.004***			0.006***
			(0.001)			(0.002)
R-sq	0.005	0.012	0.012	0.004	0.005	0.009
Obs.	1595	1595	1595	1594	1594	1594

Source: Inps and Patstat.

Table 11 Probability for applying for a patent (second step): old vs. young inventors

Dependent variable: Dummy for applying for a patent						
	(1)	(2)	(3)	(4)	(5)	(6)
Old inventors						
urban LLMs	0.001			0.004*		
	(0.001)			(0.002)		
urban $<$ 250K		-0.001			0.003	
		(0.002)			(0.003)	
250K $\leq$ urban $<$ 3M		0.001			0.003	
		(0.002)			(0.003)	
Rome & Milan		0.036***			0.029***	
		(0.005)			(0.002)	
Population			0.006***			0.010***
			(0.001)			(0.001)
R-sq	0.006	0.010	0.017	0.001	0.002	0.013
N	4930	4930	4930	4929	4929	4929
Young inventors						
urban LLMs	0.001			0.004*		
	(0.001)			(0.002)		
urban $<$ 250K		-0.002			0.003	
		(0.002)			(0.003)	
250K $\leq$ urban $<$ 3M		0.001			0.004	
		(0.002)			(0.003)	
Rome & Milan		0.033***			0.030***	
		(0.005)			(0.002)	
Population			0.004***			0.010***
			(0.001)			(0.001)
R-sq	0.004	0.008	0.008	0.001	0.002	0.012
Obs.	4950	4950	4950	4948	4948	4948

Source: Inps and Patstat.
Notes: Old inventors are those aged above the median of the age distribution, 39 years.

Table 12 Probability for applying for a patent (second step): fat cats vs. newcomers

Dependent variable: Dummy for applying for a patent						
	(1)	(2)	(3)	(4)	(5)	(6)
Fat cats						
urban LLMs	-0.002*			0.007***		
	(0.001)			(0.002)		
urban < 250K		-0.004**			0.007**	
		(0.002)			(0.003)	
250K $\leq$ urban < 3M		-0.003			0.004	
		(0.002)			(0.003)	
Rome & Milan		0.026***			0.027***	
		(0.005)			(0.002)	
Population			-0.000			0.008***
			(0.001)			(0.002)
R-sq	0.001	0.004	0.001	0.002	0.002	0.008
Obs.	4213	4213	4213	4213	4213	4213
Newcomers						
urban LLMs	0.004***			0.004*		
	(0.001)			(0.002)		
urban < 250K		0.001			0.003	
		(0.002)			(0.003)	
250K $\leq$ urban < 3M		0.005***			0.003	
		(0.002)			(0.003)	
Rome & Milan		0.039***			0.031***	
		(0.005)			(0.002)	
Population			0.006***			0.010***
			(0.001)			(0.001)
R-sq	0.002	0.005	0.015	0.001	0.001	0.011
Obs.	5619	5619	5619	5618	5618	5618

Source: Inps and Patstat.
Notes: Fat cats are those inventors who have more than the median number of lifetime applications (2); newcomers are those with less applications than the median number.