Stars and Comets: An Exploration of the Patent Universe

Carlo Menon (SERC, Department of Geography and Environment, London School of Economics)

November 2009

Stars and Comets: An Exploration of the Patent Universe

Carlo Menon[*]

November 2009

[*] SERC, Department of Geography and Environment, London School of Economics

Abstract

The analysis of patent and citation data has become a popular source of evidence on localized knowledge spillovers and innovation. Nevertheless, an aspect has been overlooked: the patent distribution across inventors is extremely skewed, as many inventors register one or a few patents, while a small number of inventors register many patents. To our knowledge, the previous empirical literature has not discussed the different kinds of local innovation from which patents may originate. A first contribution of this paper is therefore to document the issue. A second contribution is to investigate whether patents originating from different scales of innovation are located in different cities. A third contribution - which constitutes the main scope of the paper - is to test whether the concentration of the activity of star inventors is beneficial to the local productivity of other kinds of innovation - namely the ones led by more occasional, and less prolific, inventors.

1 Introduction

The analysis of patent and citation data has become a key source of evidence on localized knowledge spillovers and innovation. Nevertheless, one aspect has been generally overlooked: the patent distribution across inventors is extremely skewed, as many inventors register one or a few patents, while a small number of inventors register many patents. Innovations developed by inventors at the opposite extremes of the distribution are unlikely to be the outcome of an homogeneous innovation "black box". Interestingly, this peculiar characteristic of the patenting activity recalls the more general "innovation paradox" highlighted in the innovation literature (e.g. Acs and Audretsch, 1990): while big companies massively invest in formal R&D activities, many new products and processes are generated by small and medium firms, with little or no reported investments in R&D. The latter kind of innovation process is therefore more likely to be based on learning-by-doing and informal innovation, being thus intrinsically different from the activity of "professional scientists".

To our knowledge, none of the previous empirical literature on "local innovation" based on patent data has discussed the different "innovation scales" patents may originate from. A first contribution of this paper is therefore to document the issue. A second contribution is to investigate whether patents originating from different categories of inventors are located in different cities. A third contribution - which constitutes the main aim of the paper - is to test whether the concentration of the activity of star inventors is beneficial to the local productivity of more occasional, and less prolific, inventors.

In order to achieve that, using the USPTO/NBER database we identify two illustrative categories of inventors situated in the tails of the distribution: we define as *stars* those inventors who are highly productive in a time window of 8 years - while we define as comets those inventors that develop only one or two patents in same time window. A preliminary data inspection at MSA level shows how the association with establishment births and other MSA structural characteristics and number of patents is significantly different for the two patent categories. This confirms that the categorization is not trivial, and suggests that i) the two categories may relate to different innovation processes, and ii) stars and comets are concentrated in different cities, especially after controlling for the general distribution of the patenting activity.

The location of investments of big companies is increasingly influenced, directly or indirectly, by local policy makers: the attraction of "million dollar plants" is seen as a successful policy targeted at increasing the productivity of incumbent (small) firms through technological spillovers (Greenstone et al, 2008). Similarly, local policy makers may be keen to attract R&D labs of big companies within their jurisdiction. Our results do not seem to support the effectiveness of these policies: we find some evidence suggesting that the direct impact of stars on the local economy is negligible; however, the lack of direct effects might be compensated by indirect effects operating trough an increase of the activity of comet inventors, which in turn may justify the provision of public money to place-marketing policies.

Therefore, in the second part of our empirical analysis we assess whether the activity of star inventors is beneficial to the production of comet patents, and try to quantify this effect. More specifically, using the NBER/USPTO patent database we estimate a model where the number of comet patents produced in a given city, time period, and technological category is a function of the number of star patents developed in the same city, period, and category. We exploit the

panel dimension of our dataset to account for various fixed effects, and adopt an instrumental variable approach to avoid a potential endogeneity bias. In our preferred estimation, we find that, on average, 10% more patents developed by star inventors lead to 2-3% more patents authored by comet inventors.

2 Patents, localized knowledge spillovers, and the size of innovation

Patent data have become extremely popular in the economic literature in the last two decades, as they represent an easy and accessible way to proxy for an economic activity which is generally very hard to measure, i.e., innovation. Furthermore, the availability of citation linkages has added even more interest in patents data: for the first time, researchers had a tool to "trace" knowledge spillovers, which previously had been considered as one of the most intangible concepts in economic theory. A popular book by Jaffe and Trajtenberg (2005), and the free availability of the USTPO dataset from the NBER website, further contributed to multiply the empirical applications based on patent data.

A significant part of this literature has focused on the geographic component of innovation, with a particular interest in the spatial decay of knowledge spillovers. A seminal contribution by Jaffe et al (1993) showed that a cited-citing patent couple is twice as likely to be in the same US metropolitan area than a couple of technologically similar patents with no citation links. Similarly, Peri (2005) examined the flows of citations among 147 European and US regions to find that "only 20% of average knowledge is learned outside the average region of origin", and Jaffe (1989) demonstrated that academic research has large effects on the number of private patents developed in the same US state. Finally, Carlino et al (2007) used patent data for a cross-section of US metropolitan areas to investigate the relationship between urban density and innovation intensity (as measured by patents per capita) finding a positive and robust association, with the caveat that many omitted variables might explain the positive correlation.[1] All these contributions (and many similar which we omit for brevity) highlight that knowledge spillover have a geographically limited distance decay.

It is also important to stress that the nature and causes of knowledge spillovers are still debated. For instance, Breschi and Lissoni (2009), building on previous contributions by Breschi and Lissoni (2001), Zucker et al (1998), and Almeida and Kogut (1999), highlighted how defining localized knowledge spillovers as an *externality* can be misleading, as most of the knowledge diffusion may take place through market interactions - namely the spatially-bounded mobility of inventors among workplaces - rather than through informal contacts. Using data on US inventors' application to the European Patent Office, they were able to show that after controlling for inventors' labour mobility and the related professional network, the role of proximity in explaining knowledge diffusion is greatly reduced.

Previous contributions, however, did not take into consideration an important feature of

---

[1]The authors include a robustness test based on IV estimation, but, in our opinion, the exogeneity of all the instruments is questionable, as they may affect patenting through e.g. productivity. Also, it is not very clear how sorting of very productive inventors and companies into denser cities may influence their results.

patent data, i.e., the skewness of the distribution of patents across inventors.[2] This is in part due to the fact that until very recently an unique identifier for inventors was not available in the NBER/USPTO database and therefore calculating the distribution of patents by inventors was infeasible. Thanks to the efforts of Trajtenberg et al (2006), who "estimated" an unique inventor identifier using an ad-hoc algorithm,[3] we know that out of 1,600,000 inventors listed in the NBER dataset in the period 1975-99, 60% of them registered just one patent, 30% from 2 to 5, and only 0,15% (2,402 inventors) more then 50 patents.

The peculiar distribution of patents by inventors reveals that the innovation process which patenting is a proxy for is an extremely composite phenomenon. On one side, a large number of patents is developed by "comets", i.e., individuals who apply for a patent only once or twice over a long period. On the other side, a small group of "stars" develop individually a huge number of patents. This, beyond being an interesting fact per se, poses a number of questions related to the geography of innovation: do different categories of inventors interact with the local economic environment in the same way? Do they respond similarly to the same location determinants? Are they equally distributed over space or do they tend do concentrate? Is spatial proximity beneficial for their activity?

These questions are related to the growing interest in peer effects in science and in the spillovers originating from star scientists. Among the most interesting recent contributions, Azoulay et al. (2008) exploit the exogenous variation in the number of "superstar scientists" in US university originated by the sudden death of these individuals to estimate the loss in productivity of their collaborators. They find an average 5-10% decline in their average publication rates, starting 3-4 years after the superstars' death and enduring over time, but no differential effect for co-located collaborators. Waldinger (2009) estimates the effect of the dismissal of scientists from Germany Universities during Nazism. Similarly to Azoulay et al., he finds a strong effect on coauthors (13-18%), but no significant effects at department level. Therefore, both the studies challenge the existence of localized positive spillovers originating from stars in academic environments.

Equally on the "skeptical" side, there are the advocates of the "death of distance" theory, who argue for a decreasing importance of the role of spatial proximity following the progress of communication technologies (e.g., Friedman, 2005; Quah, 1999; Cairncross, 1997). On the other side, other economists argue that the technological progress has actually increased the scope for proximity for innovative activities due to the higher importance of face-to-face contacts and agglomeration externalities (e.g. Coyle, 1999). The few empirical assessments of the issue seem to support the "death of distance" hypothesis (Griffith et al, 2007; Ioannides et al, 2008), indeed suggesting that localized knowledge spillovers are fading over time.

Turning to industrial innovation literature, the skewed patent distribution recalls the well-

---

[2] Among the closest contributions we could find, we mention: Silverberg and Verspagen (2007), who analysed in depth the skewness of the distribution of citations across patents; Zucker and Darby (2007) looked at the linkages with private companies of a small sample of star inventors.

[3] The authors needed to face two orders of problems: first, the same author may appear in the database with different names due to spelling errors; second, different authors may have the same name (the "John Smith problem"). The complex algorithm they developed exploits all the available accessory information (dates, locations, technological fields, etc.), toghether with word sound matching routines. The validity of the procedure is confirmed by a test on a dataset of Israeli inventors.

known difference between the innovative activity of small and big companies. In particular, robust evidence on two distinct aspects of small firm innovation poses a challenging "innovation paradox". First, small firms have a much higher ratio of patents developed to R&D expenditures (Griliches,1990) than big companies. If we substitute patents with innovations introduced to the market and R&D with employment, the result is equivalent: the ratio is much higher for small firms (Acs and Audretsch, 1990). The authors argue that this can be due to the higher permeability of small companies to local public R&D inputs (e.g., university research) (Acs et al, 1992). An alternative explanation could be that small companies rely on alternative innovation inputs, based on learning-by-doing and applied innovation, rather than formal scientific research. Second, small firm innovation is all but a residual phenomenon, accounting for most of the innovative activity in many sectors (Acs and Audretsch, 1990). In passing, it is also worth mentioning that small firms account for most of the employment growth in the US in the last decades (Audretsch, 2002). Furthermore, Balasubramanian and Sivadasan (2008) in a recent working paper link patent data with Census firm data for the US, being able to assess the impact of patents on firm performance. They focus in particular on firms that patent for the first time, and find a significant and large effect of the first patent on firm growth (but, interestingly, little change in factor productivity). This would suggest that "occasional" patents have a relevant market value, although further research based on patent-firm matched datasets is needed to explore the issue. As we cannot access this kind of data, in this paper we focus only on patents and their inventors; the "innovation paradox" could be an interesting way to generalize the results whenever patent-firm matched data will be accessible to all the interested researchers.

3   Stars and Comets

Our analysis is based on the NBER/USPTO database, which lists all the patents granted in the United States from 1969 to 1999. We added to this dataset the inventors' unique ID developed by Trajtenberg et al (2006). As the latter is available only since the 1975, our period of analysis is restricted accordingly. More details on the data, including the geocoding process, are reported in Appendix A.

At a first glance, the abundance of data makes a micro analysis at inventor level the most appealing alternative. A deeper view of the data, however, suggests that this is unfeasable, in light of the simple fact that the dataset is about patents, not inventors, which implies that individual inventors are observed only when they patent. When an inventor is not patenting, we do not know their location, their possible employer (i.e., the assignee of their patents), etc. The problem would be perhaps negligible if we focused only on very productive inventors; but given we are interested also in comets, the issue is crucial.

We therefore opt for an analysis at city level, focusing on the number of *patents* produced by each group of inventors, rather than on the number of *inventors* themselves. Ideally, this would require that, for every time interval, we knew how many comet patents, star patents, and other patents are developed in a given locality. However, the data we use are rather imprecise in the time dimension, for the following reasons: first, we use the year when the patent is

granted,[4] which is generally 2-3 years later the year of application. Second, we do not know how long an inventor has been working on a patent before applying for it. Equally difficult is to time when local knowledge spillovers may have effect - it could be while the source and destination inventors are both working on their respective patents, but it could equally happen a few years after the star has applied (or has been granted) for it. By inspecting the data we found that the median and mean value of the citation lag of patents in the same MSA is four years, and we therefore choose to adopt periods of the same length.[5] This seems a reasonable choice in order to "average out" some of the measurement error in the temporal dimension. We thus identify five time periods of four years each, which are listed in table 1.

We then need to identify those inventors which we define as stars or comets. The task necessarily entails a degree of arbitrariness, which makes our quantification of the number of star and comet patents relatively noisy. However, the estimations we present in the paper (namely in section 4) are robust to measurement errors,[6] and we also check whether our results are consistent with other variable definitions, finding very little variation. We describe these alternative specifications and results in Appendix B. Therefore, although we of course aim for the most precise definition, the reader should not be excessively worried about the exact definition: we just need to define two good proxies of the quantity of star and comet patents in a given city, technological category, and period.

Potentially, we could observe inventors for their whole career, and then classify them as stars or comets according to their propensity to patent. There are, however, two problems, one conceptual and one due to data truncation. First, to the extent that we aim at assessing the effect of productivity spillovers, a definition of stars based on their whole career can be imprecise, as productivity may be highly variable along it. Second, given that our data cover the 1975-99 period, we cannot observe the whole career of the large majority of the inventors in the sample. We therefore adopt a definition that takes into account the productivity of inventors for a shorter period of time, but still long enough to approximate the average productivity of individual inventors in that stage of their career, and to smooth short term disturbances. We follow the same approach for comets as well, in order to avoid including in the category inventors who do not satisfy the requirements in the years immediately before, or after, a given period.

Therefore, for each of the five periods, we define an 8-years long, overlapping *observational window* - they are reported in the third column of table 1. In each period, a patent is defined as the outcome of a "star inventor" if its first author has developed five other patents or more (as first author) in the relative observational window, and it is therefore defined as a star patent. The threshold has been chosen as it approximately limits the top 5% of the inventors'

---

[4]The reason why we use the grant year, rather than the application one, is to avoid the bias given by data truncation. More precisely, using the application year we would automatically exclude all the patents not granted (but applied for) before the 1999, as they are not included in the dataset. This subsample could easily be non-random, e.g. better patents may take longer to be examined, etc.

[5]We restricted the calculation to patent couples with a maximum citation lag of ten years, as longer lags are unlikely to be related to knowledge spillovers. The citation lag is calculated as the difference between the grant year of the citing and cited patents.

[6]The number of star and comet patents are used as dependent and indipendent variables, respectively. In the first case, the measurement error does not affect the consistency of the estimates; in the second case, we rely on 2SLS estimates to obtain consistent coefficients.

Table 1: Period classification

Period	Years	Obs. window
1	1978-1981	1976-1983
2	1982-1985	1980-1987
3	1986-1989	1984-1991
4	1990-1993	1988-1995
5	1994-1997	1992-1999

distribution in term of patents per-capita. Similarly, we define "comet inventors" patent (first) authors who developed less than three patents in the relative observational window, and less then six till that point in time (the latter condition excludes the possibility that a star becomes a comet); the patents they develop are defined as comet patents. As a further restriction, comet patents must not have as assignee a company which is assignee of 50 patents or more in the whole dataset, in order to avoid defining as comets those inventors working for companies where many stars are potentially employed. The treshold has been chosen because 80% of star patents are assigned to an assignee which has more than 50 patents assigned. This restriction is important for our analysis, for two reasons: first, it allows us to better identify local knowledge externalities, disentangling them from co-located increases in productivity due to market mediated workplace contacts. The recent literature has indeed highlighted the risk to overestimate the positive effects of externalities by ignoring the "priced" component of the professional network of inventors, as we discussed in the previous section (e.g. Breschi and Lissoni, 2009; Zucker et al, 1998; and Almeida and Kogut, 1999). Second, our definition of comets entails inventors working for firms for which the primary activity is not the production of patented innovation. Without a patent-firm matched dataset this is hard to detect precisely, but the restriction is our best approximation. Furthermore, in order to focus on patents with a direct market application, a comet patent must be assigned to an US corporation: this leaves out around 10% of comets which are unassigned, or assigned to individuals. These latter restrictions are instead unnecessary for stars, as they are satisfied in the large majority of the cases and, in the few cases in which they are not satisfied, it is likely to be due to spelling errors in the assignee name. A summary of the definition requirements for stars and comets are reported in 2.

Table 2: Definition requirements

Inventor group	Stars	Comets
Number of patents in the relative obs window	$\geqslant 5$	$\leqslant 2$
Total number of patents of the assignee		$\leqslant 50$
Total number of patents granted to the inventor till that point in time		≤ 5
Kind of assignee		US corporation

The analysis is generally limited to the last three periods, as MSA controls are unavailable for period 1 and 2. We define five periods, however, as the first two are used to build the instrumental variables.

Star patents account for the 26% of the total patents granted in the period 1986 -1997, while the corresponding share of comet patents is equal to 11%. On the inventors' side, among all the unique inventors listed in the five periods (534,120), around 5% of them are listed as stars at least once, while for comets the same share is equal to 15%. Looking at single periods, star inventors are 7-9% of the total, while comets are 14-16%. It is worth noticing, therefore, that the majority of patents and inventors do not belong to the two categories. The "star" status appears to be quite persistent across time: around 40% of stars in given period were stars also in the previous period. The share goes down to 15% with a two periods lag. Individual inventors listed as stars cannot become comets in following periods by construction, while a comet can potentially become a star; this, however, happens for only 1% of comet inventors listed in the dataset.

Interesting facts emerge also from the analysis of citation data. Table 3 reports the flows of citations across groups, expressed as a share of the total citations originating from each group. Compared to patents that are neither comets nor stars (third row), comets (first row) are more likely to cite comets, and less likely to cite stars. The opposite is true for stars: they are more likely to cite stars, and less likely to cite comets. The pattern is similar also when looking at citations within technological categories (not shown). We interpret this as further evidence that the stars/comets categorization, although stylized and somehow arbitrary, do identify different groups of patents. On the other hand, we notice that comets do cite stars, although at a smaller rate than other patents; this in turn suggests that comets might benefit from knowledge spillovers from stars. We will explore this hypothesis in depth in the rest of the paper.

Table 3: citations' shares, comets and stars

		Cited		
		Comets	Stars	Other patents
Citing	Comets	16.2	16.8	67.0
	Stars	7.5	34.7	57.8
	Other patents	9.7	19.8	70.5

Citations may also be useful to inspect the average "value" of different categories of patents. Although quite debatable and noisy, the association of number of received citations with the market value of the patents has been convincingly argued (Hall et al, 2001). We use citation data to explore whether patents and comets significantly differ from other patents in this dimension, by regressing the number of received citations on "comet" and "star" dummies, over the whole sample of patents in period 3, 4, and 5. We also include time and technological category dummies, and a variable reporting the number of citations made to control for the heterogeneous propensity to cite among different kinds of patents (within categories and time periods). The dependent variable is de-meaned and standardized, and thus the constant is excluded. We also run the same specification with technological *sub*category dummies and MSA dummies, and excluding the top 5% cited patents. In both cases, we obtain very similar results (reported in Appendix D).

Results - reported in table 4 - show that stars are on average more cited than comets,

Table 4: Regression of citations received

Dep. var.	Citations received (standardized)
Nr. citations made	0.00763***
	(0.00017)
Star patent dummy	0.176***
	(0.0042)
Comet patent dummy	0.0974***
	(0.0051)
Other patent dummy	0.0745***
	(0.0039)
Period F.E.	YES
Tech. cat. F.E.	YES
Observations	590953
R^2	0.12

Heteroskedasticity robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

and comets are more cited than patents which are neither stars nor comets (all the pairwise differences between the three coefficients are statistically significant). A star patent is receiving, on average, 0.87 citations more than "other patents" (0.10 time 8.7, i.e., the difference of the two coefficients multiplied by the standard deviation of the dependent variable, the number of citations received). Comets, on the other side, are receiving just around one fifth of citation more (0.02 time 8.7). Results therefore suggest that star patents have a higher scientific and market value than the average patent. However, the effect is positive also for comet patents: this is important as it confirms that even comet patents have some scientific value (in other words, they are not just useless "garage patents" made for hobby).

3.1 Preliminary evidence on location of stars and comets

In this section, we present some descriptive statistics which i) show how stars and comets are located in different places, and ii) substantiate the validity of stars and comets as good proxies for the output of different innovation processes.

If we look at the distribution of comet, star, and other patents over total employment across MSAs,[7] we can see that there is a sizeable correlation (Table 5, Figure 1), which implies that innovative activity is overall spatially concentrated. When plotting the shares of comets and stars on the total of patents, however, there is a fair degree of dispersion in both the distributions, driven by a long right tail (Figure 2, 3).

---

[7]Counties are grouped into MSAs according to the 1993 definition, based on 1990 Census data. Counties not included into MSAs are also individually included in the sample. The analysis, therefore, covers the whole US territory.

Table 5: Patents by MSAs over total employment, rank correlation

	comets	stars	other patents
comets	1	0.42	0.59
stars	0.42	1	0.61
other patents	0.59	0.61	1

We can go further by looking at patterns of partial correlation with MSAs structural characteristics, setting up a simple panel regression for periods 3-4-5 based on the following equations:

$$Share(Comets)_{it} = \beta_1 X_{it} + \delta_t + \epsilon_{it} \tag{1}$$

$$Share(Stars)_{it} = \beta_2 X_{it} + \delta_t + \epsilon_{it} \tag{2}$$

where i indexes MSAs and t periods, X_{it} is a matrix of MSA-specific coovariates, β_1 and β_2 are vectors of coefficients, and δ_t is a time fixed effect. The aim of these regressions is to assess whether stars and comets show two distinctive location patterns, depending on the industrial structure of cities. The variables included in X, therefore, are a list of simple proxies of the industrial structure of the MSA. In the detail, these variables are the following:

a) the (log of) the total patents in the MSA which are neither stars or comets, in order to control for the size of the patenting sector in the city (we excluded stars and comets to avoid circularity). We included this variable as the absolute size of the patenting sector may impact differently the production of stars and comets.

b) log of total employment (totemp), to control for agglomeration economies and size effects; we expect MSAs with larger employment to produce proportionally more patents, in line with the findings of Carlino et al. (2007), but, again, we do not have any strong a-priori on the association of city size with the different kinds of invention.

c) the share of employment in manufacturing (manuf. share), in order to assess whether comets are associated with specialization in manufacturing. To the extent that comets are linked to production phases through learning-by-doing mechanisms, this variable should also have a positive effect on the number of comets.

d) the Herfindahl diversity index (Herfindahl, calculated as the sum of the squares of the share over the total of employment of 2-digit SIC sectors), as a proxy of the diversity of the economic structure. This variable can have two opposite effects: on one side, the literature has emphasized the positive effect of diversity on innovation due to Jacobian externalities (e.g., Glaeser et al., 1992; Duranton and Puga, 2005). On the other side, we do not exclude that MAR externalities,[8] rather than Jacobian, might be more beneficial for the kind of innovation which underlies the development of comets. In fact, to the extent that comets are the outcome of a

---

[8]MAR (an acronym for Marshall-Arrow-Romer) externalities are those based on within-industry knowledge spillovers, and are associated with an high degree of sectoral specialisation.

"learning-by-doing" innovation process, we may expect them to be more frequently developed where there are within-industry knowledge spillovers, as well as other economy of scale, i.e., in specialized cities.

vi) log of the number of plants with less then 500 employees (n. plants <500 emp.) as these are defined as "small plants" in the US; to the extent that comets represent a proxy for occasional and less codified innovation, we hypothesize that their number is positively affected by the presence of small plants. Conversely, we expect star inventors to work for big companies, thus the number of star patents should be negatively associated with this variable, once controlling for total MSA employment.

The sample is restricted to the last three periods and to all the MSAs or counties where at least 100 patents have been developed in the same interval of time. The equations are estimated by OLS regressions on the pooled samples, with standard errors clustered at MSA level.[9] The results - reported in table 6 - clearly show how the two vectors of coefficient are different (as confirmed by the Hausman test: the null hypothesis of equality of the coefficients of column 1-2, and 3-4, is rejected at 1% confidence level). In particular, comet patents are positively associated with the number of small firms, while the total number of other patents and the Herfindahl index have a negative coefficient (which means that a more diversified city is associated with more comets). Conversely, star patents are positively associated with both the number of other patents and the Herfindahl index, suggesting that star patents are more frequently located in specialized cities.

Our (speculative) interpretation of these results is the following: comet patents are associated with more general innovation activities, and therefore are more likely to be located in innovative hotspots with a diversified economy and many small firms; in such cities the pool of patents is not necessarily large, as innovations may be introduced to the market in other forms. On the other hand, the activity of stars is more strongly associated with formal R&D and patenting, thus it is more frequently located where the pool of patents is large, and the structure of the local economy is specialized and dominated by big companies.

We also look at the association with establishment births, by regressing the latter variable on the (log of the) number of star and comet patents developed in the same MSA, plus some other controls (log of total employment, Herfindahl index, and log of average establishment employment - all lagged by one period to avoid simultaneity bias), for period 4 and 5 (period 3 is dropped due to data restrictions). The sample is composed of the 209 MSAs for which data are available, and the model is estimated by OLS on the pooled sample, with standard errors clustered at MSA level. Again, the results (table 7) show a differentiated pattern for stars and comets: while comets have a significant effect, comparable to the effect of other patents, star patents have a negative coefficient.

We do not claim causality at this stage - many variables are potentially omitted and we cannot exclude a reverse causality bias. Nevertheless, the associations we have analysed support two statements: first, once controlling for the general distribution of patenting activities, comet and star patents are developed in different places; second, star patents seem to have a much weaker connection with the local economy than comet patents. To the extent that the

---

[9] We also estimated a SUR model to account for correlation across errors in the two equations. The significance of the regressors, however, is not affected. Results are reported in Appendix D.

Table 6: Regression of comets/stars shares at MSA level

COEFFICIENT	(1) Comets (share)	(2) Stars (share)	(3) Comets (share)	(4) Stars (share)
Tot. emp. (log)	-0.0237***	0.0116	0.00291	-0.00365
	(0.0057)	(0.011)	(0.0060)	(0.012)
Herfindahl	-0.276**	0.672**	-0.284**	0.677**
	(0.13)	(0.33)	(0.13)	(0.33)
Manuf. share	0.0904*	0.0503	0.0573	0.0694
	(0.046)	(0.090)	(0.045)	(0.090)
N. plant <500 emp. (log)	0.0286***	-0.00404	0.0351***	-0.00776
	(0.0064)	(0.013)	(0.0062)	(0.013)
Other patents (log)			-0.0412***	0.0237***
			(0.0036)	(0.0082)
Period dummies	YES	YES	YES	YES
Observations	1289	1289	1289	1289
R^2	0.11	0.03	0.23	0.04

Heteroskedasticity robust standard errors clustered at MSA level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

former are developed in R&D labs of big companies, while the latter are the by-product of the innovative activity of small firms, the finding is not surprising.

3.2   Why should stars positively affect comets?

Even though we assume comet and star patents are the outcome of substantially different innovation processes, still the activity of stars could generate positive externalities increasing the productivity of comets. We identify four main mechanisms through which the externalities may occur:

a) Informal knowledge spillovers: star inventors and comet inventors develop informal contacts due to residential proximity, which in turn facilitate the activity of the latter (e.g., they may obtain hints on their work).

b) Formal knowledge spillovers: star inventors may transfer their expertise to comet inventors in more formal ways, e.g. during seminars, conferences, and the like.

c) Workplace contacts: (future) comet inventors may have the opportunity to work in an institution where stars are employed, without necessarily becoming stars themselves (they may be employed in different duties, or they may leave the institution at an early stage of their career).

d) Display/attraction effects: the presence of many labs of big companies may attract comets to a locality, as they may expect to enjoy the effects of points a, b, and c.

Although all the mechanisms may, in theory, work also in the opposite direction (from

Table 7: Regression of establishment births at MSA level

COEFFICIENT	(1) Estab. births (log)	(2) Estab. births (log)
Total comets (log)	0.304***	0.151***
	(0.062)	(0.048)
Total stars (log)	-0.119***	-0.0818***
	(0.037)	(0.027)
Total oth. patents (log)	0.487***	0.230***
	(0.072)	(0.054)
Herfindahl Index t-1		-3.297
		(2.30)
Tot. emp. t-1 (log)		0.500***
		(0.058)
Manuf. share t-1		-0.473
		(0.46)
N. plant <500 emp. t-1 (log)		-0.105*
		(0.063)
Constant	5.302***	4.984***
	(0.16)	(0.18)
Period dummies	YES	YES
Observations	418	418
R^2	0.71	0.85

Heteroskedasticity robust standard errors clustered at MSA level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

comets to stars), we expect that the main direction of the knowledge spillovers to go from the star scientist to the "occasional" inventor. However, we understand that this may not be true *a priori* and we consequently design our empirical methodology to be robust to reverse causality.

On the other side, we mentioned earlier that a few recent contributions are downsizing the role of localized knowledge spillovers, either arguing for the weakness of local peer effects (Azoulay et al., 2008; Waldinger, 2009), or for the fading of these effects over time in the light of the "death of distance" hypothesis. Thus, the aforementioned mechanisms - and especially a, b, and c - may also play a negligible role in our context.

We therefore test whether the activity of star inventors leads to higher production of comet patents. Unfortunately, the data do not allow us to disentangle the different mechanisms (e.g., a citation may be output of a, b, or c), thus in the following analysis we will generally test for positive spillovers from stars to comets. The definition and empirical identification of the channels thorough which knowledge spillovers take place is probably one of the most challenging and interesting topics in urban economics research agenda, and we hope that the increasing availability of microgeographic data may lead to some progress in the field.

4   Analysis

In the present section we investigate whether the production of star patents in a city affects the production of comet patents in the same city and period, and try to quantify this effect. We therefore estimate the following model:

$$Comets_{ikt} = \beta \cdot Stars_{ikt} + \gamma X_{it} + \delta_k + \tau_t + \phi_i + \delta\tau_{kt} + \varepsilon_{ikt} \qquad (3)$$

where i, k, and t index MSAs, categories, and periods, respectively; Stars and Comets are the number of patents in the respective group, X is a set of MSA time-variant controls, and δ, τ, ϕ are category, time, and MSA fixed effects. The six technological categories are the following: Chemical (excluding Drugs); Computers and Communications (C&C); Drugs and Medical (D&M); Electrical and Electronics (E&E); Mechanical; and Others.

The unit of observation is the MSA-category pair; the choice is motivated by the assumption that knowledge flows in the patenting activity are mostly contained within the same technological category. This is confirmed by citation data: 80% of citation linkages are bounded within the same category. Furthermore, this allows us to exploit a useful source of variation within MSA and period. The analysis is limited to periods 3-4-5, as MSA controls are not available for previous periods, and the sample is restricted to the MSA-category pairs in which at least 25 patents have been granted in the given period.[10]

We opt for a log-linear specification because the dependent variable is an extended count variable (with a long right tail and skewed to the left), which approximates the normal distribution after the log transformation. The side effect of the log transformation is the loss of the zeros, which, however, are less then 5% of observations. In the following section, we

---

[10]The restriction is made in order to exclude small counties where only a few patents are developed, which are likely to act as outliers. This also bring the advantage of reducing drastically the number of zeros and to speed calculations. Robustness tests show that the sample selection is not affecting the results.

perform some robustness tests on the whole sample based on a Negative Binomial model with the natural count variable and we find compatible results.

We suppose that there are two groups of time-variant variables which may potentially affect the number of comet patents produced in a given city, technological category, and time period.

The first group of variables is specific to the patenting activity, and it includes i) the relative size of the given technological category - as the number of comets may increase because the category as a whole is growing; and ii) the total number of patents in the given city, as the number of comets in a given category may increase because the city patent sector is expanding. Omitting these two variables will introduce an important source of spurious positive correlation between the number of comets and stars, which in turn will lead to an overestimate of the main coefficient of interest.

The second group of variables relate to general city characteristics, and includes a few variables measuring the total employment and the industrial structure of the MSA. This group of variables is motivated by the findings we presented previously, namely the strong association of comet patents with a few specific MSA structural characteristics, and from theoretical insights suggesting that comets are more likely to be associated with small companies and a high share of manufacturing employment. We anticipate, however, that this group of variables is rarely significant in our regressions. This is due to the inclusion of the MSA fixed effects, which absorb most of the effect of variables with small variations across time.

In detail, the variables included in the matrix X are the following:

i) number of other patents (neither stars or comets) in the technological category, over the other patents in the other five categories *(share other patents cat.)*; this variable controls for the relative size of the given technological category, and for idiosyncratic (i.e., specific to the category/city pair) productivity shocks. We expect it to be positively correlated with the number of comets.

ii) total number of patents developed in the MSA - excluding all comets to avoid circularity, and stars of the given category to avoid double counting - as a control for the size of the patenting activity *(tot. MSA patents)* in the whole city. Again, we expect a positive coefficient on this variable.

We then include four MSA-specific variables, as proxies for the industrial structure of the city. These variables are exactly the same as in equations 1 and 2:

iii) Log of total employment (*totemp*), to control for agglomeration economies and size effects.

iv) The share of employment in manufacturing (*manuf. share*).

v) The Herfindahl diversity index (*Herfindahl*, calculated as the sum of the squares of the share over the total of employment of 2-digit SIC sectors), as a proxy of the diversity of the economic structure.

vi) Log of the number of plants with less then 500 employees (*n. plants <500 emp.*).

Finally, we include a number of fixed effects, controlling for technological category and MSA time invariant factors, for time-specific shocks, and for technological category shocks. In a few specification, we include also a MSA-period fixed effect. Potentially, we could also include a MSA-category fixed effect but in this case identification will arise only from within MSA-category pairs variation, which is too limited in the data to give significant results. Standard errors are clustered at the MSA-category pair level (i.e., at every cross-sectional

unit of observation). Alternative estimates based on clustering at the State-year pairwise combination gave almost identical standard errors.

4.1 Instrumental Variable Estimation

Estimates of equation 3 can be inconsistent due to reverse causality or omitted variable biases, especially for the main variable of interest (the number of star patents). We therefore create two different instrumental variables for the number of star patents to deal with the issue. The two instruments share a similar intuition: an exogenous variation in the productivity of star inventors in a given MSA and period may arise from the interaction of two factors: i) an historical presence of inventors working in a given technological category or for given companies in that MSA, and ii) an US-wide increase of productivity of these sectors or companies in the given period. To the extent that the first factor is path-dependent and exhibits some inertia over time, it is exogenous to contemporaneous MSA-specific factors once MSA fixed effects are introduced in the specification. At the same time, we expect the productivity of stars inventors working in the same subcategories or companies (but in different cities) to be correlated, due to sharing a similar competition pressure, regulatory framework, market demand, etc. Therefore, we presuppose that US-wide productivity shift in a given sector or company will translate into MSA-specific productivity shocks in proportion to the number of inventors working in that sector or company in the given MSA.

For example, we assume that the total number of star patents developed in the MSA of New York in the year 1994-97 entails an exogenous component due to the interaction of a) the historical presence in New York of many R&D labs in semiconductor devices, and b) the US-wide growth in (patent) productivity of the semiconductor devices sector in the period 1994-97, relatively to other sectors.

The IV strategy is close in spirit to the approach of Bartik (1991) and Blanchard and Katz (1992), among others, who instrumented regional economic growth interacting the lagged sectoral structure of a region with the contemporaneous national sectoral trend. In what follows, the construction of the instruments is explained in detail.

4.1.1   First instrument

The first instrument is calculated through the following steps:

a)   For each period, we calculate the total number of star inventors active in a given MSA and technological subcategory (patents are classified into 6 categories and 36 subcategories). If an inventor developed patents classified into different subcategories, he/she is assigned corresponding weights summing to one, accordingly to the subcategories' shares. If they have been recorded as resident in several MSAs, the modal one is chosen.

b)   For each period, each subcategory, and each MSA, we calculate the average number of patents produced by star inventors in the whole US, excluding the given MSA.

c)   For each MSA, each period, and each subcategory, we multiply the number of inventors in period *n-2* at point a) by the productivity in the respective technological subcategories in period n calculated in b). Subsequently, we sum the outcome by MSA, period, and technological

category. The result is the instrumental variable for total number of star patents in period 3-4-5, by MSA and category.

Formally, it can be expressed with the following equation:

$$IV1_{ikt} = \Sigma_s(StarsInv_{ikst-2} \cdot AvPat_{ikst}) \tag{4}$$

where i indexes MSAs, t periods, k technological categories, and s technological subcategories within the category k. The first element of the product is calculated at point a), and the second one at point b).

The validity of the IV relies on an assumption of excludability for point a), i.e., once MSA fixed effects and the share of patents in a given category are controlled for, the number of star inventors active in a given MSA/category in period *n-2* (on average ten years before) has no independent effect on the number of comet patents developed in period n in the same MSA/category; and on an assumption of exogeneity for b), i.e., the average productivity in the whole US is exogenous to MSA-specific unobserved factors.

There is, however, a reason of concern about the exogeneity assumption for point b). To the extent that comets in a given MSA are specialized in the same subcategories of stars, the US-wide variation in productivity in a subcategory can be correlated with the error term of equation 3. This in turn will compromise the validity of the instrument. We therefore build a second IV in order to improve the robustness of our estimate.

4.1.2 Second instrument

The second instrument follows a methodology similar to the first one, but the technological subcategories are substituted with the assignees of the patents. The steps are the following:

a) For the first period, we calculate the total number of star inventors active in a given MSA and with a given assignee. In case of star inventors with multiple MSAs or assignees in the same period, the modal one is chosen.

b) For each period, each assignee, and each MSA, we calculate the average number of patents produced by star inventors in that period in the whole US, excluding the given MSA.

c) For each MSA, period, and assignee, we multiply the number of inventors in the first period calculated at point a) by the average number of patents produced by star inventors sharing the same assignee in period t calculated in b). Subsequently, we sum the outcome by MSA, period, and technological category (if an inventor has patented in different categories in the same period, the modal one is chosen). The result is the second instrumental variable for total number of star patents in period t, by MSA and category.

Formally, it can be summarized by the following equation:

$$IV2_{ikt} = \Sigma_a(StarsInv_{ika1} \cdot AvPat_{iat}) \tag{5}$$

where i indexes MSAs, t periods, k technological categories, and a the assignees. In the few cases in which the value of point b was missing (because there were not other stars with the

same assignee in other MSAs), it was replaced with the contemporaneous US-wide average productivity of stars in the same technological category.

The excludability condition is identical to the one for the first instrument, while the exogeneity assumption is similar: given that stars and comets generally have different assignees (the assignee is very often the employer of the inventor, and comets have, by definition, assignees which less than 50 patents assigned in total - while, on average, assignees of stars have 4010 assigned patents) we assume that the average productivity of an assignee in the whole US (calculated excluding the given MSA) has no independent effect on the productivity of comets of that MSA.

5   Results

In table 8 we report mean and standard deviation of the patent variables for the 2113 MSA/category pairs which compose our sample. As it is possible to see, the distribution of the variables in natural form (first two rows) is very skewed. All the count variables (number of patents, number of firms) and total employment enter the regression equations in logarithmic form, thus the coefficients can be interpreted as elasticities. The variables which express continous shares (the share of other patents in the same category, the Herfindahl index, and the share of manufacturing employment) are reported in natural form (thus the coefficients reflects percentage changes in the dependent variable following unit changes in the regressors).

Table 8: Summary statistics of stars and comets

Variable	Obs	Mean	Std. Dev.	Min	Max
comets	2113	27.165	53.46	1	626
stars	2113	68.80	159.71	1	2125
log(comets)	2113	2.38	1.29	0	6.43
log(stars)	2113	3.07	1.46	0	7.66

Results from the OLS estimation are reported in col. 1, 2, and 4 in table 9. The effect of star patents on comets is always positive, but overall quite small: when the MSA fixed effect is included, the coefficient ranges from 0.03 to 0.11. Among the other controls, the share of patents in the category have a positive sign, as expected, although the latter is significant only in the specification without MSA fixed effects (col. 1). The same is true for the small plants variable. The total MSA employment is positive but significant in only one specification, while the Herfindahl index and the manufacturing share are always insignificant. The inclusion of the MSA-period fixed effects reduces the size of the star coefficient, which becomes insignificant (col. 4), and magnify the effect of the share of patents in the category. This is due to the fact that now the only variation left is within-MSA (i.e., across different technological categories) in the same period; which is probably too small to allow us to identify precisely any significant effect of stars (at least with OLS), considering also the strong collinearity of the two explanatory variables included (once other factors are controlled for).

Results from 2SLS regressions are reported in col. 3 and 5 of table 9. For brevity, here we report only the results obtained with the second instrument, as it is assumed to be the

Table 9: regression of comet patents

VARIABLES	(1) comets (log)	(2) comets (log)	(3) comets (log)	(4) comets (log)	(5) comets (log)
	OLS	OLS	IV2	OLS	IV2
stars (log)	0.114***	0.0980***	0.273***	0.0334	0.303***
	(0.0215)	(0.0189)	(0.0680)	(0.0286)	(0.0949)
Share other patents cat.	0.289***	0.437***	0.114	1.242***	0.635***
	(0.0797)	(0.0909)	(0.144)	(0.168)	(0.241)
Tot. MSA patents (log)	0.369***	0.0397	0.00107		
	(0.0362)	(0.0874)	(0.0880)		
Total MSA empl. (log)	0.0755	0.400*	0.384		
	(0.0529)	(0.242)	(0.238)		
Plants <500 emp. (log)	0.411***	0.0444	0.0454		
	(0.0609)	(0.193)	(0.185)		
herfindahl	-1.030	2.676	3.048		
	(2.023)	(3.070)	(3.079)		
Manuf. share	0.401	0.0214	-0.253		
	(0.432)	(0.569)	(0.557)		
Constant	-3.239***	-1.462	-0.981	3.532***	-1.475**
	(0.180)	(1.208)	(1.951)	(0.378)	(0.732)
MSA f.e.	NO	YES	YES	YES	YES
Tech. cat.*Period f.e.	YES	YES	YES	YES	YES
MSA*period f.e.	NO	NO	NO	YES	YES
Observations	2113	2113	2113	2113	2113
R^2	0.764	0.861	0.852	0.834	0.817

Heteroskedasticity robust standard errors clustered at MSA-category level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

most exogenous. In Appendix C we report more specifications using also the first instrument, together with first stage estimates and other diagnostics; all the tests reported there confirm the validity of the IV specification and the strength of the instruments.

Instrumented coefficients are still positive and significant, and now the elasticity of comet to star patents is around 0.3. The value is around three times bigger than OLS estimates (and ten times bigger when the MSA-time fixed effect is included). We explain the downward bias of the OLS as originating from negative selection: it is likely that, in general, patenting activities in a given city and category are specialized in one of the two groups of patents (comets or stars) for unobserved reasons, which in turn creates a negative, spurious association between the number of comet and star patents developed in a given MSA-category pair, thus leading to the downward bias in the OLS estimates. Another plausible explanation for the downward bias is the presence of a measurement error in the star variable: we proxy the intensity of activity of star inventors in a locality with the number of patents they produce, but the measure is clearly noisy, as patents are heterogeneous in quality. To the extent that the measurement error of the instrumental variable is independent from the one in the endogenous variable, IV estimates eliminate the "attenuation bias" of the OLS coefficient. The independence of the two errors is actually plausible as the variables are measured using patents in different localities (in the specific city and in the whole US excluding that city, respectively). We instead rule out that the increase in the coefficients may be due to weak instruments, as the hypothesis is rejected by first-stage results reported in Appendix C.

5.1 Robustness tests

We run a series of robustness tests to check the validity of our estimates. In table 10, we report the estimates of the model reported in equation 3 applying a Negative Binomial count model to different selection of the sample: the OLS one, the OLS one plus the observation with zero comets, the OLS one plus the observation with zero comets and less than 25 patents in the MSA/technological category pair, and all the observation (thus adding also the observation with zero stars; to easy comparability, this is done by applying the logarithmic transformation to the natural variable augmented by one). We opted for a Negative Binomial, rather than a Poisson model, as the dependent variable shows a remarkable degree of overdispersion.

Results show that the coefficient of star patents is substantially unaffected by the different sample selections. Furthermore, is size is almost identical the OLS one. We therefore exclude sample selection biases in our OLS estimations, due to either the exclusion of observations with zero comets or the threshold of 25 patents.

A further robustness test involves the inclusion of spatially lagged variables. Although the empirical literature on patents and knowledge spillovers has argued that urban agglomerations are a good approximation of the relevant spatial decay, we cannot exclude *a priori* that some of the effects we are looking at may go beyond the MSA borders. On the other hand, the exact identification of true spatial effects is complex in this context, as unobserved local factors may, in fact, create spurious evidence of spatial dependence. For instance, two contiguous cities may have similar numbers of comet patents because they share other, unobserved attributes, but failing to recognize that would lead to conclude that the number of comet patents in contiguous cities has a *causal* effect on city comets (this is a classic and well known identification problem

19

Table 10: Negative Binomial count regressions

VARIABLES	(1) comets (count)	(2) comets (count)	(3) comets (count)	(4) comets (count)
Sample	OLS	OLS + 0s	OLS + 0s+ <25 pat.	All
stars (log)†	0.132***	0.129***	0.147***	0.153***
	(0.0167)	(0.0169)	(0.0140)	(0.0135)
Share other patents cat	0.558***	0.524***	0.410***	0.401***
	(0.0830)	(0.0815)	(0.0517)	(0.0460)
Tot. MSA patents (log)	0.129**	0.131**	0.0592	0.0274
	(0.0655)	(0.0667)	(0.0515)	(0.0441)
Tot. MSA patents (log)	0.300*	0.345*	0.683***	0.822***
	(0.172)	(0.176)	(0.163)	(0.145)
Plants <500 emp. (log)	0.158	0.196	-0.0480	-0.149
	(0.136)	(0.140)	(0.117)	(0.0990)
herfindahl	2.842	3.306	2.841	-0.0433
	(2.535)	(2.503)	(2.096)	(1.535)
Manuf. share	-0.164	0.157	-0.306	0.0317
	(0.474)	(0.520)	(0.467)	(0.375)
Constant	-1.408	-2.201	-2.832**	-3.150***
	(1.339)	(1.406)	(1.346)	(0.446)
MSA f.e.	YES	YES	YES	YES
Tech. cat.*Period f.e.	YES	YES	YES	YES
MSA*period f.e.	NO	NO	NO	NO
Observations	2113	2202	4191	7589

Heteroskedasticity robust standard errors clustered at MSA-category level in parentheses
*** p<0.01, ** p<0.05, * p<0.1
†This variable is equal to [log (stars +1)] in the regression of column 4

in spatial economics, and more generally in social sciences, as discussed by Manski, 1999). Nevertheless, totally ignoring spatial effects might be also an important omission. In this section, we apply some standard spatial econometrics tools, in order to check whether our results are robust to the inclusion of spatially lagged variables.

We therefore create a set of spatially lagged variables - namely the number of stars, comets, and other patents - calculated by weighting neighbouring observations - within a radius of 300 miles - by the inverse of their distance. Results are reported in 11. The inclusion of the spatial variables leaves the other coefficients almost unaffected, while the spatially lagged variables have generally significant coefficients, especially the "other patents" one. Including the spatial lag of the comets makes OLS estimations inconsistent as a spatial lag of the dependent variable is endogenous by construction (Anselin, 1988). Therefore, we opt for an IV estimation, instrumenting both the endogenous variables, i.e., the number of star patents and the spatial lag of comets. Regarding the choice of the instrument for the latter variable, a popular option in spatial econometrics literature is the spatial lag of one or a few independent variables, as long as one assumes that they do not have any independent effect on the dependent variable. However, in this case we have a better candidate promptly available, i.e, the spatial lag of the instrument. The fourth and fifth columns of 11 therefore report the results of an IV regression where stars and the spatial lag of comets are the endogenous variables, and the second IV and its spatial lag are the instruments. The stars coefficient is extremely similar to the previous IV regressions; the lagged comets have a sizeable coefficient when they are the only spatially lagged variable included in the specification (col. 4), although it is barely significant. Once the other spatial variables are included, it becomes insignificant; similarly, also the other spatial lags are not statistically different from zero. One possible reason for that can be the high correlation among these three variables, which may introduce problem of collinearity; another source of concern can be the weakness of the instrument for the lagged comets once the other lagged variables are included, although standard first stage diagnostic seems to exclude the problem.[11] However, given that our main concern is to assess whether omitted spatially lagged variables may affect our results, rather than obtaining precise point estimates for these variables, for simplicity we decide not to complicate the specification any further; for the same reason, we omit the calculation of a spatial error model (robust to spatial correlation in the error term), as we believe that the large number of fixed effects included in the specifications, as well as the clustered structure of the estimated standard errors, make unlikely this kind of problem - which, however, would affect only the efficiency, and not the consistency, of our estimates. We therefore conclude that we cannot reject the presence of spatial effects in the context under analysis, but, at the same time, their omission is not affecting our main results.

Finally, we run two other robustness tests, which are:

i) the exclusion of the sixth category, which includes all the patents not classifiable under the other five categories;

ii) allowing for different effects of stars in each of the three time periods.

In the first case, results are unaffected. In the second case, coefficients are not significantly

---

[11]The Kleibergen-Paap rk Wald F statistic for regression of col. 5 is equal to 5.86, which corresponds to a bias lower than 15% of the coefficient according to Stock-Yogo critical values. The same statistic for the regression of column 4 is equal to 45.947.

Table 11: Regressions with spatially lagged variables

VARIABLES	(1) comets (log)	(2) comets (log)	(3) comets (log)	(4) comets (log)	(5) comets (log)
	OLS	OLS	OLS	IV	IV
stars (log)	0.0877***	0.0856***	0.0854***	0.252***	0.251***
	(0.0190)	(0.0189)	(0.0190)	(0.0696)	(0.0695)
Share other patents cat.	0.432***	0.421***	0.421***	0.127	0.137
	(0.0906)	(0.0901)	(0.0902)	(0.141)	(0.142)
Tot. MSA patents (log)	0.0363	0.0368	0.0362	0.00594	0.00415
	(0.0882)	(0.0883)	(0.0882)	(0.0874)	(0.0884)
sp. lag stars (log)	0.119***		0.0156		0.0176
	(0.0305)		(0.0512)		(0.0536)
sp. lag oth. pat. (log)		0.193***	0.175**		0.297
		(0.0452)	(0.0769)		(0.231)
sp. lag comets (log)				0.141*	-0.248
				(0.0734)	(0.333)
Total MSA empl. (log)	0.369	0.387	0.382	0.355	0.460*
	(0.245)	(0.243)	(0.243)	(0.237)	(0.254)
Plants <500 emp. (log)	0.0499	0.0513	0.0513	0.0519	0.0519
	(0.188)	(0.187)	(0.187)	(0.178)	(0.180)
herfindahl	2.182	1.952	1.963	2.513	2.464
	(3.057)	(3.057)	(3.056)	(3.045)	(3.081)
Manuf. share	0.0344	0.0342	0.0344	-0.228	-0.205
	(0.563)	(0.560)	(0.560)	(0.547)	(0.552)
Constant	-1.427	-1.734	-1.690	-1.460***	-2.233***
	(1.241)	(1.233)	(1.232)	(0.529)	(0.776)
MSA f.e.	YES	YES	YES	YES	YES
Tech. cat.*Period f.e.	YES	YES	YES	YES	YES
MSA*period f.e.	NO	NO	NO	NO	NO
Observations	2096	2096	2096	2096	2096
R^2	0.863	0.864	0.864	0.856	0.854

Heteroskedasticity robust standard errors clustered at MSA level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

different across periods, although the last one is generally slightly bigger. The hypothesis of a fading effect over time is therefore rejected.

Given the close similarity of these results with the ones already presented, they are not reported for brevity (they are however available from the author upon request).

6   Conclusions

This paper builds on the analysis of a very peculiar aspect of the patent data, i.e., the skewness of the distribution of patents among inventors. We therefore identify two illustrative categories of patents - stars and comets - based on the average productivity of their inventors. Two main conclusions emerge from the analysis: first, once controlling for the overall concentration of patenting activity, stars and comets are associated with cities with different structural characteristics. In particular, comets are associated with a diversified economic structure, concentration of small plants, and establishment births; while stars are more likely to be found in metropolitan areas with a large pool of patents and a specialized economic structure. Second, we show that the activity of star inventors is beneficial to the activity of comet inventors: in our preferred specifications, we find that the elasticity of comet patents to star patents is approximately equal to 0.3, which means that, on average, a 10% increase in the number of star patents leads approximately to a 3% increase in the number of comets.

More research is needed to expand both the conclusions we reached, in order to better identify the characteristics of cities associated with concentrations of the two categories of inventors; regarding the second, and to investigate the channels through which the spillovers take place. Also, the availability of a patent-firm matched dataset will allow i) to check our speculative hypothesis that comets are more likely to be employed by small firms, while stars work for the R&D labs of big companies; and ii) to assess more in depth the impact of the different categories of patents on the local economy.

The policy recommendations are not one-way. On one side, given the strong effect of stars on the productivity of comets, the attraction of stars to a city may be highly beneficial to the local economic environment: stars will benefit comets, which in turn will foster the birth of new plants, the innovation output of small businesses, and the generation of new employment. Thus, even though R&D labs of big corporations may have only a limited direct effect on the local economy, as most the of employment and value added is located elsewhere, they may be highly beneficial in the light of the aforementioned indirect effect.

On the other side, we know that stars and comets are concentrating in different places, which might imply that attracting stars where comets are might not be a successful policy, as stars in "comets' places" may be less productive. In other words, the same location for comets and stars will end up to be sub-optimal for (at least) one of the two categories. Therefore, interfering on the location choice of stars (or comets) in order to increase the spatial proximity may introduce perverse incentives and lead to much weaker effect than expected.

References

Acs ZJ, DB Audretsch and MP Feldman, 1992, Real Effects of Academic Research: Comment, American Economic Review, 82:1

Acs ZJ, DB Audretsch, 1990, Innovation and small firms, the MIT Press Real Effects of Academic Research: Comment

Almeida, P., and B. Kogut. 1999. Localization of knowledge and the mobility of engineers in regional networks. Management Science 45

Audretsch DB, 2002, The dynamic role of small firms: evidence from the US, Small Business Economics, 18:1-3

Anselin L., 1988, Spatial Econometrics: Methods and Models, Dordrecht, Kluwer Academic Publishers.

Azoulay, P., J. S. Graff Zivin, and J. Wang, 2008, Superstar Extinction, NBER Working Paper No. 14577

Balasubramanian N, J Sivadasan, 2008, What Happens when Firms Patent? New Evidence from US Economic Census Data, working paper

Bartik T.J., 1991. Who Benefits from State and Local Economic Development Policies? W.E. Upjohn Institute

Bessen J, 2008, The value of U.S. patents by owner and patent characteristics, Research Policy 37:5

Breschi, S. and Lissoni, F.,2001, Knowledge spillovers and local innovation systems: a critical survey. Industrial and Corporate Change, 10: 975–1005.

Breschi S., and F. Lissoni, 2009, Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows,Journal of Economic Geography 9 pp. 439–468

Cairncross F., 1997, The death of distance: how the communications revolution will change our lives. Harvard Business School Press, Boston.

Carlino G.A., S. Chatterjee, R.M. Hunt, 2007, Urban density and the rate of invention, Journal of Urban Economics, 61: 389–419

Coyle, D., 1999, The Weightless World: Strategies for Managing the Digital Economy, MIT Press

Duranton G., D. Puga, 2005, From sectoral to functional urban specialisation, Journal of Urban Economics, 57:2, 343-370

Ellison, G.D., E.L. Glaeser, and W.R. Kerr, forthcoming. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." The American Economic Review

Friedman, T., 2005, The World is Flat, New York: Farrar, Strauss and Giroux

Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., and Shleifer, A., 1992. Growth in cities, Journal of Political Economy 100, 1126-1152.

Greenstone M, R Hornbeck, E Moretti, 2008, Identifying agglomeration spillovers: evidence from million dollar plants, NBER Working Paper n. W13833

Griffith R., L. Sokbae, J. Van Reenen, 2007, Is Distance Dying at Last? Falling Home Bias in Fixed Effects Models of Patent Citations, NBER Working Paper 13338

Griliches Z, 1990, Patent Statistics as Economic Indicators: A Survey, Journal of Economic Literature, 28: 4

Hall, B.H., A. B. Jaffe, M. Trajtenberg, 2001, The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools, NBER Working Paper 8498

Ioannides Y., H.G. Overman, E. Rossi-Hansberg, and K. Schmidheiny, 2008, The effect of information and communication technologies on urban structure, Economic Policy, 23, pp 201-242

Jaffe A.B., 1989, Real Effects of Academic Research, The American Economic Review, 79:5, pp. 957-970

Jaffe A.B., M. Trajtenberg, 2005, Patents, Citations, and Innovations: A Window on the Knowledge Economy, The MIT Press

Jaffe A.B., M. Trajtenberg, and R. Henderson, 1993, "Geographic localization of knowledge spillovers as evidenced by patent citations", Quarterly Journal of Economics 10, 577-598

Manski, C.F., 1999. Identification problems in the social sciences. Harvard University Press

Peri G, 2005, Determinants of Knowledge Flows and Their Effect on Innovation,The Review of Economics and Statistics, vol. 87, issue 2, pages 308-322

Silverberg G, B Verspagen, 2007, The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance, Journal of Econometrics, 139:2

Quah, D., 1999, The Weightless Economy in Economic Development, CEP Discussion Paper No. 417

Trajtenberg M., G. Shiff, R. Melamed, 2006, The "Names Game": Harnessing Inventors' Patent Data for Economic Research, NBER Working Paper No. 12479

Waldinger F., 2009, Peer Effects in Science - Evidence from the Dismissal of Scientists in Nazi Germany, CEP Discussion Paper No 910

Zucker, L G. and MR Darby, 2007, Star Scientists, Innovation and Regional and National Immigration NBER Working Paper Series no. 13547

Zucker, L. G., Darby, M. R., Armstrong, J.,1998, Geographically localized knowledge:spillovers or markets? Economic Inquiry, 36: 65–86.

Appendix A: Data

Patent data come from the United States Patent and Trademark Office (USTPO) database as processed by the National Bureau of Economic Research (NBER), described in Hall et al, 2001. To the original dataset we added the inventors' unique identifier developed by Trajtenberg et al (2006) and the standardized assignee name available in the Prof. Bronwyn H. Hall website.[12] We are aware that the latter is not always reliable as i) the complex ownership structure of companies may imply that differently named assignees correspond, in fact, to the same company, and ii) the same company name can be spelled in different ways (and the standardization routines cannot completely solve the problem).

We eliminated patents granted to inventors residing outside US and geolocated all the cities of residence of inventors through the ArcGis geolocator tool (based on the 2000 gazzetter of US places from US Census) and the Yahoo! Maps Web Services. In case more authors are listed for the same patents and they live in different cities, the city of residence of the first author is chosen; this is a standard procedure in patent literature, and Carlino et al. (2007) show that the approximation is substantially innocuous. The geocoding operation was successful for 1,161,650 patents, which correspond to 97% of the database. We then assigned cities to counties using the ArcGis spatial join tool, and subsequently counties into MSAs (1993 definition). Those counties which are not included in the MSAs dataset are reported singularly - the geographical units are therefore a mix of counties and MSAs (for simplicity in the paper we do not distinguish between the two entities and call all the spatial units "MSAs"). This is a sensible choice to the extent that small counties not included in the MSAs definition do not exhibit strong commuting flows and are therefore self-contained functional entities. To our knowledge, is the first time that patent data are geocoded (almost) entirely, without ignoring small counties.

Other County and MSA specific variables for employment and industrial structure are calculated from the County Business Pattern dataset, while data on establishment births come from Company Statistics. Both the databases are freely available from the US Census webpage.

---

[12]http://elsa.berkeley.edu/~bhhall/

Appendix B: Alternative definitions of comets and stars

In this appendix we present various alternative definitions of the patent variables, and we briefly discuss how the main results of the paper are affected.

The first concern about the definition we adopted regards the choice of considering only the first author of the patent. Looking at table 12, we can see that authors whose surname is starting with one of the first letters of the alphabet are only slightly more likely to be reported as first author, as compared to second or third authors. However, as a further robustness test, we followed a different procedure, defining a patent as a "star patent" if at least one inventor satisfies the requirements listed in section 3, and as a "comet patent" if all the inventors satisfy the relative requirements. The new variables are highly correlated with the single-author ones (99% pairwise, and 98% partial correlation when including also the total number of patents in the same MSA and category), and lead to extremely similar results: coefficients are only sligthly (20-30%) smaller (table 13). Therefore, to the extent that the first author is generally the project leader, defining comet and star patents based only on her/him probably increase the precision of the estimates.

We then build three other definitions of the comet variable. They are the following:

1) Standard definition (described in Section 3) but including patents assigned to all the assignee types (not only to US corporations), or not assigned.

2) Same as in 1, but excluding not assigned patents.

3) Same as in section 3, but relaxing the constraint on the maximum number of 50 patents for assignee.

We then calculated the results of the specification 3 with both OLS and IV (reported in table 14 and 15, respectively) and checked whether the results were affected. In the first case, the coefficients are reduced by around 50%, although they keep their significance. This is explained by the inclusion in the comet group of many patents not assigned or assigned to individuals, which are likely to bear less scientific and market value than other patents, and therefore they should benefit less from spillovers from stars (assuming that if the quality of patents is lower, there are less points of contact with excellent patents). The second definition gives coefficients that are around 20% smaller than the adopted definition; the difference is therefore small and due to similar reasons. The third comet variable gives a coefficient around twice as higher in the OLS, and similar to the one obtained with the standard comet variable in the IV specification. Again, this is not surprising, as comets defined in this way are more likely to work for the same employers of stars, which in turn leaves room for spurious positive correlation which pushes OLS estimates upward (which reduces the downward bias in the specific case).

To conclude, results are always qualitatively similar to the ones obtained with the standard definition of comets, and none of the (small) quantitative differences is unexpected.

Table 12: Inventors' surname initial and patent authors' sequence

Initial	first author		second author		third author	
	Freq.	Percent	Freq.	Percent	Freq.	Percent
A	42,942	3.58	14,683	2.69	5,697	2.45
B	115,093	9.6	43,904	8.03	16,242	6.99
C	86,866	7.25	36,552	6.69	13,911	5.99
D	57,310	4.78	24,773	4.53	9,614	4.14
E	23,823	1.99	10,272	1.88	3,941	1.7
F	45,165	3.77	20,096	3.68	7,891	3.4
G	63,038	5.26	28,161	5.15	11,123	4.79
H	85,751	7.16	39,656	7.26	16,097	6.93
I	5,838	0.49	2,606	0.48	1,087	0.47
J	28,038	2.34	12,922	2.36	5,387	2.32
K	63,828	5.33	30,438	5.57	12,917	5.56
L	63,088	5.26	30,152	5.52	13,138	5.65
M	98,633	8.23	47,858	8.76	20,944	9.01
N	24,425	2.04	11,712	2.14	5,365	2.31
O	16,422	1.37	7,974	1.46	3,541	1.52
P	55,056	4.59	27,231	4.98	12,197	5.25
Q	1,854	0.15	970	0.18	386	0.17
R	55,828	4.66	26,368	4.82	12,045	5.18
S	124,636	10.4	60,864	11.14	27,666	11.9
T	37,138	3.1	18,570	3.4	8,690	3.74
U	3,582	0.3	1,769	0.32	928	0.4
V	17,480	1.46	8,525	1.56	4,342	1.87
W	63,419	5.29	30,428	5.57	14,356	6.18
X	304	0.03	247	0.05	120	0.05
Y	9,540	0.8	5,055	0.92	2,481	1.07
Z	9,282	0.77	4,735	0.87	2,297	0.99
Total	**1,198,379**	**100**	**546,521**	**100**	**232,403**	**100**

Table 13: regression of comet patents, multi-author

VARIABLES	(1) Comets (log) OLS	(2) Comets (log) OLS	(3) Comets (log) IV2
stars (log)	0.0795***	0.0830***	0.228***
	(0.0205)	(0.0181)	(0.0645)
Share other patents cat.	0.355***	0.543***	0.249*
	(0.0950)	(0.0969)	(0.148)
Tot. MSA patents (log)	0.410***	0.137	0.0954
	(0.0372)	(0.0936)	(0.0913)
Total MSA empl. (log)	0.120**	0.358	0.368
	(0.0536)	(0.243)	(0.239)
Plants <500 emp. (log)	0.352***	0.0712	0.0574
	(0.0620)	(0.178)	(0.174)
herfindahl	-2.151	4.111	5.577*
	(1.931)	(3.057)	(3.037)
Manuf. share	0.516	-0.178	-0.521
	(0.442)	(0.596)	(0.598)
Constant	-3.431***	-2.036	-1.875***
	(0.194)	(1.273)	(0.558)
MSA f.e.	YES	YES	YES
Tech. cat.*Period f.e.	YES	YES	YES
MSA*period f.e.	NO	NO	NO
Observations	2088	2088	2088
R^2	0.763	0.864	0.857

Heteroskedasticity robust standard errors clustered at MSA level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: regression of comet patents, alternative definitions, OLS

VARIABLES	(1) comets def 1 (log)	(2) comets def 2 (log)	(3) comets def 3 (log)
	OLS	OLS	OLS
stars (log)	0.0488***	0.0821***	0.153***
	(0.0137)	(0.0190)	(0.0147)
Share other patents cat.	0.465***	0.566***	1.164***
	(0.0664)	(0.0933)	(0.102)
Tot. MSA patents (log)	0.210***	0.163**	0.344***
	(0.0619)	(0.0824)	(0.0458)
Total MSA empl. (log)	-0.0964	-0.0614	0.429***
	(0.163)	(0.232)	(0.123)
Plants <500 emp. (log)	0.194	0.291	0.0382
	(0.124)	(0.185)	(0.127)
herfindahl	3.460	2.867	1.735
	(2.232)	(3.000)	(1.868)
Manuf. share	-0.950*	-0.777	-0.409
	(0.517)	(0.747)	(0.401)
Constant	0.673	-0.265	-2.039***
	(0.849)	(1.120)	(0.678)
MSA f.e.	YES	YES	YES
Tech. cat.*Period f.e.	YES	YES	YES
MSA*period f.e.	NO	NO	NO
Observations	2113	2113	2113
R^2	0.764	0.861	0.852

Heteroskedasticity robust standard errors clustered at MSA level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15: regression of comet patents, alternative definitions, IV

VARIABLES	(1) comets def 1 (log)	(2) comets def 2 (log)	(3) comets def 3 (log)
	IV2	IV2	IV2
stars (log)	0.156***	0.275***	0.361***
	(0.0495)	(0.0718)	(0.0427)
Share other patents cat.	0.275***	0.218	0.806***
	(0.0983)	(0.143)	(0.101)
Tot. MSA patents (log)	0.191***	0.126	0.306***
	(0.0616)	(0.0817)	(0.0505)
Total MSA empl. (log)	-0.114	-0.0843	0.386***
	(0.160)	(0.225)	(0.135)
Plants <500 emp. (log)	0.189	0.292	0.0435
	(0.118)	(0.181)	(0.116)
herfindahl	3.292	3.125	1.492
	(2.078)	(2.997)	(1.767)
Manuf. share	-1.015**	-1.045	-0.566
	(0.493)	(0.732)	(0.365)
Constant	2.498*	-1.049**	-2.886**
	(1.365)	(0.488)	(1.239)
MSA f.e.	YES	YES	YES
Tech. cat.*Period f.e.	YES	YES	YES
MSA*period f.e.	NO	NO	NO
Observations	2113	2113	2113
R^2	0.764	0.861	0.852

Heteroskedasticity robust standard errors clustered at MSA level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

31

Appendix C: IV estimation diagnostics

In this appendix we present the results from alternative specifications relative to the IV estimations, following the recommendations reported in Angrist and Pischke (2008, p. 212). We report the results of the first stage regressions, in order to test the strength of the excluded instruments; subsequently, we test the exogeneity of the instruments, by comparing our main results presented in table 9 with the overidentified specifications (thus including also the first instrument) estimated through 2-stages least squares (2SLS) and Limited Information Maximum Likelihood models (LIML).

In table 16 we report the results from the first stage regressions; columns 1-2 are the specifications correspondent to our preferred IV estimations reported in columns 3 and 5 of table 9, respectively. In both the specifications, the coefficients on the instrument are highly significant, and the F statistics is well above the rule-of-thumb value of 10. In columns 3 and 4 we calculate two other first-stage regressions which do not have any direct correspondence to any of the 2SLS estimates we presented in the paper, but are meant to be a further test on the strength and exogeneity of the instrument: specifically, we added a MSA-category fixed effect (which we did not include in the main model), which absorbs every time-invariant component specific to a given MSA-category pair. As it is possible to see, the coefficient is less precise (but this is not surprising, given the little variability left) but it is still significant, and its size is even bigger than in columns 1 and 2. Column 5, instead, reports the result from the first-stage regression including both the excluded instruments. Again, the F statistic confirms the strength of the instruments. We also calculated a F-test on null hypothesis that the instrumental variables are jointly equal to zero, which is clearly rejected (F=46.21).

In table 17 we report some diagnostics on the exogeneity of the instruments. Specifically, we estimate the overidentified regression (thus including also the first instrument) by means of Limited Information Maximum Likelihood models (LIML) as well as Two Stages Least Squares (2SLS). As Angrist and Pischke (2008) argue, LIML models are less precise but also less biased, thus sizeable differences in the point estimates with 2SLS equivalent specifications should be a reason of concern. However, in this case the coefficient values are very close either among them, and to the ones estimated in the main model of table 9. Therefore, we can conclude that the validity of the IV estimation is not a concern in our case.

Table 16: First stage regression

VARIABLES	(1) stars (log)	(2) stars (log)	(3) stars (log)	(4) stars (log)	(5) stars (log)
IV1 (logs)					0.102***
					(0.019)
IV2 (logs)	0.268***	0.243***	0.362*	0.528***	0.192***
	(0.0327)	(0.0291)	(0.215)	(0.149)	(0.032)
Share other patents cat.	1.352***	1.656***	1.038***	0.0288	1.243***
	(0.216)	(0.127)	(0.261)	(0.161)	(0.21)
Tot. MSA patents (log)	0.287**			0.812***	0.298**
	(0.115)			(0.106)	(0.12)
Total MSA empl. (log)	-0.0352			-0.223	-0.153
	(0.311)			(0.265)	(0.32)
Plants <500 emp. (log)	0.0190			-0.119	-0.0315
	(0.247)			(0.210)	(0.25)
herfindahl	-2.627			-1.747	-3.691
	(4.135)			(3.885)	(4.36)
Manuf. share	1.597*			1.125	1.588*
	(0.920)			(0.747)	(0.90)
Constant	0.298	6.376***	1.748***	-1.366	1.091
	(1.758)	(0.260)	(0.506)	(1.520)	(1.82)
MSA f.e.	YES	YES	YES	YES	YES
Tech. cat.*Period f.e.	YES	YES	YES	YES	YES
MSA*Period f.e.	NO	YES	YES	NO	NO
MSA*cat. f.e.	NO	NO	YES	YES	NO
Observations	2113	2113	2113	2113	2113
R^2	0.790	0.831	0.664	0.642	0.79
F-stat	27.04	24.60	N.A.	532.08	29.06

Heteroskedasticity robust standard errors clustered at MSA level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 17: IV, overidentified regressions, 2SLS and LIML

VARIABLES	(1) Comets (log)	(2) Comets (log)	(3) Comets (log)	(4) Comets (log)
	LIML	LIML	2SLS	2SLS
Stars (log)	0.310***	0.322***	0.308***	0.322***
	(0.0606)	(0.0776)	(0.0600)	(0.0775)
Share other patents cat.	0.0456	0.592***	0.0495	0.593***
	(0.133)	(0.216)	(0.132)	(0.215)
Tot. MSA patents (log)	-0.00715		-0.00669	
	(0.0897)		(0.0896)	
Total MSA empl. (log)	0.381		0.381	
	(0.242)		(0.242)	
Plants <500 emp. (log)	0.0456		0.0456	
	(0.188)		(0.188)	
herfindahl	3.127		3.123	
	(3.148)		(3.144)	
Manuf. share	-0.311		-0.308	
	(0.569)		(0.568)	
Constant	-1.113	-1.358**	-1.106	-1.361**
	(1.985)	(0.679)	(1.983)	(0.679)
Observations	2113	2113	2113	2113
R^2	0.848	0.814	0.848	0.814

Heteroskedasticity robust standard errors clustered at MSA level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix D: Ancillary tables and results

Table 18: citations' shares, comets and stars, within tech. category

		Cited		
		Comets	Stars	Other patents
	Comets	16.5	16.8	66.7
Citing	Stars	7.3	35.6	57.5
	Other patents	9.7	19.8	70.5

Table 19: Regression of citations received with tech. subcat. fixed effects

Dep. var.	Citations received
	(standardized)
Nr. citations made	0.00244***
	(0.00014)
Star patent dummy	-0.125***
	(0.012)
Comet patent dummy	-0.183***
	(0.012)
Other patent dummy	-0.206***
	(0.012)
Period F.E.	YES
Tech. subcat. F.E.	YES
Observations	590953
R^2	0.06

Heteroskedasticity robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 20: Regression of citations received excluding the top 5 per cent cited patents

Dep. var.	Citations received (standardized)
Nr. citations made	0.00817***
	(0.00018)
Star patent dummy	0.243***
	(0.0045)
Comet patent dummy	0.151***
	(0.0053)
Other patent dummy	0.135***
	(0.0041)
Period F.E.	YES
Tech. subcat. F.E.	YES
Observations	564339
R^2	0.13

Heteroskedasticity robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 21: Regression of comets/stars shares at MSA level, SUR

COEFFICIENT	(1) Comets (share)	(2) Stars (share)	(3) Comets (share)	(4) Stars (share)
Tot. emp. (log)	-0.0237***	0.0116	0.00291	-0.00365
	(0.0049)	(0.0088)	(0.0049)	(0.0094)
Herfindahl	-0.276**	0.672***	-0.284***	0.677***
	(0.12)	(0.21)	(0.11)	(0.21)
Manuf. share	0.0904**	0.0503	0.0573*	0.0694
	(0.036)	(0.064)	(0.033)	(0.064)
N. plant <500 emp. ((log)	0.0286***	-0.00404	0.0351***	-0.00776
	(0.0061)	(0.011)	(0.0056)	(0.011)
Other patents (log)			-0.0412***	0.0237***
			(0.0029)	(0.0055)
Period dummies	YES	YES	YES	YES
Observations	1289	1289	1289	1289
R^2	0.11	0.03	0.23	0.04

Heteroskedasticity robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

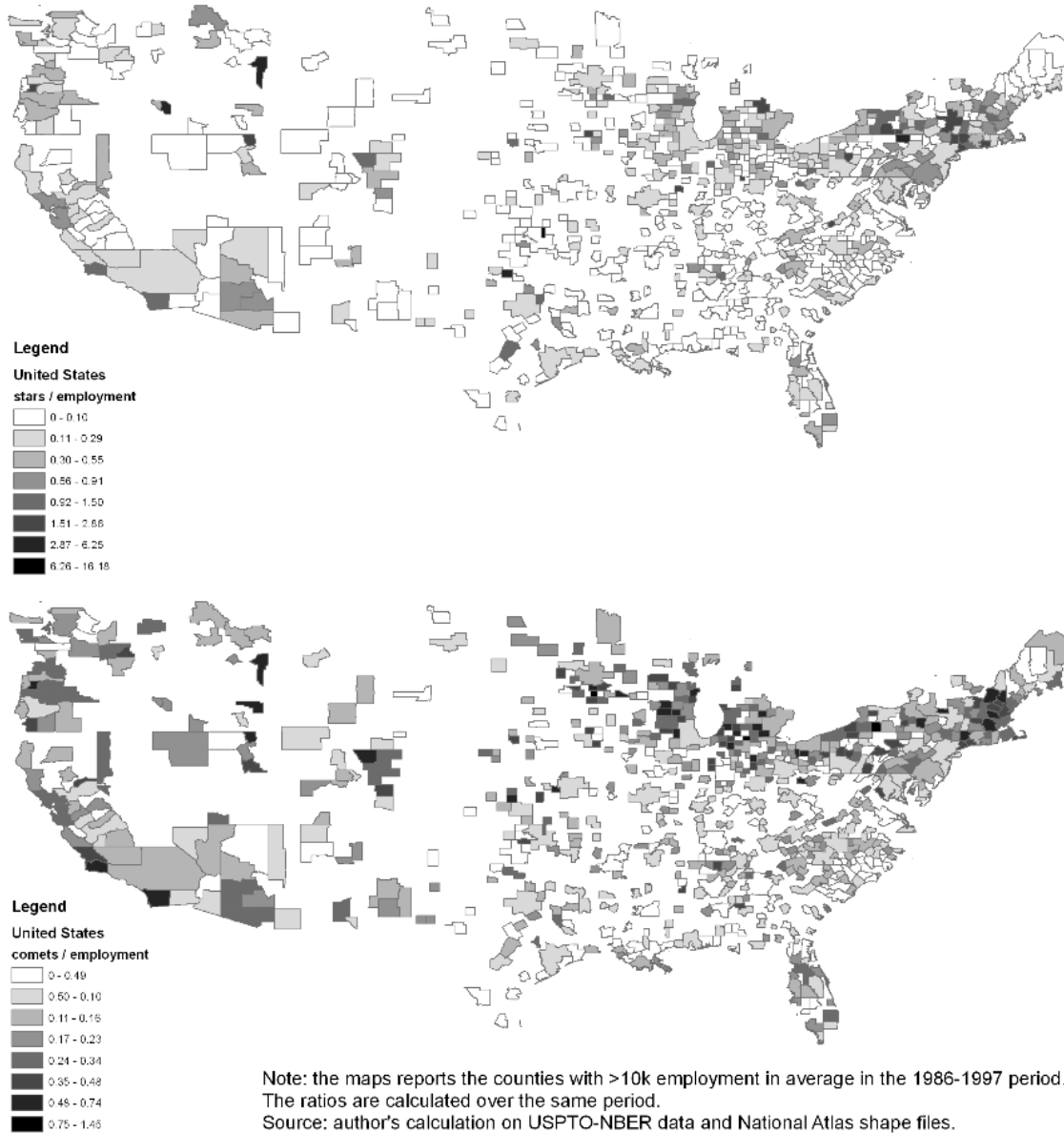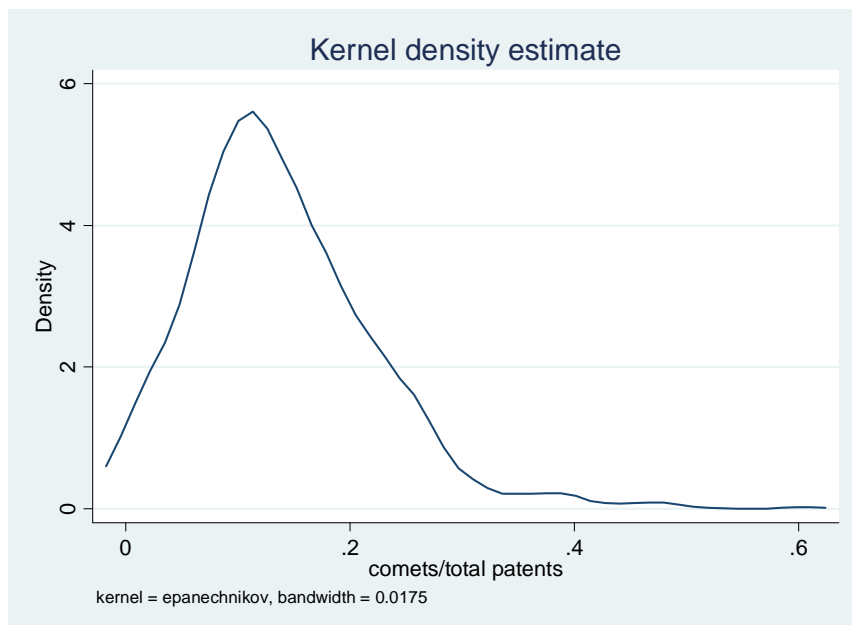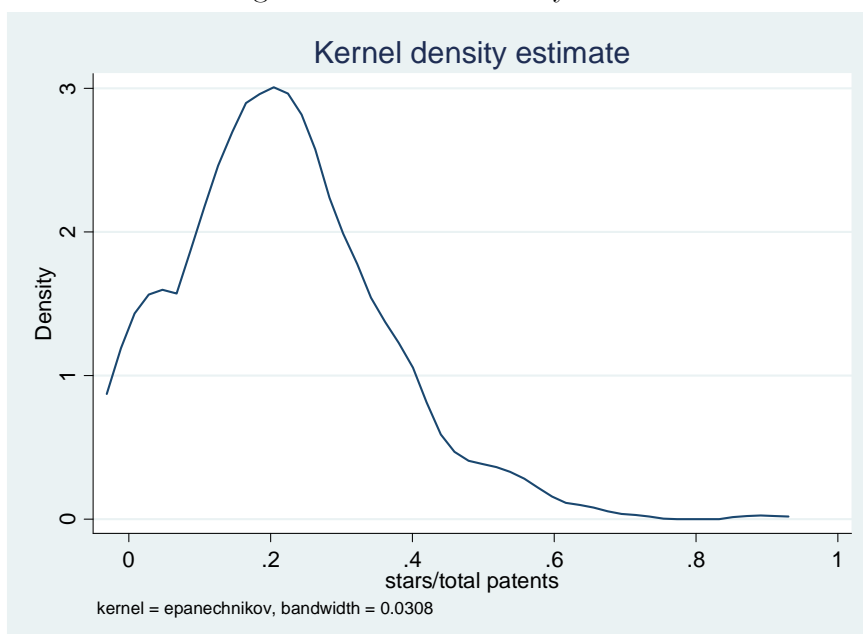Figure 1: Star and comet patents over employment, MSAs

Legend
United States
stars / employment
- 0 - 0.10
- 0.11 - 0.29
- 0.30 - 0.55
- 0.56 - 0.91
- 0.92 - 1.50
- 1.51 - 2.86
- 2.87 - 6.25
- 6.26 - 16.18

Legend
United States
comets / employment
- 0 - 0.49
- 0.50 - 0.10
- 0.11 - 0.16
- 0.17 - 0.23
- 0.24 - 0.34
- 0.35 - 0.48
- 0.48 - 0.74
- 0.75 - 1.46

Note: the maps reports the counties with >10k employment in average in the 1986-1997 period. The ratios are calculated over the same period.
Source: author's calculation on USPTO-NBER data and National Atlas shape files.

Figure 2: Share of comets by MSAs

Kernel density estimate

kernel = epanechnikov, bandwidth = 0.0175

Figure 3: Share of stars by MSAs

Kernel density estimate

kernel = epanechnikov, bandwidth = 0.0308

38