BANCA D'ITALIA

EUROSISTEMA

Temi di Discussione

(Working Papers)

Stars and comets: an exploration of the patent universe

by Carlo Menon

January 2011

Number 784

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

Stars and comets: an exploration of the patent universe

by Carlo Menon

Number 784 - January 2011

STARS AND COMETS: AN EXPLORATION OF THE PATENT UNIVERSE

by Carlo Menon*

Abstract

The analysis of patent and citation data has become a popular source of evidence on localized knowledge spillovers and innovation. Nevertheless, one aspect has been overlooked: the patent distribution across inventors is extremely skewed, as many inventors -- the comets -- register one or few patents, while a small number of inventors -- the stars -- register many patents. This raises a number of questions relating to the geography of innovation: do different categories of inventors interact with the local economic environment in the same way? Are they equally distributed over space or do they tend to concentrate? Is spatial proximity beneficial for their activity? Using a rich database on US inventors, we provide evidence suggesting that the two categories of patents are associated with different kinds of cities. We then test whether the activity of stars is beneficial for local comets, finding that a 10% increase in the number of patents authored by star inventors leads to a 3% increase in the number of patents developed by comet inventors.

JEL Classification: R10, 031.
Keywords: localized knowledge spillovers, patents, innovation.

Contents

* Bank of Italy, Economic Research Unit, Roma Branch.

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 few patents, while a small number of inventors register many patents. This, apart from being an interesting fact per se, raises a number of questions relating to the geography of innovation: do different categories of inventors interact with the local economic environment in the same way? Are they equally distributed over space or do they tend to concentrate? Is spatial proximity beneficial for their activity?

Innovations developed by inventors at the opposite extremes of the distribution are unlikely to be the outcome of a homogeneous innovation “black box”. 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 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 this, we use the USPTO/NBER database to 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. Similarly, we define comet and star *patents,* according to the classification of the respective inventors. A preliminary data inspection at metropolitan area level shows that the association with establishment births and other local structural characteristics 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 patenting activity.

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 attempt 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 around 3% more patents authored by comet inventors.

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 only partially support the effectiveness of these policies: we find some evidence 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 through an increase in the activity of comet inventors. At the same time, we find that stars and comets tend to concentrate in different localities; thus attracting stars where there are comets may reduce the former's productivity.

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 further spurred more interest in patent data: for the first time, researchers had a tool to "trace" knowledge spillovers, which previously had been considered 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, also 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) shows that a cited-citing patent couple is twice as likely to be in the same US metropolitan area as a couple of technologically similar patents with no citation links. Similarly, Peri (2005) examines 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) demonstrates that academic research has large effects on the number of private patents developed in the same US state. Finally, Carlino et al. (2007) use 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. All these contributions (and many similar which we omit for brevity) highlight that knowledge spillovers have a geographically limited distance decay.

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), highlight that 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' applications 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.

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. (2010) exploit the exogenous variation in the number of "superstar scientists" in US universities due to 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. (2010), 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 "sceptical" 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, some economists argue that technological progress has actually increased the scope for proximity for innovative activities due to the greater 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.

Previous contributions on star scientists, however, did not look at the other tail of the distribution of patents across inventors or, more generally, discuss the strong skewness of the distribution.[1] This is in part due to the fact that until very recently a unique identifier for inventors was not available in the NBER/USPTO database and therefore calculating the distribution of patents by inventors was not feasible. Thanks to the efforts of Trajtenberg et al (2006), who "estimate" a unique inventor identifier using an ad-hoc algorithm,[2] we know that out of 758,000 inventors listed in the NBER dataset in the period 1978-99, 28% registered just one patent, 34% from 2 to 5, and only 5% more than 20 patents. A snapshot of the skewness of the distribution of patents across inventor is reported in Figure 1.

The peculiar distribution of patents by inventors reveals that patenting is a proxy of many different innovation activities. On one side, a large number of patents is developed by "comet" inventors, i.e., individuals who apply for a patent only once or twice over a long period. On the other side, a small group of "star" inventors develop individually a huge number of patents. We can reasonably suppose that comet patents are inventions made by individuals or firms whose primary activity is not scientific research (although we would need a patent-firm matched dataset to validate this hypothesis). This does not mean that comet patents are less important: they may give birth to new entrepreneurial

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

[2] The authors need to tackle two orders of problems: first, the same author may appear in the database under 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.), together with word sound matching routines. The validity of the procedure is confirmed by a test on a dataset of Israeli inventors.

projects and to spin-offs of new firms. This seems to be confirmed by empirical evidence: Balasubramanian and Sivadasan (2008) in a recent working paper link patent records to Census firm data for the US, in order 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.

On the other side, star patents are likely to be the outcome of specialized labs of big companies with a constant flow of patents. This activity is still economically relevant, but may have weaker implications on the local economic environment: the productivity gains of these inventions are likely to be spread across the different sites where the company is located. In the light of that, the first part of our empirical analysis will be aimed at i) describing the geography of stars and comets across US Metropolitan Statistical Areas (MSAs) and ii) assessing the linkages between stars, comets, and the local economy. Since we find some evidence suggesting that stars have indeed little connection with the local economy, it becomes crucial to assess whether there are substantial, locally bounded knowledge spillovers from stars to comets. If this were the case, stars would still have an important (indirect) effect on the local economy. The second part of the empirical analysis will therefore address this question.

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 have 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 first glance, the abundance of data makes a micro analysis at inventor level the most appealing alternative. A deeper view, however, suggests that this is not feasable, in light of the simple fact that the dataset is about patents, not inventors, implying 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 that 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 in which the patent is granted,[3] which is generally 2-3 years after the

[3] The reason why we use the grant year, rather than the application year, 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 1999, as they are not included in the dataset. This subsample

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 timing 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 also happen a few years after the star has applied for (or been granted) 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.[4] 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 (in Section 4) are robust to measurement errors,[5] 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 aim for the highest degree of precision, 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.

Ideally, we would like to observe inventors for their whole period of actvity, and then classify them as stars or comets according to their propensity to patent. However, since our data cover the 1975-99 period only, we cannot observe the whole career of the large majority of the inventors in the sample; furthermore, the productivity of inventors varies substantially along their lifework, and so does the intensity of the relative knowledge spillovers. We therefore adopt a definition that takes into account the productivity of inventors in that stage of their career, and to smooth short-term disturbances.

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

Therefore, for each of the five periods we define an 8-years long, overlapping *observational window,* 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

could easily be non-random, e.g. better patents may take longer to be examined, etc.

[4]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.

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

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' distribution in terms 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 than six up to 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 threshold 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 of overestimating 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 mainly entails inventors working for firms whose 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 a 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, this 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 Table 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 up to 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 periods 1 and 2. We define five periods, however, as the first two are used to build the instrumental variables.

Star patents account for 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% are listed as stars at least once, while for comets the 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 over time: around 40% of stars in a given period are stars also in the period before. The share goes down to 15% with a two period 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 with 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 (the table is reported in Appendix D). We interpret this as further evidence that the stars/comets categorization, although stylized and somehow arbitrary, does 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: citation 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 stars 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 periods 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, 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 receives, on average, 0.87 citations more than "other patents" (0.10 time 8.7, i.e., the difference between the two coefficients multiplied by the standard deviation of

Table 4: Regression of citations received

VARIABLES	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

the dependent variable, the number of citations received). Comets, on the other hand, 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 as a hobby).

3.1 Preliminary evidence on the 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,[6] we can see that there is a sizeable correlation (table 5), which implies that innovative activity is overall spatially concentrated. When plotting the shares of comet and star patents on the total of patents, however, there is a fair degree of dispersion in both the distributions, especially for the stars (figure 2).

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

[6] 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

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 are a list of simple proxies of the industrial structure of the MSA: log of total employment (*totemp*), share of employment in manufacturing (*manuf. share*), Herfindahl diversity index (*Herfindahl*, calculated as the sum of the squares of the share over the total of employment of 2-digit SIC sectors), and log of the number of plants with less then 500 employees (*n. plants <500 emp.*). We also include 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 exclude stars and comets to avoid circularity). 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 MSA structural variables refer to the first year of the time period, while the patent variables correspond to the sum over the period.

The results - reported in table 6 - clearly show how the two vectors of coefficient are different. 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, ceteris paribus). 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, and is thus 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

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

VARIABLES	(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

of average establishment employment - all lagged by one period to avoid simultaneity bias), for periods 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 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 affect comets positively?

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

Table 7: Regression of establishment births at MSA level

VARIABLES	(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

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.

It is worth noticing that all the mechanisms may, in theory, work also in the opposite direction (from comets to stars); we therefore 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., 2010; 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 the 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 through 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:

$$Comets_{ikt} = \beta \cdot Stars_{ikt} + \theta Z_{ikt} + \gamma X_{it} + \delta_k + \tau_t + \phi_i + \delta_k \tau_t + \varepsilon_{ikt} \tag{3}$$

where i, k, and t index MSAs, categories, and periods, respectively; Stars and Comets are the number of patents in the respective group, Z is a control specific to the MSA/category pair, 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; attempts to estimate the same model at aggregate (MSA) level did not produce any significant coefficient (results are available upon request). This is likely to be due to the fact that knowledge flows in the patenting are mostly contained within the same technological category, as confirmed by data: 80% of citation linkages are bounded within the same category. Furthermore, this also 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.[7]

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 drawback of the log transformation is the loss of the zeros, which, however, are less than 5% of observations. In the following section, we perform some robustness tests on the whole sample based on a Negative Binomial model with the natural count variable and we find compatible results.

The MSA/category control (Z) is the number of other patents granted (neither stars nor comets) in the technological category, over the other patents in the other five categories (*share other patents cat.*); it controls for the relative size of the specific technological category, and for idiosyncratic productivity shocks. It is worth noticing that this variable might be endogenous: although unlikely, we cannot a priori exclude that comet inventors produce knowledge spillovers benefiting other inventors in the same technological category. Unfortunately, we do not have any instrument available; however, we find that its inclusion has little effect on other coefficients, especially with the 2SLS estimator. As the latter is robust to omitted variables bias, the estimate of the coefficient for the variable of interest (the number of star patents) is consistent even excluding the (endogenous) control.

The second group of variables, included in the matrix X, relate to general city characteristics. Their inclusion is motivated by the findings we presented previously, namely the strong association of comet patents with a few specific MSA structural characteristics. 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, these variables are the following:

i) 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. We expect a positive coefficient on this variable.

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

[7] 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.

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

iv) 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.

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

The last four variables are exactly the same as in equations 1 and 2.

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 specifications, we include also an MSA-period fixed effect: although demanding in terms of degree of freedoms, it allows us to exclude all the time-varying, MSA-specific variables, which may potentially be endogenous. We could also include a MSA-category fixed effect but in this case identification would 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 give 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 intuition for the first instrument spurs from the fact that assignees of stars are generally multilocated. More than 60% of star patents are assigned to companies which are located in several MSAs (two or more); the companies which own most of the patents tend to employ inventors who are distributed across many different MSAs and states. Table 8 lists the 21 assignees which own more than 5,000 patents in the period under examination (1976-1999), reporting the number of different MSAs and states where their inventors are located, and the highest share of patents developed in an individual MSA: only one company is located in just one MSA (Ford Motor), while all the remaining assignees are located in several different states. Smaller assignees of star patents show a similar pattern. Therefore, 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 for a given company in that MSA, and ii) a US-wide increase in the productivity of this company 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 star inventors working for the same companies (but in different cities) to be correlated, due to their sharing a similar competition pressure, regulatory framework, market demand, technological shocks, company strategy, etc. We then suppose that a US-wide productivity shift in a given company will translate into MSA-specific productivity shocks in proportion to the number of inventors working for that 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 years 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 second instrument shares a similar intuition: instead of using the assignees, it is calculated interacting the lagged number of star inventors patenting in each of the 36 technological *sub*categories (as identified by Hall et al, 2001) and the US-wide variation in productivity within the same subcategory.

Table 8: The location of big patent companies

Company	**Nr of MSAs**	**Nr of US States**	**Share 1st MSA**
GEN ELECTRIC	36	19	0.45
DU PONT DE NEMOURS	21	13	0.63
INT BUSINESS MACHINES	21	15	0.17
WESTINGHOUSE ELECTRIC	17	9	0.38
AT & T	16	13	0.60
GEN MOTORS	15	5	0.46
HEWLETT PACKARD	14	8	0.35
ALLIED SIGNAL	13	11	0.40
MOTOROLA	12	8	0.35
UNISYS	12	9	0.16
DOW CHEM	9	9	0.48
TEXAS INSTR	8	3	0.72
UNITED TECH	8	5	0.58
EASTMAN KODAK	6	5	0.85
MONSANTO	6	6	0.66
RCA	6	4	0.38
XEROX	6	4	0.76
HUGHES AIRCRAFT	5	3	0.87
MINNESOTA MINING & MFG	5	4	0.88
MOBIL OIL	5	4	0.46
PHILLIPS PETROLEUM	5	3	0.91
EXXON RES & ENG	4	3	0.68
FORD MOTOR	1	1	1.00

The IV strategy is close in spirit to the approach of Bartik (1991) and Blanchard and Katz (1992), among others, who instrument 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 the first period, we calculate the total number of star inventors active in a given MSA and with a given assignee. In the 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:

$$IV2_{ikt} = \Sigma_a(StarsInv_{ika1} \cdot AvPat_{iat}) \quad (4)$$

where i indexes MSAs, t periods, k technological categories, and a the assignees. In the few cases in which the value of point b is missing (because there are not other stars with the same assignee in other MSAs), it is replaced with the contemporaneous US-wide average productivity of stars in the same technological category.

The validity of the IV relies on an assumption of excludability for point a), i.e., once MSA fixed effects are controlled for, the number of star inventors working for a given assignee in the first period 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., given that stars and comets have different assignees (the assignee is very often the employer of the inventor, and comets have, by definition, assignees with 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.

4.1.2 Second instrument

The second instrument is built in a similar way:

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 develops patents classified into different subcategories, he/she is assigned corresponding weights summing to one, according 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 periods 3-4-5, by MSA and category.

Formally, it can be expressed with the following equation:

$$IV1_{ikt} = \Sigma_s(StarsInv_{ikst-2} \cdot AvPat_{ikst}) \tag{5}$$

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 for 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, might compromise the validity of the instrument. We will rely on a test of overidentifying restrictions to rule out this concern.

5 Results

In Table 9 we report mean and standard deviation of the patent variables for the 2113 MSA/category pairs which make up 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 9: 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 in table 10. 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 has 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 magnifies 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 table 11. In Appendix C we report more specifications with IVs, 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 positive and significant, and the elasticity of comet to star patents now ranges between 0.28 and 0.32 (col. 1-3). The value is thus significantly greater than OLS estimates. We explain the downward bias of the OLS as originating from negative selection: it is likely that, in general, those star inventors that are more "exposed" to comet inventors have a lower potential for knowledge spillovers than the average star inventor. In other words, star inventors localiled in "comet-cities" may be "worse" than star inventors localized in "star cities". As this "lower quality" is unobserved, it introduces a (downward) bias in the OLS estimates.

However, another plausible explanation for the downward bias could be 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 may eliminate the "attenuation bias" of the OLS coefficient. The independence of the two errors is 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 try to assess whether the measurment error may indeed explain the OLS bias by building two new proxy variables for star patents: the average number of citations received by all the patents in the MSA/category, and the number of patents authored by those inventors who have developed more than 15 patents before. We then use these two variables as instruments; as they are clearly endogenous, the difference with the OLS estimates can give an indication of the amount of the "attenuation bias". Results (table 11, col. 4-5) suggest that the attuenuation bias is quite small, as the difference with the OLS coefficient is minimal. This seems to suggest that the underestimate of the OLS estimates is mainly due to negative selection.

Table 10: regression of comet patents, OLS

	(1)	(2)	(3)	(4)
VARIABLES	Comets (log)	Comets (log)	Comets (log)	Comets (log)
	OLS	OLS	OLS	OLS
Stars (log)	0.114***	0.142***	0.0980***	0.114***
	(0.0215)	(0.0188)	(0.0189)	(0.0237)
Share other patents cat.	0.289***		0.437***	0.455***
	(0.0797)		(0.0909)	(0.123)
Tot. MSA patents (log)	0.369***		0.0397	
	(0.0362)		(0.0874)	
Total MSA empl. (log)	0.0755	0.388	0.400*	
	(0.0529)	(0.244)	(0.242)	
Plants <500 emp. (log)	0.411***	0.0377	0.0444	
	(0.0609)	(0.189)	(0.193)	
Herfindahl	-1.030	2.424	2.676	
	(2.023)	(3.254)	(3.070)	
Manuf. share	0.401	-0.00163	0.0214	
	(0.432)	(0.565)	(0.569)	
Constant	-3.239***	-1.114	-1.462	1.037***
	(0.180)	(1.153)	(1.208)	(0.194)
MSA f.e.	NO	YES	YES	YES
Tech. cat.*Period f.e.	YES	YES	YES	YES
MSA*period f.e.	NO	NO	NO	YES
Observations	2,113	2,113	2,113	2,113
R-squared	0.764	0.858	0.861	0.890

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

Table 11: Regression of comet patents, IVE

VARIABLES	(1) comets (log)	(2) comets (log)	(3) comets (log)	(4) comets (log)	(5) comets (log)
	IV	IV	IV	IV - M.E.†	IV - M.E.†
Stars (log)	0.302***	0.280***	0.323***	0.130***	0.116***
	(0.0441)	(0.0661)	(0.0899)	(0.0268)	(0.0299)
Share other patents cat.		0.101	-0.0713	0.378***	0.449***
		(0.141)	(0.242)	(0.0901)	(0.116)
Tot. MSA patents (log)	-0.0222	-0.000546		0.0326	
	(0.0882)	(0.0884)		(0.0824)	
Total MSA empl. (log)	0.386	0.384		0.397*	
	(0.242)	(0.239)		(0.226)	
Plants <500 emp. (log)	0.0461	0.0454		0.0446	
	(0.187)	(0.186)		(0.180)	
Herfindahl	3.022	3.063		2.744	
	(3.157)	(3.092)		(2.882)	
Manuf. share	-0.291	-0.264		-0.0287	
	(0.564)	(0.560)		(0.531)	
Constant	-1.213*	-1.344*	0.110	-1.529**	1.026***
	(0.723)	(0.736)	(0.420)	(0.689)	(0.189)
MSA f.e.	YES	YES	YES	YES	YES
Tech. cat.*Period f.e.	YES	YES	YES	YES	YES
MSA*period f.e.	NO	NO	YES	NO	YES
Observations	2,113	2,113	2,113	2,113	2,113
R-squared	0.849	0.851	0.881	0.861	0.890
F-stat of excl. instr.	90.83	36.76	24.64	489.53	465.78

Heteroskedasticity robust standard errors clustered at MSA-category level in parentheses

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

†The instruments used in column 5 are meant to correct only for the measurment error

5.1 Robustness tests

We run a series of robustness tests to check the validity of our estimates. In table 12, we report the estimates of the model in equation 3 applying a Negative Binomial count model to different selections of the sample: the OLS one, the OLS one plus the observations with zero comets, the OLS one plus the observations with zero comets and less than 25 patents in the MSA/technological category pair, and all the observations (thus adding also the observations with zero stars; for ease of comparison, 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 largely unaffected by the different sample selections. Furthermore, its size is almost identical to 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 wellknown identification problem in spatial economics, and more generally in social sciences, as discussed by Manski, 1999). Nevertheless, totally ignoring spatial effects might also be 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 table 13. 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 they are assumed not to have any direct 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 column of table 13 therefore reports the results of an IV regression where stars and the spatial lag of comets are the endogenous variables, and the first IV and its spatial

Table 12: 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

lag are the instruments. The main coefficient of interest is almost unaffected compared with 2SLS regressions without spatial lags; all the other variables are insignificant. We therefore conclude that we cannot reject the presence of spatial effects in the context under analysis, but, at the same time, their omission does not affect our main results.[8]

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 different across periods, although the last one is generally slightly larger. 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 an increase of 3% in the number of comets.

More research is needed to expand both the conclusions we reach, in order to better identify the characteristics of cities associated with concentrations of the two categories of inventors, and to investigate the channels through which the spillovers take place. Also, the availability of a patent-firm matched dataset would allow us 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

[8] We omit the calculation of a spatial error model, robust to spatial correlation in the error term, as the large number of fixed effects included in the specifications and the clustered structure of the estimated standard errors already address the issue.

Table 13: Regressions with spatially lagged variables

VARIABLES	(1) Comets (log)	(2) Comets (log)	(3) Comets (log)	(5) Comets (log)
	OLS	OLS	OLS	IV
Stars (log)	0.0877***	0.0856***	0.0854***	0.322***
	(0.0190)	(0.0189)	(0.0190)	(0.0695)
Share other patents cat.	0.432***	0.421***	0.421***	0.00874
	(0.0906)	(0.0901)	(0.0902)	(0.06)
Tot. MSA patents (log)	0.0363	0.0368	0.0362	-0.00999
	(0.0882)	(0.0883)	(0.0882)	(-0.11)
Sp. lag stars (log)	0.119***		0.0156	0.0105
	(0.0305)		(0.0512)	(0.19)
Sp. lag oth. pat. (log)		0.193***	0.175**	0.264
		(0.0452)	(0.0769)	(1.11)
Sp. lag comets (log)				-0.222
				(-0.65)
Total MSA empl. (log)	0.369	0.387	0.382	0.455
	(0.245)	(0.243)	(0.243)	(1.74)
Plants <500 emp. (log)	0.0499	0.0513	0.0513	0.0525
	(0.188)	(0.187)	(0.187)	(0.28)
Herfindahl	2.182	1.952	1.963	2.661
	(3.057)	(3.057)	(3.056)	(0.83)
Manuf. share	0.0344	0.0342	0.0344	-0.317
	(0.563)	(0.560)	(0.560)	(-0.55)
Constant	-1.427	-1.734	-1.690	-2.084**
	(1.241)	(1.233)	(1.232)	(1.008)
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	2096	2096	2096	2096
R^2	0.863	0.864	0.864	0.854

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

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 the 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 concentrate 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 would end up to be sub-optimal for (at least) one of the two categories. Therefore, interfering in the location choice of stars (or comets) in order to increase the spatial proximity may introduce perverse incentives and lead to a much weaker effect than expected.

References

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

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

Azoulay, P., J. S. Graff Zivin, and J. Wang, 2010. Superstar Extinction. *Quaterly Journal of Economics* 125:2, 549-589.

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.

Blanchard O., L. F. Katz, 1992. Regional Evolutions, Brookings *Papers on Economic Activity.* 23:1. pp 1-76.

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.

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

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.

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 add the inventors' unique identifier developed by Trajtenberg et al (2006) and the standardized assignee name available in Prof. Bronwyn H. Hall's website.[9] 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 eliminate 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 the case where several 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, this 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.

[9] 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.

A first point of concern is the choice of considering only the first author of the patent. Looking at table 14, we can see that authors whose surname begins with one of the first letters of the alphabet are only slightly more likely to be reported as first author, compared with second or third authors. However, as a robustness test, we followed a different procedure, defining a patent a "star patent" if at least one inventor satisfies the requirements listed in section 3, and a "comet patent" if all the inventors satisfy the relative requirements. The new variables are highly correlated with the single-author ones across MSAs/categories (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 15). Therefore, to the extent that the first author is generally the project leader, defining comet and star patents based only on her/him probably increases the precision of the estimates.

A second point of concern is the criteria used to define comet patents We thus build three other definitions of these 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 calculate the results of the specification 3 with both OLS and IV (reported in table 16 and 17, respectively) and check whether the results are 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 should benefit less from spillovers from stars (assuming that if the quality of patents is lower, there will be points of contact with excellent patents). The second definition gives coefficients that are around 20% lower than the adopted definition; the difference is therefore small and due to similar reasons. The third comet variable gives a slightly higher coefficient in the OLS. 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 that pushes OLS estimates upwards (reducing 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.

We also re-estimate the model using a different definition of star patents, namely the one previously adopted to quantify the effect of the measurment error, i.e., the number of patents authored by an inventor who has developed 15 other patents up to that point

Table 14: 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 15: 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 and clustered at MSA-category level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 16: regression of comet patents, alternative definitions of comets, OLS

	(1)	(2)	(3)
VARIABLES	comets def 1 (log)	comets def 2 (log)	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 and clustered at MSA-category level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

in time. This definition is more raw and inprecise than the one we adopted and it is affected by a data truncation bias; nevertheless, is a proxy of the same phenomenon, and is useful to check the sensitivity of our results to variable definition. The results of both the OLS and IV estimations of the main model with the alternative star definition are reported in table 18. Again, the outcome is qualitatively very similar to our main specifications.

Table 17: regression of comet patents, alternative definitions of comets, IV

	(1)	(2)	(3)
VARIABLES	comets def 1 (log)	comets def 2 (log)	comets def 3 (log)
	IV	IV	IV
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 and clustered at MSA-category level in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 18: regression of comet patents, alternative definitions of stars

VARIABLES	(5) comets (log)	(6) comets (log)	(7) comets (log)	(8) comets (log)	(9) comets (log)
	OLS	OLS	IV	OLS	IV
Stars def. 2 (log)	0.125***	0.0502***	0.189***	0.0358	0.186**
	(0.0172)	(0.0187)	(0.0677)	(0.0231)	(0.0726)
Share other patents cat.		1.067***	0.511*	1.157***	0.519
		(0.161)	(0.301)	(0.189)	(0.332)
Tot. MSA patents (log)	-0.0607	0.120	0.0653		
	(0.0874)	(0.0878)	(0.0892)		
Total MSA empl. (log)	0.429*	0.362	0.380		
	(0.247)	(0.243)	(0.235)		
Plants <500 emp. (log)	-0.0120	-0.00460	-0.0576		
	(0.188)	(0.187)	(0.181)		
Herfindahl	1.717	2.274	2.101		
	(3.248)	(3.043)	(2.944)		
Manuf. share	0.215	0.0790	0.122		
	(0.551)	(0.551)	(0.519)		
Constant	-0.591	-1.430	-0.993	1.221***	0.741***
	(1.208)	(1.211)	(0.798)	(0.182)	(0.278)
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	YES	YES
Observations	2,113	2,113	2,113	2,113	2,113
R-squared	0.857	0.863	0.857	0.891	0.886

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

Appendix C: IV estimation diagnostics

In this appendix we present the results from alternative specifications regarding 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 10 with the overidentified specifications (thus including also the first instrument) estimated through 2-stage least squares (2SLS) and Limited Information Maximum Likelihood models (LIML).

In table 19 we report the results from the first stage regressions; columns 3-4 are the specifications correspondent to our preferred IV estimations reported in table 11. In both the specifications, the coefficients on the instrument are highly significant. In columns 1 and 2 we calculate the first-stage regressions including the instruments individually, while in col. 4 we estimate a regression which does not have any direct correspondence to any of the 2SLS estimates we present in the paper, but is meant to be a further test of the strength and exogeneity of the first instrument: specifically, we add a MSA-category fixed effect (not included 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 still significant, and its size is even bigger than in columns 1 and 2.

In table 20 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 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 for concern. However, in this case the coefficient values are very close to the ones estimated in the main model of table 10. Therefore, we can conclude that the validity of the IV estimation is not a concern in our case. In col. 3-4, finally, we report the 2SLS results using only one instrumental variable each time; again, the point estimates are only miniamally affected.

Table 19: First stage regression

VARIABLES	(1) stars (log)	(2) stars (log)	(3) stars (log)	(4) stars (log)	(5) stars (log)
IV1 (logs)		0.268***	0.221***	0.135***	0.528***
		(0.0327)	(0.0394)	(0.0406)	(0.149)
IV2 (logs)	0.169***		0.0507**	0.0692***	
	(0.0218)		(0.0233)	(0.0261)	
Share other patents cat.	1.521***	1.352***	1.341***	1.991***	0.0288
	(0.217)	(0.216)	(0.215)	(0.151)	(0.161)
Tot. MSA patents (log)	0.271**	0.287**	0.291**		0.812***
	(0.120)	(0.115)	(0.116)		(0.106)
Total MSA empl. (log)	-0.282	-0.0352	-0.125		-0.223
	(0.328)	(0.311)	(0.315)		(0.265)
Plants <500 emp. (log)	0.0836	0.0190	0.0415		-0.119
	(0.259)	(0.247)	(0.248)		(0.210)
Herfindahl	-2.436	-2.627	-2.633		-1.747
	(4.454)	(4.135)	(4.179)		(3.885)
Manuf. share	1.721*	1.597*	1.638*		1.125
	(0.973)	(0.920)	(0.917)		(0.747)
Constant	1.642	0.298	0.649	3.502***	-1.366
	(1.838)	(1.758)	(1.755)	(0.210)	(1.520)
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	YES	NO
MSA*cat. f.e.	NO	NO	NO	NO	YES
Observations	2113	2113	2113	2113	
R^2	0.784	0.790	0.791	0.846	0.297

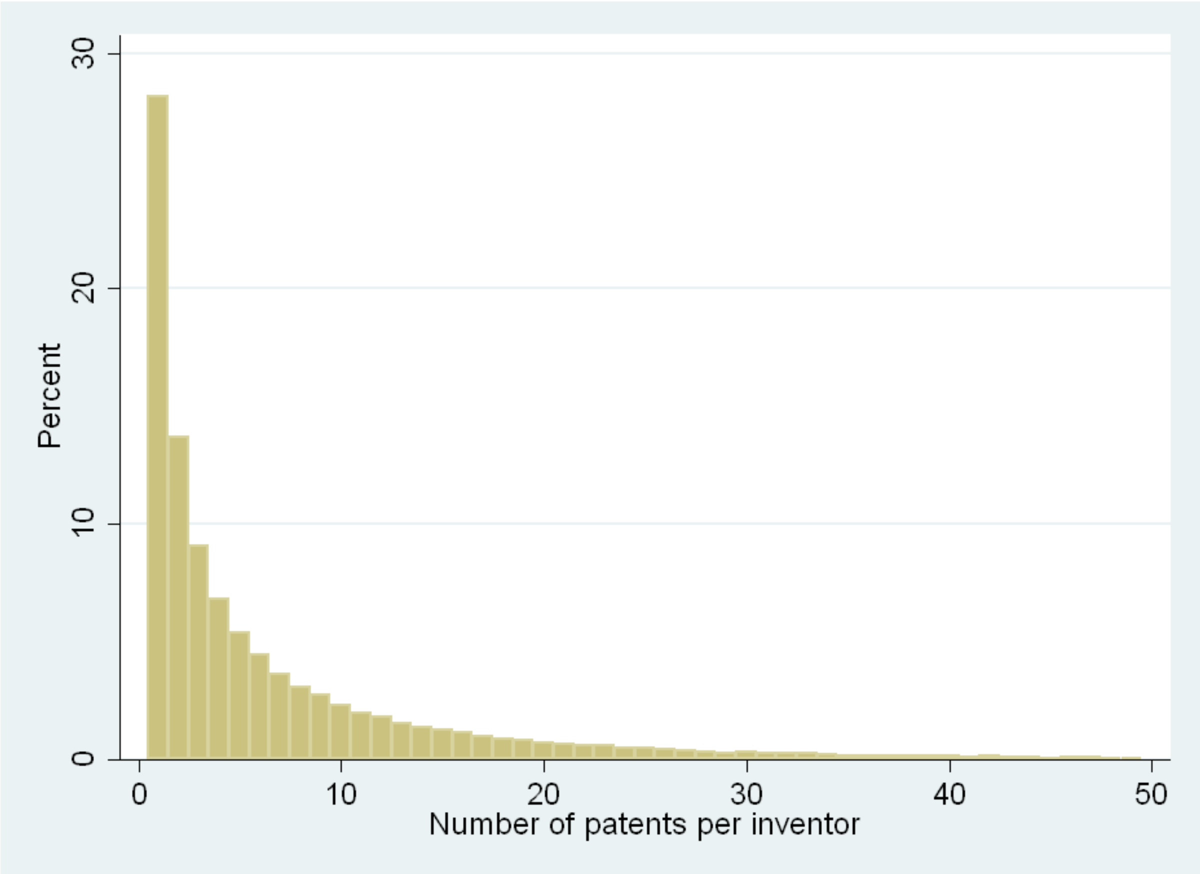
Heteroskedasticity robust standard errors and clustered at MSA-category level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 20: IV, overidentified regressions, 2SLS and LIML

VARIABLES	(1) Comets (log)	(2) Comets (log)	(3) Comets (log)	(4) Comets (log)
	LIML	LIML	2SLS - IV2	2SLS - IV1
Stars (log)	0.303***	0.281***	0.310***	0.273***
	(0.0441)	(0.0663)	(0.0748)	(0.0680)
Share other patents cat.		0.0997	0.0470	0.114
		(0.141)	(0.155)	(0.144)
Tot. MSA patents (log)	-0.0221	-0.000681	-0.00698	0.00107
	(0.0883)	(0.0884)	(0.0906)	(0.0880)
Total MSA empl. (log)	0.386	0.384	0.381	0.384
	(0.242)	(0.239)	(0.242)	(0.238)
Plants <500 emp. (log)	0.0461	0.0454	0.0456	0.0454
	(0.187)	(0.186)	(0.188)	(0.185)
Herfindahl	3.023	3.065	3.125	3.048
	(3.158)	(3.093)	(3.149)	(3.079)
Manuf. share	-0.292	-0.265	-0.310	-0.253
	(0.564)	(0.560)	(0.576)	(0.557)
Constant	-1.214*	-1.343*	-1.308*	-1.353*
	(0.723)	(0.736)	(0.748)	(0.733)
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	2113	2113	2113
R^2	0.849	0.851	0.848	0.852

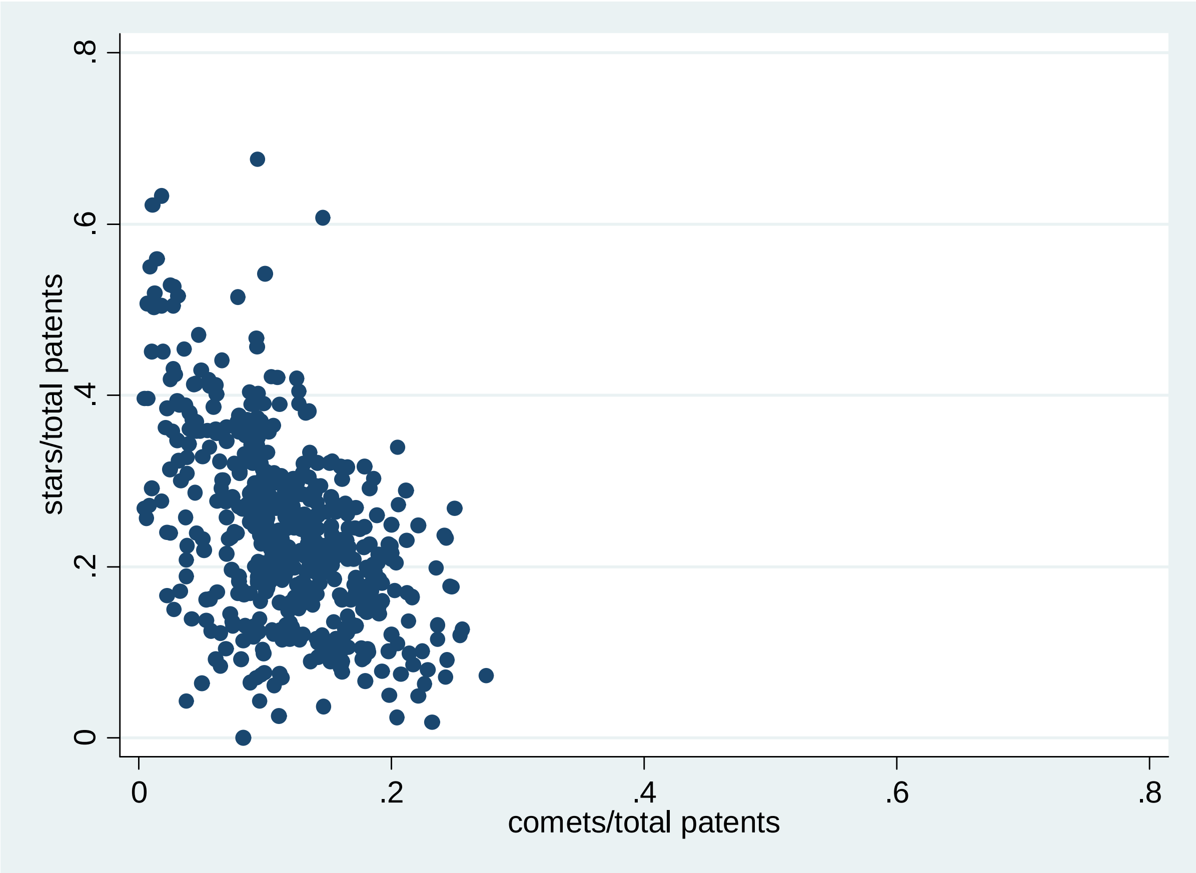
Heteroskedasticity robust standard errors and clustered at MSA-category level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 1: The distribution of patents across inventor



Note: the figure reports the histogram of the distribution of patents across US inventors in the period 1978-98.
Source: USTPO/NBER patents database.

Figure 2: share of comets and stars on total patents, US MSAs



Note: the figure reports the scatter plot of the share of star and comet patents over total patents across US MSAs 1986-98. The sample is limited to MSAs with at list 100 patents granted in the 4-years period.
Source: USTPO/NBER patents database.

RECENTLY PUBLISHED "TEMI" (*)

N. 762 – *A public guarantee of a minimum return to defined contribution pension scheme members*, by Giuseppe Grande and Ignazio Visco (June 2010).

N. 763 – *Debt restructuring and the role of lending technologies*, by Giacinto Micucci and Paola Rossi (June 2010).

N. 764 – *Disentangling demand and supply in credit developments: a survey-based analysis for Italy*, by Paolo Del Giovane, Ginette Eramo and Andrea Nobili (June 2010).

N. 765 – *Information uncertainty and the reaction of stock prices to news*, by Paolo Angelini and Giovanni Guazzarotti (July 2010).

N. 766 – *With a little help from abroad: the effect of low-skilled immigration on the female labor supply*, by Guglielmo Barone and Sauro Mocetti (July 2010).

N. 767 – *Real time forecasts of inflation: the role of financial variables*, by Libero Monteforte and Gianluca Moretti (July 2010).

N. 768 – *The effect of age on portfolio choices: evidence from an Italian pension fund*, by Giuseppe G. L. Cappelletti, Giovanni Guazzarotti and Pietro Tommasino (July 2010).

N. 769 – *Does investing abroad reduce domestic activity? Evidence from Italian manufacturing firms*, by Raffaello Bronzini (July 2010).

N. 770 – *The EAGLE. A model for policy analysis of macroeconomics interdependence in the euro area*, by Sandra Gomes, Pascal Jacquinot and Massimiliano Pisani (July 2010).

N. 771 – *Modelling Italian potential output and the output gap*, by Antonio Bassanetti, Michele Caivano and Alberto Locarno (September 2010).

N. 772 – *Relationship lending in a financial turmoil*, by Stefania De Mitri, Giorgio Gobbi and Enrico Sette (September 2010).

N. 773 – *Firm entry, competitive pressures and the US inflation dynamics*, by Martina Cecioni (September 2010).

N. 774 – *Credit ratings in structured finance and the role of systemic risk*, by Roberto Violi (September 2010).

N. 775 – *Entrepreneurship and market size. The case of young college graduates in Italy*, by Sabrina Di Addario and Daniela Vuri (September 2010).

N. 776 – *Measuring the price elasticity of import demand in the destination markets of Italian exports*, by Alberto Felettigh and Stefano Federico (October 2010).

N. 777 – *Income reporting behaviour in sample surveys*, by Andrea Neri and Roberta Zizza (October 2010).

N. 778 – *The rise of risk-based pricing of mortgage interest rates in Italy*, by Silvia Magri and Raffaella Pico (October 2010).

N. 779 – *On the interaction between market and credit risk: a factor-augmented vector autoregressive (FAVAR) approach*, by Roberta Fiori and Simonetta Iannotti (October 2010).

N. 780 – *Under/over-valuation of the stock market and cyclically adjusted earnings*, by Marco Taboga (December 2010).

N. 781 – *Changing institutions in the European market: the impact on mark-ups and rents allocation*, by Antonio Bassanetti, Roberto Torrini and Francesco Zollino (December 2010).

N. 782 – *Central bank's macroeconomic projections and learning*, by Giuseppe Ferrero and Alessandro Secchi (December 2010).

N. 783 – *(Non)persistent effects of fertility on female labour supply*, by Concetta Rondinelli and Roberta Zizza (December 2010).

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2008

P. ANGELINI, *Liquidity and announcement effects in the euro area*, Giornale degli Economisti e Annali di Economia, v. 67, 1, pp. 1-20, **TD No. 451 (October 2002).**

P. ANGELINI, P. DEL GIOVANE, S. SIVIERO and D. TERLIZZESE, *Monetary policy in a monetary union: What role for regional information?*, International Journal of Central Banking, v. 4, 3, pp. 1-28, **TD No. 457 (December 2002).**

F. SCHIVARDI and R. TORRINI, *Identifying the effects of firing restrictions through size-contingent Differences in regulation*, Labour Economics, v. 15, 3, pp. 482-511, **TD No. 504 (June 2004).**

L. GUISO and M. PAIELLA,, *Risk aversion, wealth and background risk*, Journal of the European Economic Association, v. 6, 6, pp. 1109-1150, **TD No. 483 (September 2003).**

C. BIANCOTTI, G. D'ALESSIO and A. NERI, *Measurement errors in the Bank of Italy's survey of household income and wealth*, Review of Income and Wealth, v. 54, 3, pp. 466-493, **TD No. 520 (October 2004).**

S. MOMIGLIANO, J. HENRY and P. HERNÁNDEZ DE COS, *The impact of government budget on prices: Evidence from macroeconometric models*, Journal of Policy Modelling, v. 30, 1, pp. 123-143 **TD No. 523 (October 2004).**

L. GAMBACORTA, *How do banks set interest rates?,* European Economic Review, v. 52, 5, pp. 792-819, **TD No. 542 (February 2005).**

P. ANGELINI and A. GENERALE, *On the evolution of firm size distributions*, American Economic Review, v. 98, 1, pp. 426-438, **TD No. 549 (June 2005).**

R. FELICI and M. PAGNINI, *Distance, bank heterogeneity and entry in local banking markets,* The Journal of Industrial Economics, v. 56, 3, pp. 500-534, **No. 557 (June 2005).**

S. DI ADDARIO and E. PATACCHINI, *Wages and the city. Evidence from Italy*, Labour Economics, v.15, 5, pp. 1040-1061, **TD No. 570 (January 2006).**

S. SCALIA, *Is foreign exchange intervention effective?,* Journal of International Money and Finance, v. 27, 4, pp. 529-546, **TD No. 579 (February 2006).**

M. PERICOLI and M. TABOGA, *Canonical term-structure models with observable factors and the dynamics of bond risk premia,* Journal of Money, Credit and Banking, v. 40, 7, pp. 1471-88, **TD No. 580 (February 2006).**

E. VIVIANO, *Entry regulations and labour market outcomes. Evidence from the Italian retail trade sector*, Labour Economics, v. 15, 6, pp. 1200-1222, **TD No. 594 (May 2006).**

S. FEDERICO and G. A. MINERVA, *Outward FDI and local employment growth in Italy,* Review of World Economics, Journal of Money, Credit and Banking, v. 144, 2, pp. 295-324, **TD No. 613 (February 2007).**

F. BUSETTI and A. HARVEY, *Testing for trend*, Econometric Theory, v. 24, 1, pp. 72-87, **TD No. 614 (February 2007).**

V. CESTARI, P. DEL GIOVANE and C. ROSSI-ARNAUD, *Memory for prices and the Euro cash changeover: an analysis for cinema prices in Italy,* In P. Del Giovane e R. Sabbatini (eds.), The Euro Inflation and Consumers' Perceptions. Lessons from Italy, Berlin-Heidelberg, Springer, **TD No. 619 (February 2007).**

B. H. HALL, F. LOTTI and J. MAIRESSE, *Employment, innovation and productivity: evidence from Italian manufacturing microdata*, Industrial and Corporate Change, v. 17, 4, pp. 813-839, **TD No. 622 (April 2007).**

J. SOUSA and A. ZAGHINI, *Monetary policy shocks in the Euro Area and global liquidity spillovers,* International Journal of Finance and Economics, v.13, 3, pp. 205-218, **TD No. 629 (June 2007).**

M. DEL GATTO, GIANMARCO I. P. OTTAVIANO and M. PAGNINI, *Openness to trade and industry cost dispersion: Evidence from a panel of Italian firms,* Journal of Regional Science, v. 48, 1, pp. 97-129, **TD No. 635 (June 2007).**

P. DEL GIOVANE, S. FABIANI and R. SABBATINI, *What's behind "inflation perceptions"? A survey-based analysis of Italian consumers,* in P. Del Giovane e R. Sabbatini (eds.), *The Euro Inflation and Consumers' Perceptions. Lessons from Italy,* Berlin-Heidelberg, Springer, **TD No. 655 (January 2008).**

R. BRONZINI, G. DE BLASIO, G. PELLEGRINI and A. SCOGNAMIGLIO, *La valutazione del credito d'imposta per gli investimenti*, Rivista di politica economica, v. 98, 4, pp. 79-112, **TD No. 661 (April 2008).**

B. BORTOLOTTI, and P. PINOTTI, *Delayed privatization,* Public Choice, v. 136, 3-4, pp. 331-351, **TD No. 663 (April 2008).**

R. BONCI and F. COLUMBA, *Monetary policy effects: New evidence from the Italian flow of funds,* Applied Economics , v. 40, 21, pp. 2803-2818, **TD No. 678 (June 2008).**

M. CUCCULELLI, and G. MICUCCI, *Family Succession and firm performance: evidence from Italian family firms,* Journal of Corporate Finance, v. 14, 1, pp. 17-31, **TD No. 680 (June 2008).**

A. SILVESTRINI and D. VEREDAS, *Temporal aggregation of univariate and multivariate time series models: a survey,* Journal of Economic Surveys, v. 22, 3, pp. 458-497, **TD No. 685 (August 2008).**

2009

F. PANETTA, F. SCHIVARDI and M. SHUM, *Do mergers improve information? Evidence from the loan market,* Journal of Money, Credit, and Banking, v. 41, 4, pp. 673-709, **TD No. 521 (October 2004).**

M. BUGAMELLI and F. PATERNÒ, *Do workers' remittances reduce the probability of current account reversals?,* World Development, v. 37, 12, pp. 1821-1838, **TD No. 573 (January 2006).**

P. PAGANO and M. PISANI, *Risk-adjusted forecasts of oil prices*, The B.E. Journal of Macroeconomics, v. 9, 1, Article 24, **TD No. 585 (March 2006).**

M. PERICOLI and M. SBRACIA, *The CAPM and the risk appetite index: theoretical differences, empirical similarities, and implementation problems*, International Finance, v. 12, 2, pp. 123-150, **TD No. 586 (March 2006).**

U. ALBERTAZZI and L. GAMBACORTA, *Bank profitability and the business cycle,* Journal of Financial Stability, v. 5, 4, pp. 393-409, **TD No. 601 (September 2006).**

S. MAGRI, *The financing of small innovative firms: the Italian case,* Economics of Innovation and New Technology, v. 18, 2, pp. 181-204, **TD No. 640 (September 2007).**

V. DI GIACINTO and G. MICUCCI, *The producer service sector in Italy: long-term growth and its local determinants,* Spatial Economic Analysis, Vol. 4, No. 4, pp. 391-425, **TD No. 643 (September 2007).**

F. LORENZO, L. MONTEFORTE and L. SESSA, *The general equilibrium effects of fiscal policy: estimates for the euro area*, Journal of Public Economics, v. 93, 3-4, pp. 559-585, **TD No. 652 (November 2007).**

R. GOLINELLI and S. MOMIGLIANO, *The Cyclical Reaction of Fiscal Policies in the Euro Area. A Critical Survey of Empirical Research*, Fiscal Studies, v. 30, 1, pp. 39-72, **TD No. 654 (January 2008).**

P. DEL GIOVANE, S. FABIANI and R. SABBATINI, *What's behind "Inflation Perceptions"? A survey-based analysis of Italian consumers,* Giornale degli Economisti e Annali di Economia, v. 68, 1, pp. 25-52, **TD No. 655 (January 2008).**

F. MACCHERONI, M. MARINACCI, A. RUSTICHINI and M. TABOGA, *Portfolio selection with monotone mean-variance preferences,* Mathematical Finance, v. 19, 3, pp. 487-521, **TD No. 664 (April 2008).**

M. AFFINITO and M. PIAZZA, *What are borders made of? An analysis of barriers to European banking integration,* in P. Alessandrini, M. Fratianni and A. Zazzaro (eds.): The Changing Geography of Banking and Finance, Dordrecht Heidelberg London New York, Springer, **TD No. 666 (April 2008).**

A. BRANDOLINI, *On applying synthetic indices of multidimensional well-being: health and income inequalities in France, Germany, Italy, and the United Kingdom,* in R. Gotoh and P. Dumouchel (eds.), Against Injustice. The New Economics of Amartya Sen, Cambridge, Cambridge University Press, **TD No. 668 (April 2008).**

G. FERRERO and A. NOBILI, *Futures contract rates as monetary policy forecasts,* International Journal of Central Banking, v. 5, 2, pp. 109-145, **TD No. 681 (June 2008).**

P. CASADIO, M. LO CONTE and A. NERI, *Balancing work and family in Italy: the new mothers' employment decisions around childbearing,* in T. Addabbo and G. Solinas (eds.), Non-Standard Employment and Qualità of Work, Physica-Verlag. A Sprinter Company, **TD No. 684 (August 2008).**

L. ARCIERO, C. BIANCOTTI, L. D'AURIZIO and C. IMPENNA, *Exploring agent-based methods for the analysis of payment systems: A crisis model for StarLogo TNG,* Journal of Artificial Societies and Social Simulation, v. 12, 1, **TD No. 686 (August 2008).**

A. CALZA and A. ZAGHINI, *Nonlinearities in the dynamics of the euro area demand for M1,* Macroeconomic Dynamics, v. 13, 1, pp. 1-19, **TD No. 690 (September 2008).**

L. FRANCESCO and A. SECCHI, *Technological change and the households' demand for currency,* Journal of Monetary Economics, v. 56, 2, pp. 222-230, **TD No. 697 (December 2008).**

G. ASCARI and T. ROPELE, *Trend inflation, taylor principle, and indeterminacy*, Journal of Money, Credit and Banking, v. 41, 8, pp. 1557-1584, **TD No. 708 (May 2007).**

S. COLAROSSI and A. ZAGHINI, *Gradualism, transparency and the improved operational framework: a look at overnight volatility transmission*, International Finance, v. 12, 2, pp. 151-170, **TD No. 710 (May 2009).**

M. BUGAMELLI, F. SCHIVARDI and R. ZIZZA, *The euro and firm restructuring*, in A. Alesina e F. Giavazzi (eds): Europe and the Euro, Chicago, University of Chicago Press, **TD No. 716 (June 2009).**

B. HALL, F. LOTTI and J. MAIRESSE, *Innovation and productivity in SMEs: empirical evidence for Italy,* Small Business Economics, v. 33, 1, pp. 13-33, **TD No. 718 (June 2009).**

2010

A. PRATI and M. SBRACIA, *Uncertainty and currency crises: evidence from survey data*, Journal of Monetary Economics, v, 57, 6, pp. 668-681, **TD No. 446 (July 2002).**

S. MAGRI, *Debt maturity choice of nonpublic Italian firms* , Journal of Money, Credit, and Banking, v.42, 2-3, pp. 443-463, **TD No. 574 (January 2006).**

R. BRONZINI and P. PISELLI, *Determinants of long-run regional productivity with geographical spillovers: the role of R&D, human capital and public infrastructure,* Regional Science and Urban Economics, v. 39, 2, pp.187-199, **TD No. 597 (September 2006).**

E. IOSSA and G. PALUMBO, *Over-optimism and lender liability in the consumer credit market,* Oxford Economic Papers, v. 62, 2, pp. 374-394, **TD No. 598 (September 2006).**

S. NERI and A. NOBILI, *The transmission of US monetary policy to the euro area,* International Finance, v. 13, 1, pp. 55-78, **TD No. 606 (December 2006).**

F. ALTISSIMO, R. CRISTADORO, M. FORNI, M. LIPPI and G. VERONESE, *New Eurocoin: Tracking Economic Growth in Real Time,* Review of Economics and Statistics, v. 92, 4, pp. 1024-1034, **TD No. 631 (June 2007).**

A. CIARLONE, P. PISELLI and G. TREBESCHI, *Emerging Markets' Spreads and Global Financial Conditions*, Journal of International Financial Markets, Institutions & Money, v. 19, 2, pp. 222-239, **TD No. 637 (June 2007).**

U. ALBERTAZZI and L. GAMBACORTA, *Bank profitability and taxation*, Journal of Banking and Finance, v. 34, 11, pp. 2801-2810, **TD No. 649 (November 2007).**

M. IACOVIELLO and S. NERI, *Housing market spillovers: evidence from an estimated DSGE model,* American Economic Journal: Macroeconomics, v. 2, 2, pp. 125-164, **TD No. 659 (January 2008).**

F. BALASSONE, F. MAURA and S. ZOTTERI, *Cyclical asymmetry in fiscal variables in the EU*, Empirica, **TD No. 671**, v. 37, 4, pp. 381-402 **(June 2008).**

F. D'AMURI, O. GIANMARCO I.P. and P. GIOVANNI, *The labor market impact of immigration on the western german labor market in the 1990s,* European Economic Review, v. 54, 4, pp. 550-570, **TD No. 687 (August 2008).**

A. ACCETTURO, *Agglomeration and growth: the effects of commuting costs,* Papers in Regional Science, v. 89, 1, pp. 173-190, **TD No. 688 (September 2008).**

S. NOBILI and G. PALAZZO, *Explaining and forecasting bond risk premiums*, Financial Analysts Journal, v. 66, 4, pp. 67-82, **TD No. 689 (September 2008).**

A. B. ATKINSON and A. BRANDOLINI, *On analysing the world distribution of income*, World Bank Economic Review , v. 24, 1 , pp. 1-37, **TD No. 701 (January 2009).**

R. CAPPARIELLO and R. ZIZZA, *Dropping the Books and Working Off the Books,* Labour, v. 24, 2, pp. 139-162 ,**TD No. 702 (January 2009).**

C. NICOLETTI and C. RONDINELLI, *The (mis)specification of discrete duration models with unobserved heterogeneity: a Monte Carlo study*, Journal of Econometrics, v. 159, 1, pp. 1-13, **TD No. 705 (March 2009).**

V. DI GIACINTO, G. MICUCCI and P. MONTANARO, *Dynamic macroeconomic effects of public capital: evidence from regional Italian data,* Giornale degli economisti e annali di economia, v. 69, 1, pp. 29-66, **TD No. 733 (November 2009).**

F. COLUMBA, L. GAMBACORTA and P. E. MISTRULLI, *Mutual Guarantee institutions and small business finance,* Journal of Financial Stability, v. 6, 1, pp. 45-54, **TD No. 735 (November 2009).**

A. GERALI, S. NERI, L. SESSA and F. M. SIGNORETTI, *Credit and banking in a DSGE model of the Euro Area*, Journal of Money, Credit and Banking, v. 42, 6, pp. 107-141, **TD No. 740 (January 2010).**

M. AFFINITO and E. TAGLIAFERRI, *Why do (or did?) banks securitize their loans? Evidence from Italy*, Journal of Financial Stability, v. 6, 4, pp. 189-202, **TD No. 741 (January 2010).**

S. FEDERICO, *Outsourcing versus integration at home or abroad and firm heterogeneity*, Empirica, v. 37, 1, pp. 47-63, **TD No. 742 (February 2010).**

V. DI GIACINTO, *On vector autoregressive modeling in space and time*, Journal of Geographical Systems, v. 12, 2, pp. 125-154, **TD No. 746 (February 2010).**

A. DI CESARE and G. GUAZZAROTTI, *An analysis of the determinants of credit default swap spread changes before and during the subprime financial turmoil*, Journal of Current Issues in Finance, Business and Economics, v. 3, 4, pp., **TD No. 749 (March 2010).**

A. BRANDOLINI, S. MAGRI and T. M SMEEDING, *Asset-based measurement of poverty*, Journal of Policy Analysis and Management, v. 29, 2 , pp. 267-284, **TD No. 755 (March 2010).**

G. CAPPELLETTI, *A Note on rationalizability and restrictions on beliefs*, The B.E. Journal of Theoretical Economics, v. 10, 1, pp. 1-11,**TD No. 757 (April 2010).**

S. DI ADDARIO and D. VURI, *Entrepreneurship and market size. the case of young college graduates in Italy*, Labour Economics, v. 17, 5, pp. 848-858, **TD No. 775 (September 2010).**

FORTHCOMING

L. MONTEFORTE and S. SIVIERO, *The Economic Consequences of Euro Area Modelling Shortcuts*, Applied Economics, **TD No. 458 (December 2002).**

M. BUGAMELLI and A. ROSOLIA, *Produttività e concorrenza estera*, Rivista di politica economica, **TD No. 578 (February 2006).**

G. DE BLASIO and G. NUZZO, *Historical traditions of civicness and local economic development*, Journal of Regional Science, **TD No. 591 (May 2006).**

S. DI ADDARIO, *Job search in thick markets*, Journal of Urban Economics, **TD No. 605 (December 2006).**

F. SCHIVARDI and E. VIVIANO, *Entry barriers in retail trade*, Economic Journal, **TD No. 616 (February 2007).**

G. FERRERO, A. NOBILI and P. PASSIGLIA, *Assessing excess liquidity in the Euro Area: the role of sectoral distribution of money,* Applied Economics, **TD No. 627 (April 2007).**

P. E. MISTRULLI, *Assessing financial contagion in the interbank market: maximun entropy versus observed interbank lending patterns,* Journal of Banking & Finance, **TD No. 641 (September 2007).**

Y. ALTUNBAS, L. GAMBACORTA and D. MARQUÉS, *Securitisation and the bank lending channel*, European Economic Review, **TD No. 653 (November 2007).**

E. CIAPANNA, *Directed matching with endogenous markov probability: clients or competitors?*, The RAND Journal of Economics, **TD No. 665 (April 2008).**

M. BUGAMELLI and F. PATERNÒ, *Output growth volatility and remittances*, Economica, **TD No. 673 (June 2008).**

V. DI GIACINTO e M. PAGNINI, *Local and global agglomeration patterns: two econometrics-based indicators*, Regional Science and Urban Economics, **TD No. 674 (June 2008).**

P. SESTITO and E. VIVIANO, *Reservation wages: explaining some puzzling regional patterns*, Labour, **TD No. 696 (December 2008).**

P. PINOTTI, M. BIANCHI and P. BUONANNO, *Do immigrants cause crime?*, Journal of the European Economic Association, **TD No. 698 (December 2008).**

L. FORNI, A. GERALI and M. PISANI, *Macroeconomic effects of greater competition in the service sector: the case of Italy,* Macroeconomic Dynamics, **TD No. 706 (March 2009).**

Y. ALTUNBAS, L. GAMBACORTA, and D. MARQUÉS-IBÁÑEZ, *Bank risk and monetary policy*, Journal of Financial Stability, **TD No. 712 (May 2009).**

P. ANGELINI, A. NOBILI e C. PICILLO, *The interbank market after August 2007: What has changed, and why?*, Journal of Money, Credit and Banking, **TD No. 731 (ottobre 2009).**

L. FORNI, A. GERALI and M. PISANI, *The macroeconomics of fiscal consolidations in euro area countries,* Journal of Economic Dynamics and Control, **TD No. 747 (March 2010).**

A. DI CESARE and G. GUAZZAROTTI, *An analysis of the determinants of credit default swap spread changes before and during the subprime financial turmoil,* in C. V. Karsone (eds.), Finance and Banking Developments, Nova Publishers, New York., **TD No. 749 (March 2010).**

G. GRANDE and I. VISCO, *A public guarantee of a minimum return to defined contribution pension scheme members,* Journal of Risk, **TD No. 762 (June 2010).**

S. MAGRI and R. PICO, *The rise of risk-based pricing of mortgage interest rates in Italy*, Journal of Banking and Finance, **TD No. 696 (October 2010).**