Research Policy 37 (2008) 1892–1908

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Does it matter where patent citations come from? Inventor vs. examiner citations in European patents

Paola Criscuolo[a], Bart Verspagen[b,c,*]

[a] Tanaka Business School, Imperial College, South Kensington Campus, London SW7 2AZ, UK
[b] Maastricht University, Department of Economics, and UNU-Merit, PO Box 616, NL-6200 MD Maastricht, The Netherlands
[c] TIK, University of Oslo, Postboks 1108, Blindern N-0317 Oslo, Norway

ARTICLE INFO

Article history:
Received 3 October 2006
Received in revised form 11 July 2008
Accepted 14 July 2008
Available online 3 September 2008

JEL classification:
O30
O34

Keywords:
Patent citations
Local knowledge spillovers

ABSTRACT

This paper addresses the question of whether patent citations are useful indicators of technology flows. We exploit the distinction between citations added by inventors and patent examiners. We use information from the search reports of European Patent Office patent examiners to construct our dataset of patenting activity in Europe and the US, and apply various econometric models to investigate what determines the probability that a citation is added by the inventor rather than the examiner. Contrary to previous work which uses US Patent and Trademark Office data, we find that geographical distance is a factor that strongly diminishes the probability of knowledge flows. We find other significant effects of such factors as cognitive distance, time and strategic factors on citing behaviour.

1. Introduction

Patent citations have been used extensively as indicators of technology spillovers, and technology flows more generally. However, this is a very indirect use of patent citation data; citations are not intended to be an indication of technology flows or spillovers. They are instead a response to the legal requirement to supply a complete description of the state of the art in the field of the invention. Thus, citations limit the scope of an inventor's claim to novelty and represent a link to the pre-existing knowledge upon which the invention is built. The latter fact has been used to justify their use as indicators of knowledge spillovers. An inventor's citing of a patent or scientific article may indicate that the knowledge contained in the cited document has been useful in the development of the citing patent, and therefore that the citation might be a proxy for knowledge flow.

A criticism that has been levelled at the use of patent citations as an indicator of spillovers is that citations are a very 'noisy' indicator (Jaffe et al., 1998), i.e., they can be interpreted in several different ways and do not always point to the actual flow of knowledge from cited to citing inventor. A crucial factor here is that patent citations can be included by the applicant (or his/her patent lawyer) and also can be added by the patent examiner responsible for judging the degree of novelty of the patent. Where citations are added by the patent examiner, we cannot judge whether or not the applicants were aware of the cited patent. Jaffe et al. (1998) show that in many instances they were not and, hence, citation data are a 'noisy' indicator of spillovers or knowledge flows.

Alcacer and Gittelman (2006) joined the debate by proposing two scenarios for examiner citing behaviour: that the patent examiner might add citations that differ in

* Corresponding author at: Maastricht University, Department of Economics, PO Box 616, NL-6200 MD Maastricht, The Netherlands.
Tel.: +31 43 3883691.
E-mail addresses: p.criscuolo@imperial.ac.uk (P. Criscuolo), b.verspagen@algec.unimaas.nl (B. Verspagen).

0048-7333/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.respol.2008.07.011

nature from the inventor/applicant citations ('gap-filling'), or that the examiner might add similar citations ('tracking'). When patent citations are used in econometric analyses as indicators of inventor behaviour, gap-filling implies that failure to acknowledge the source of the citation may produce statistical results that are biased. Tracking does not lead to any bias but it may cause standard errors in statistical estimations to be inflated. Moreover, tracking raises doubts about patent citations as indicators of knowledge flows (Alcacer and Gittelman, 2006, p. 775). A priori, examiner citations may be taken as a valid reflection of technological and legal relatedness. But, since much of the literature argues that knowledge flows are a limited subset of potential technological relations (e.g., Jaffe et al., 1993 argue that knowledge flows are more likely where short geographical distances are involved), one would not expect inventor citations to be similar to examiner citations. If they are similar (tracking), this may indicate that these citations reflect expectations in examiners' opinions rather than knowledge flows that played a role during the invention process.

We build on the tracking vs. gap-filling distinction by formulating a research question in terms of potential and actual spillovers. Our research question is aimed at identifying the factors that influence whether an observed patent-to-patent citation was added by the applicant/inventor. We assume a citation to be an indicator of a *potential* spillover or knowledge flow, and whether or not the inventor/applicant added the citation as a property indicating whether a knowledge flow *actually* occurred. If the Alcacer and Gittelman (2006) gap-filling model holds, we should be able to explain the occurrence of actual spillovers (i.e., inventor/applicant citation) by the existing theoretical models on spillovers. For example, if we find that geographical distance impacts negatively on the likelihood of an inventor (vs. examiner) citation, this will indicate that inventors tend to choose their citations from within a narrower geographical space than do examiners. In that case, we can conclude that inventors make a selection from the set of technologically and legally relevant citations, which is consistent with the idea of patent citations being indicators of knowledge flows.

The explanatory variables, i.e., those variables expected to have an influence on whether or not a potential spillover or technology flow occurs, are geographical distance, technological or cognitive distance (between the 'sender' and 'receiver' of the flow), time lapse between the citing and cited patents, strategic behaviour of the applicant, self-citations and international patent application. Geographical distance is the variable that has been the focus of most econometric work in the area. A large body of empirical studies has exploited the use of patent citations to assess the spatial nature of technological spillovers (Jaffe et al., 1993, 1998; Jaffe and Trajtenberg, 1996, 1999; Maurseth and Verspagen, 2002). These authors looked at whether or not knowledge spillovers between firms, or from (semi-) public knowledge institutes to firms, depend on geographical distance, i.e., whether citing occurs, *ceteris paribus*, more frequently between inventors located close to one another. The results show that in both the US and Europe, such a relationship exists. Here, we test the hypothesis that proximity between parties increases the probability of a knowledge flow.

With the exception of the European Patent Office (EPO) dataset that we used in this study and recent US Patent and Trademark Office (USPOT) data, it is not possible to identify precisely those citations chosen by the inventor. Moreover, the role of examiner vs. inventor[2] citations differs among patent systems. And, ultimately, the final decision about which documents are cited in the published patent is made by the patent examiner. The patent examiner might decide to retain the citations proposed by the applicant and/or add new references, which will lead to the bias identified above that patent citations might not reflect an actual source of knowledge spillovers.

Two recent studies investigated the citations in patents granted by the USPTO. The studies by Alcacer and Gittelman (2006) and Thompson (2006) exploit the fact that, since 2001, the USPTO provides information on the source of patent citations. Thompson (2006) is aimed primarily at investigating whether or not knowledge spillovers are geographically concentrated. Alcacer and Gittelman's (2006) investigation is closer to the present study, and looks at how inventor and examiner citations differ.

In this study we explore the inventor/examiner origin of patent citations in EPO data, where it has been possible to identify the source of the citations since the EPO was established in 1979. This allows us to test whether the results obtained by Alcacer and Gittelman (2006), based on US patents, are typical of a different patent system. We would expect to find some differences for two reasons. First, the nature of the geographical space, which is different to that in the US due to cultural factors, language, the existence of national borders, etc. In order to identify more clearly the effect of the patent system, we implement estimations related to US-based inventors applying for EPO patents. Second, as Alcacer and Gittelman (2006) stress, USPTO and EPO patent examiner practices differ substantially, and particularly with regard to disclosure of prior art. This has a strong effect on the relative number of citations included by the inventor (see below), and therefore could have strong implications for the use of patent citations as a proxy for knowledge spillovers.

2. Patent citations

Patents contain references to prior patents and the scientific literature.[3] The legal purpose of references in patents is to indicate which parts of the knowledge described are claimed in the patent and which parts have been claimed

[2] We use the term 'inventor citation' to indicate a citation in the original patent application, irrespective of whether the inventor, the patent lawyer or someone else involved in the application added the citation.

[3] Patent citations in EPO patents are contained in the search report, which is a separate document attached to the patent and completed by the examiner. In USPTO patents, citations are reported on the front page of the patent document. In the case of both EPO and USPTO patents there may be citations to patents and non-patent literature embedded in the text of the patent document (Narin et al., 1988), but these citations are difficult to examine because they are not available in electronic format.

by previous patents or non-patent. As Collins and Wyatt (1988) explain, the applicant 'must set out the background in such a way as to show how the claimed invention relates to, but is innovatively different from what was already public knowledge', and his/her task is also to identify work 'either related to but significantly different from, or else a useful step towards, the new invention or a use of the invention'.

Although similar to references in journal articles, patent citations differ in two respects. First, while academic citations are mainly the prerogative of the author, citations in patents are the results of a highly mediated process which involves the inventor, the patent attorney and the patent examiner (Meyer, 2000). Second, articles in journals may be cited for a variety of reasons, not all of them reflecting recognition of work done previously or knowledge transfer. Authors may cite articles for strategic reasons, e.g., because the authors of the cited article might be potential reviewers. Inventors, on the other hand, have an incentive not to cite patents unnecessarily, as it may reduce their claims to novelty of the invention and therefore affect the scope of the monopoly rights granted by the patent.

In principle, when a patent cites another patent, this indicates that the knowledge embodied in the cited patent has been useful in some way for developing the new knowledge described in the citing patent and that the citing patent has no claim over that particular knowledge. This is the line of reasoning in Jaffe et al. (1993), and Jaffe and Trajtenberg (1996, 1999) for USPTO patents. Thus, patent citations represent a 'paper trail' of the knowledge flows between citing and cited inventors, although, as pointed out by Jaffe and Trajtenberg (2002), 'one that is incomplete and mixed with a fair amount of noise'.

Patent citations are an *incomplete measure of knowledge flows* because they capture only those flows that result in a novel and patentable technology and therefore they cannot be used to make inferences about knowledge transfers that do not result in a patent, such as tacit forms of knowledge, learning via imitation, or reverse engineering. It should also be emphasised that knowledge flows are a much broader concept than is captured simply by patent citations. In terms of the distinction introduced by Griliches (1992), patent citations focus on a specific form of pure knowledge spillovers. Rent spillovers, which reflect the fact that intermediate input prices do not completely embody product innovations or the quality improvements resulting from research and development (R&D) activities, are completely ignored. However, as pointed out by Breschi and Lissoni (2001), although in theory, patent citations try to measure pure knowledge spillovers, empirically it is hard to exclude those knowledge flows (giving rise to patent citations) that are mediated by markets or market mechanisms. Even within the category of pure knowledge spillovers, patent citations (to the extent that they are related to spillovers) are only a part of the story. For example, in order for patents to be cited, both the spillover-receiving and spillover-generating firms must be actively engaged in R&D and apply for patent protection. Therefore, knowledge flows can occur without generating citations.

Patent citations are a *noisy measure of knowledge flows* because, although suggested by the inventor and/or the inventor's attorney, the final decision on which patents to cite ultimately lies with the patent examiner. This implies that the inclusion of a given citation does not necessarily indicate that the inventor has knowledge of the technology underlying the cited patent and, thus, it does not represent an actual knowledge source utilised by the inventor in the development of the invention. In this study we eliminate this source of noise, although there are three others that may have an effect (Jaffe et al., 1998). The first derives from the intervention of the patent attorney who might decide to cite a patent not considered by the inventor to constitute prior art. The attorney may include it to avoid the risk of any future legal battle (strictly legal citation). The second relates to the possibility that inventors might have learnt about the cited invention only after the development of their own invention (after-fact citation). In this case the citation cannot be interpreted as a source of knowledge contributing to the development of the invention, but still represents a knowledge flow between citing and cited inventors. The third source of noise is associated with a citation to a patent, which, while not drawn on directly by the inventor in the inventing process, is nonetheless seen as basic to the process (teaching citation).

Breschi and Lissoni (2004) also argued against the use of patent citations as a proxy for inter-personal knowledge spillovers. These authors distinguish between two types of innovative efforts resulting in patents: cumulative efforts, i.e., the citing inventor built upon the knowledge developed by the cited patent, and duplicative efforts, i.e., the citing inventor duplicated the cited inventor's research. In the latter case citation might not involve any exchange of knowledge between the inventors and might not be associated with either awareness or intellectual debt between the cited and citing patent. When patents are the result of cumulative innovative efforts citations might be the result of either the citing inventor's search in a patents database, which does not correspond to inter-personal knowledge flow, or a word of mouth diffusion process, which does represent a knowledge flow.

Despite these limitations, recent studies have shown that patent citations can be used as a proxy for knowledge flows. Jaffe et al. (2000) surveyed a sample of USPTO patent inventors and asked about the extent and mode of their communication with the inventors they cited and about the extent to which patent citation was indicative of this communication. The authors found evidence that a significant fraction of the links indicated by patent citations reflect some kind of spillover. Almost 40% of the inventors surveyed indicated that they learnt about the cited invention either before or during the development of their invention. But in one-third of cases they were unaware of the cited patent, which indicates it was included as a result of the intervention in the citation process of the patent attorney or patent examiner. Duguet and MacGarvie (2005) provide evidence related to the legitimacy of citations in EPO patents as a measure of knowledge flows. Matching a sample of French firms' responses to the European Community Innovation Survey (CIS) with a count of citations made and received by their EPO patents, the authors were able to explore the relationship between patent citations and firms' technology sourcing behaviour. They found that citations are correlated significantly with the way firms acquire and disseminate new technologies. In particular, their results indicate that backward citations, i.e., citations made to other patents by the surveyed firms, were positively and significantly correlated with learning through R&D collaboration, licensing of foreign technology, mergers and acquisitions (M&A) and equipment purchases.

The evidence from these two studies goes some way towards justifying the use of patent citations involving USPTO and EPO patents as a reasonable proxy for knowledge flows, despite the differences that in these two patent systems in terms of the examination process and requirements concerning description of the state of the art.

In the USPTO the inventor and his/her attorney are obliged to provide a list of those references describing the state of the art that are considered relevant to the patentability of the invention – the so called 'duty of candour' – and non-compliance with this requirement is interpreted as fraud and can be grounds for invalidating the patent.[4] Nevertheless, applicants to the USPTO might be very strategic about what prior art to disclose since they might be awarded broader patents if certain prior art material is not considered by the patent examiner (Sampat, 2005; Hedge and Sampat, 2005).

In contrast, the EPO has no requirement similar to the duty of candour (Akers, 2000; Meyer, 2000; Michel and Bettels, 2001). Rule 27(1)(b) of the European Patent Convention requires that the description in a European patent application should 'indicate the background art which, as far as known to the applicant, can be regarded as useful for understanding the invention, for drawing up the European search report and for the examination, and, preferably, cite the documents reflecting such art'. However, Akers (2000) argues there are a number of reasons why inventors would include all prior art documents in their EPO applications. For example, applicants might want all relevant documents to be considered by the examiner to avoid future patent objections being filed by a third party. Similarly, should the patent be enforced in court or have its validity challenged, the applicant might derive stronger bargaining power from having all pertinent prior art considered during the examination procedure.[5] Akers (2000, p. 314) reports that 'many applicants take the time and trouble to disclose the most relevant prior art and discuss the relevance of its disclosure to the invention being claimed'. However, in EPO patents, it is the examiners rather than the inventors or applicants, that add the majority of patent citations. The obvious implication is that in the EPO system more often than in the USPTO system, inventors are more likely to be unaware of the patents that are (ultimately) cited in their patents.

The different legal requirements of the two systems also imply that there are significant differences in the number of citations in the patents[6] and their technological relevance. As pointed out by Michel and Bettels (2001, p. 192), applicants to the USPTO

> rather than running the risk of filing an incomplete list of references,. . .they tend to quote each and every reference even if it is only remotely related to what is to be patented. Since most US examiners apparently do not bother to limit the applicants' initial citations to those references which are really *relevant in respect to patentability*, this initial list tends to appear in unmodified form on the front page of most US patents. (emphasis added).

On the other hand, in the EPO citations are strictly related to descriptions of prior art relevant to the patentability of the invention (Michel and Bettels, 2001).

Thus, EPO citations, although fewer in number, may be less 'noisy' than USPTO citations, since it can be assumed that they have been scrutinised and chosen by the patent examiner, and citing-cited patent pairs might be 'closer' both in time and technological content than those extracted from the USPTO (Breschi and Lissoni, 2004). Moreover EPO citations might be broader in scope because patent examiners do not limit their search to prior art written in English and/or to patents issued by one particular patent office. Michel and Bettels (2001) found that, while 90% citations in USPTO patents are to other USPTO patents, in EPO patents contain citations to patents from a wide range of patent offices: 23.3% EPO patents, 30.9% USPTO patents, 16.3% WIPO patents, 13.1% German patents, 6.2% British patents, 5.2% Japanese patents, and 5% other patents. This bias against foreign patents in the USPTO was explored in more detail by Sampat (2005). In this study the author found that references to foreign patents are 27% less likely to be added by the patent examiners. On the basis of these findings the author suggests that patents granted by the USPTO might be of lower quality if they cover technological fields where most prior art is not contained in US patents.

Another important difference between the EPO and the USPTO systems is that in European search reports, cited documents are classified by the patent examiner within a particular citation category according to their relevance. Table 1 reports these citation categories. As explained by Schmoch (1993), in assessing the novelty of patent applications the examiner searches for earlier documents which have the same or almost the same features as the patent concerned. Thus, there are two important types of citations: documents of particular relevance which restrict the claims of the inventor (citation categories X and Y); and references related to technological background (citation category A). Citations in category X are those that 'already show essential features of the invention or at least question the inventive step of these features if taken alone'

[4] US patent law 37 C.F.R. 156 establishes that 'each individual associated with the filing and prosecution of a patent application has a duty of candour and good faith in dealing with the (US Patent) Office, which includes a duty to disclose to the Office all information known to that individual to be material to patentability. . . no patent will be granted on an application in connection with which fraud on the Office was practiced or attempted or the duty of disclosure was violated through bad faith or intentional misconduct'.

[5] A similar motivation can be advanced for the inclusion of citations by the applicant in US patents: 'citing more prior art will make a patent more valuable in litigation, as it is much harder to prove a patent is invalid if the patent office has already considered it and rejected the relevant prior art' Allison et al. (2004).

[6] The US patent office cites 3 times as many patents as the EPO (Michel and Bettels, 2001).

Table 1
Description of category of citations

Category of citations	Description	Fraction of all citations	Fraction of all EPO-to-EPO citations	Fraction of all EPO-to-EPO inventor citations
X	Particularly relevant documents if taken alone; citations classified under this category are such that when taken alone a claimed invention cannot be considered novel or cannot be considered to involve an inventive step	0.20	0.22	0.13
Y	Particularly relevant documents if combined with another document, such a combination being obvious to a person skilled in the art	0.16	0.16	0.17
A	Documents defining the state of the art and not prejudicing novelty or inventive step	0.62	0.60	0.69
D	Documents cited in the application.	0.09	0.11	-
P	Intermediate documents; Documents published on dates falling between the date of filing of the application being examined and the date of priority claimed	0.04	0.07	0.04
E	Earlier patent documents, but published on, or after the filing date	0.01	0.02	0.00
O	Documents that refer to a non-written disclosure	0.00	0.00	0.00
T	Documents relating to the theory or principle underlying the invention	0.00	0.00	0.00
L	Documents cited for other reasons	0.00	0.01	0.00

Source: EPO examination guides lines part B chapter X.

(Schmoch, 1993, p. 194). In other words, the examiner considers that such citations anticipate the claims in the patent application (Akers, 1999). Citations in category Y are considered to question the inventive steps claimed in the patent being examined, when combined with one or more documents in the same category. This implies that Y citations never occur singly.

The X and Y citations are related to some of the strategic incentives that Sampat (2005) and Hedge and Sampat (2005) discuss. They argue that applicants may deliberately leave out certain citations in an attempt to get broader patents. As Sampat (2005) shows, the extent of this strategic behaviour varies among technology fields. Using the EPO dataset, we can control for these strategic motives, at least as far as they are captured by X and Y type citations, which is an important advantage compared with using USPTO data.

Documents in category A describe the state of the art and, according to the patent examiner, are important for assessing the inventive step. Thus, many patents might be cited because they provide a good description of prior art and are used in the technical background of a patent. Finally, and most importantly, category 'D' documents include citations already mentioned in the patent application for which the search is being carried out, i.e., those proposed by the applicant. This is our source of inventor citations. Note that we include only those citations added by the applicant that the examiner deems relevant with respect to the patentability of the invention.

3. Data

Our primary data sources are the EPO database on patent applications (Bulletin CD) and patent citations to other patents within the EPO over the period 1985–2000 (all citations are taken from the EPO REFI database). We complement these data with information from the OECD citations database on patent applications filed under the Patent Cooperation Treaty (PCT) and on equivalent patents (Webb et al., 2004).[7] Our dataset includes citations by EPO patents to patents issued by all national and regional patent offices. However, in our analysis we focus only on EPO-to-EPO citations because it is only for this sample of patents that we have complete information on our independent variables (particularly self-citation, which is an important control variable). Among these citations we distinguish between citations by inventors with addresses in one of the European countries[8] ('within-Europe' sample) and involving inventors residing in the US ('within-US' sample).

The last three columns in Table 1 show the distribution of citations over citation categories, for the entire sample of citations, for the sample of EPO-to-EPO citations, and for the sample of EPO-to-EPO citations added by the inventor. Note that cited patents can be classified into up to three categories (e.g., ADL). For the entire sample, the largest share (62%) of citations describes the state of the art (*A*), followed by particularly relevant documents (X, 20% and Y, 16%). Similar proportions are found in the sample of EPO-to-EPO citations. 9% of all citations in EPO patents are inventor citations (D), but this share goes up to 11% when we restrict our sample to citations to other EPO patents only. All other citation categories are less than 5% of the total. It is interesting that the predominance of A citations is greater in the sample of EPO-to-EPO inventor citations: 69% of all inventor citations are categorised as A, vs. 60% for the total EPO-to-EPO sample. Also there is a smaller proportion of X citations among the sample of EPO-to-EPO inventor citations (13% vs. 22% for the total EPO-to-EPO sample), indicating that inventors have less inclination to cite patents that 'already show essential features of the invention'. This seems to indicate the (expected) tendency for inventors not to cite

[7] In a longer version of this paper available on the web as a working paper, we report in more detail how we combined the REFI and OECD datasets to build our database.

[8] Our countries include the EU-16 plus Norway and Switzerland.

patents that may compromise the novelty of their own patents. In contrast, Y category, which also includes patents compromising novelty but only in combination with other patents, occurs almost as frequently in the sample of inventor citations as in the total sample (17% vs. 16% for the total EPO-to-EPO sample). The 11% of inventor citations in our EPO-to-EPO sample is small compared to the proportion found in USPTO patents by Alcacer and Gittelman (2006): in their study inventor citations represent 60% of all citations. This finding can be explained by the different legal requirements of the two patent offices concerning description of the state of the art.

Figs. 1 and 2 report the distribution of share of inventor citations by the citing patent's priority year and technological field, for the sample of EPO-to-EPO citations. We use the IPC classes of technological fields provided by the Observatoire des Sciences et des Techniques (OST) and the Fraunhofer Institute (FhG-ISI) (see OST, 2002 appendix A5a-1 p. 346).

Two things stand out. First, as shown in Fig. 1, the share of inventor citations has been declining from almost 14% in 1985 to 9% in 2000. Second, there is a quite large variation across fields: more than 20% of citations in organic chemistry patents were added by the inventor, while for information technology patents this share drops to 4%. In general it appears that the share of inventor citations is higher for patents related to chemistry and materials than for patents in other technological fields, in particular semiconductors, telecommunication, audiovisual and IT. This is in line with the findings in Sampat (2004) for a sample of

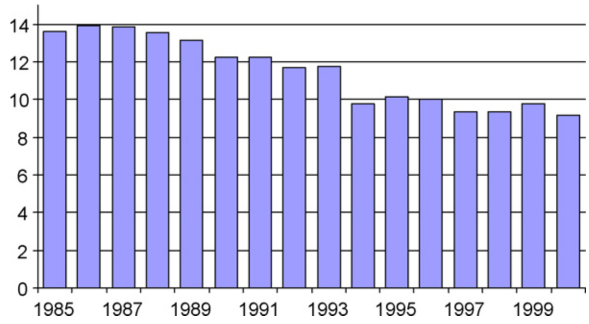

Fig. 1. Temporal pattern in the share of inventor citations.

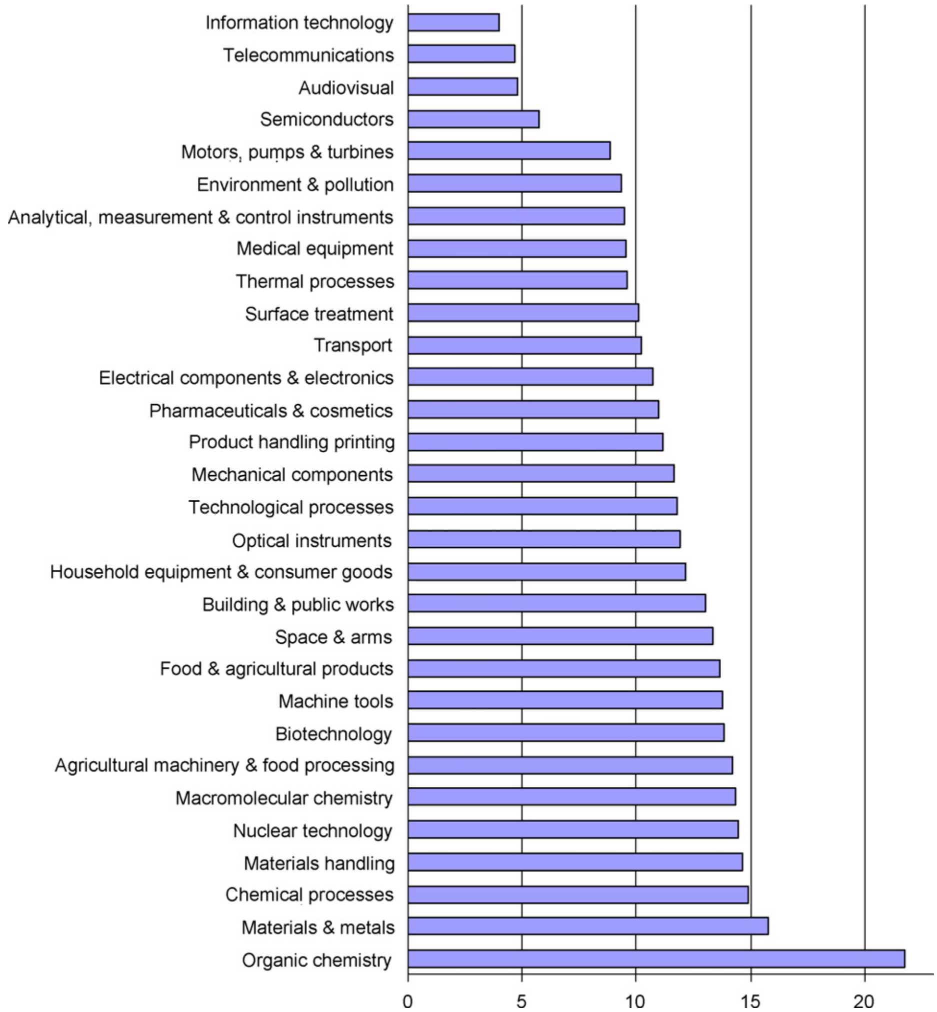

Fig. 2. Share of inventor citations by technological field.

Table 2
Summary statistics

Total sample	
Number of citing patents	700,674
Number of citations	2,859,714
Citations per patent (mean)	3.25
Fraction of citing patents with all citations added by the examiner	75.77
Fraction of citing patents with all citations added by the inventor	2.37
Sample of EPO-to-EPO citations	
Number of citing patents	490,230
Number of citations	982,826
Citations per patent (mean)	1.91
Fraction of citing patents with all citations added by the examiner	81.72
Fraction of citing patents with all citations added by the inventor	7.05

USPTO patents classified in six broader technology fields. Although not directly comparable, our results differ from Sampat's in terms of the proportion of inventor citations in USPTO 'drug and medical' patents, which is higher than for chemical patents. Sampat's interpretation, which could be valid also in the context of EPO citations, is that inventors are more likely to carry out patent searches and to disclose prior art in fields such as chemical and pharmaceuticals, where patent protection is the most important mechanism for appropriating returns from R&D investment.

Table 2 presents further summary statistics. The top part of the table provides information on the total citations database, while the bottom panel gives information for the sample we use in the regressions. Table 2 shows that our sample of EPO-to-EPO citations varies slightly from the total sample. Obviously, the number of citations per patent is lower, while the proportion of patents with only citations added by the examiner, and the fraction of patents with all citations added by the inventor, are higher.

We explore these patterns in more detail by examining the distribution of these shares over time (Fig. 3), and across technology fields (Fig. 4). The trend depicted in Fig. 3 shows that the share of patents with all citations included by the inventor has been constantly declining, from 10% in 1985 to 5% in 2000, while the fraction of patents with all citations added by the examiner has been fairly constant.

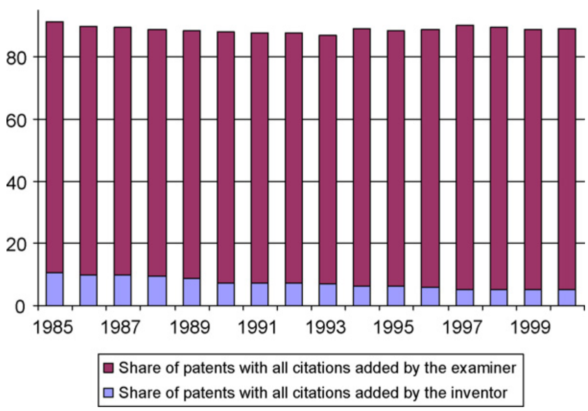

Fig. 3. Temporal pattern in the share of patents with all citations added by the examiner and by the inventor.

Some of the patterns in Fig. 2 are also evident in Fig. 4. In particular, organic chemistry stands out with almost 15% (65%) of patents with all citations added by the inventor (examiner), while the share of information technology patents where inventors (examiners) inserted all citations is only 2% (93%). Thus, our descriptive evidence indicates that there are some notable differences across technology fields in terms of both share of inventor citations and proportion of patents with all citations added by inventor or by examiner, but that there is little temporal variation.

4. Econometric approach

As discussed in the introduction, we want to investigate the factors that explain whether a potential knowledge flow (spillover) actually occurs. The dependent variable is citation type (examiner or inventor). This is a binary variable that is equal to 1 if the citation was added by the inventor. A zero value indicates that the potential knowledge flow did not occur because the examiner and not the inventor provided the citation, a value of 1 indicates that the potential knowledge flow did occur.

With regard to the explanatory factors, our first hypothesis is that the probability of a spillover taking place is higher when the *geographical distance* between the two parties decreases. Alcacer and Gittelman (2006) found mixed evidence in this respect, i.e., they found that geography is only relevant for explaining inventor citations when both cited and citing patent are within the US; in addition, their statistical results differed according to the types of geographic indicators. Thompson (2006) found more unequivocal evidence that patent citations added by the inventor are more geographically concentrated than those added by the examiner, and that this effect is particularly strong when short distances are involved (e.g., within-state or even closer), and is non-existent when geographical coverage extends beyond US borders.

Our main variable for geographical proximity is a standardised measure of regional distance in kilometres (Distance (km)) between the regions of the citing and cited inventors. We calculate this for EU-regions, and for US States, but not between Europe and the US. The variable is standardised to units of 173 km, which is the average distance between European regions in our sample. In terms of the effect of localisation on knowledge spillovers, we expect Distance (km) to be negatively correlated with inventor citations (i.e., the shorter the distance between two patents, the more likely it is that inventors actually include a citation).

In the distance calculations, assignment of patents to a region is based on the inventor's address. However, a single patent may have more than one inventor, and if these inventors are located in different regions, the distance between them is not unambiguous. We followed Alcacer and Gittelman (2006) by using the minimum distance between any citing-cited inventor pair, but our results are also robust to using average distance for the citation pairs.

Although our initial estimations are limited to within-EU and within-US, we also use a discrete distance variable that tests for the relevance of overseas distances. Alcacer and Gittelman (2006) found that geographic distance was

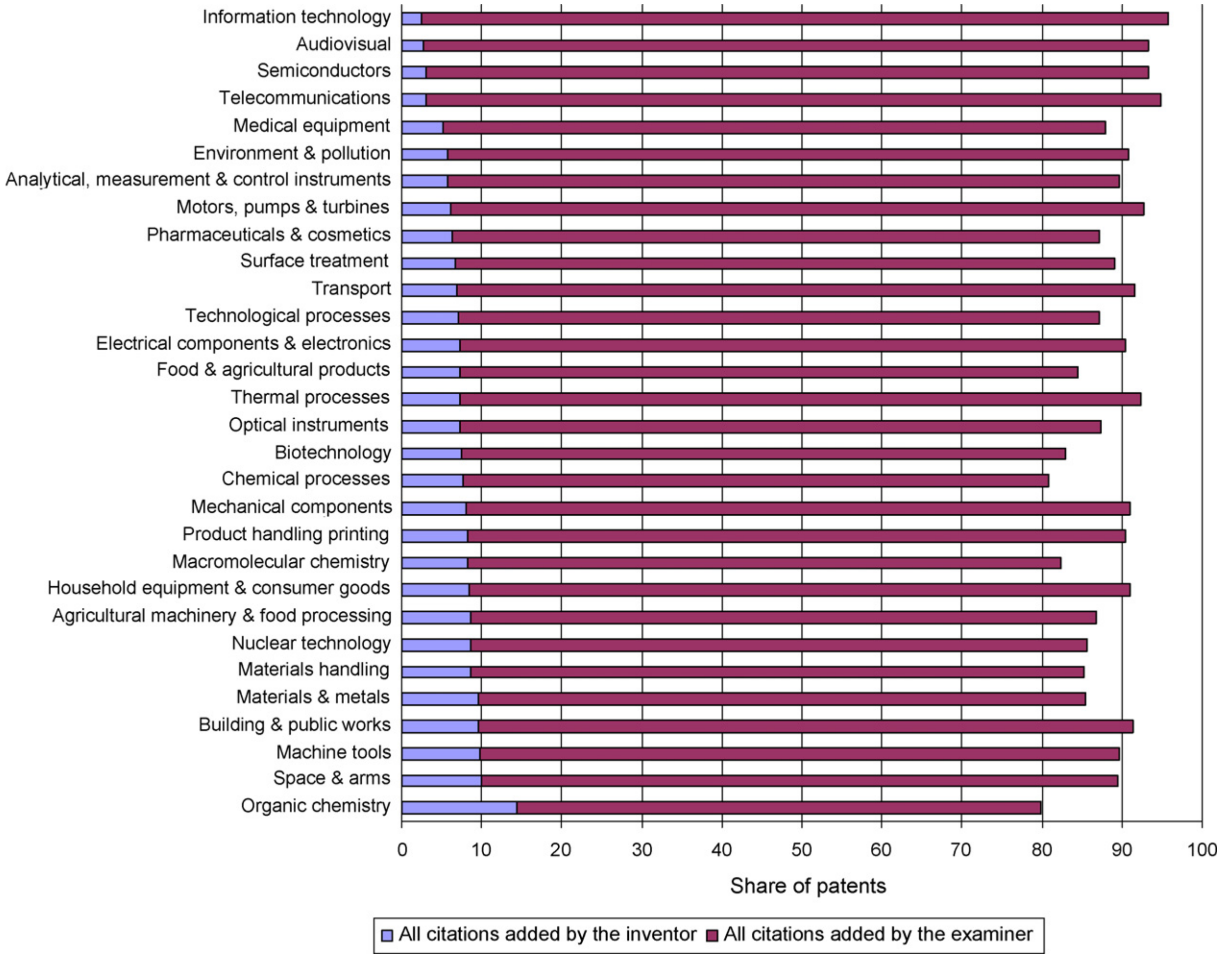

Fig. 4. Share of patents with all citations added by the examiner and by the inventor by technological fields.

no longer significantly correlated with inventor citations when citations were included that spanned beyond the US space. Because we apply a finer distance grid to the EU, our discrete distance variable has different scales for the EU and the US. In the EU, we define *KMclEU* as 0 for minimum distances between cited and citing inventors in the range 0–100 km, 1 for distances 100–250 km, 2 for 250–500 km, 3 for distances within Europe over 500 km, and 4 for citations where the citing inventor(s) is in Europe and the cited inventor(s) is not.[9] For the US space, we have a similar definition, *KMclUS* = 0 for cited and citing inventors less than 100 km apart, 1 in the range of 100–650 km, 2 for 650–1500 km, 4 for distances within the US over 1500 km and 4 for citations where the citing inventor(s) is in the US and the cited inventor(s) is outside the US.

KMclEU and *KMclUS* are obviously ordinally scaled, and this makes their use as an independent variable somewhat complicated. We include each of the values of *KMclEU* or *KMclUS* as a dummy variable, using the 0 class as the reference. Hence, we would expect that the dummies associated with the lower classes (small distances) yield relatively higher probabilities of inventor citations than those associated with the higher classes (larger distances).

Our next hypothesis is that the amount of time that has elapsed between the two patents increases the probability of a knowledge flow occurring. Temporal distance is measured by the variable Citation lag (in years), which is the time that has elapsed between the priority dates of the citing and cited patents. The idea behind this is that over time, an invention will become more well known and, hence, the probability of a spillover will increase (see also Verspagen and Schoenmakers, 2004).

Cognitive similarity is the subject of the next hypothesis, which states that a spillover is more likely to take place if the two inventions come from similar knowledge domains. This measure (*Same technology*) is a dummy variable equal to 1 if the citing and cited patent are classified in the same 4-digit IPC class, and zero otherwise.[10] If the geographic

[9] If a cited patent has one inventor in Europe and one outside Europe, we assign it to one of the categories 0–3 (rather than 4). This is consistent with using minimum distance within Europe.

[10] We also experimented with technological similarity at higher levels of IPC aggregation, but this did gave no additional insights (results available on request).

distance and the cognitive similarity variables are both significant, this is evidence of a localisation effect in addition to geographical concentration of R&D activities of a specific kind.

Self-citations, i.e., those where the citing and cited firm (or even the inventor) are the same, are a special case within cognitive similarity. It could be argued that in such cases, cognitive similarity is very strong, although a degree of cognitive distance will always be present, especially if the citing and cited firms, but not the inventor, are the same. In this case, the *Same Technology* variable does account for the full degree of cognitive similarity, and we need also to control for self-citation. Unfortunately, we cannot do this in a perfect way, because of the large number of patents in our dataset. Many patents are applied for under the names of subsidiaries and divisions that are different from those of the parent companies. In addition the names of companies are not unified, in the sense that the same company may appear several times in the data, but with a slightly different name in each case. This requires that self-citations would need to be identified manually, which is not feasible given the more than 360,000 citation pairs involved in some of the regressions.

But we can control for self-citation for two subsets of patents in our sample.[11] The first includes patents owned by 169 high-tech multinational enterprises (MNEs) listed in the Fortune 500 in 1997, mostly American, European and Japanese companies. These patents were consolidated at the level of the firm, using the Dun and Bradstreet Linkages 'Who Owns Whom' (1998) database, which contains 1997 group ownership structures.[12] We consolidated the patents for the complete period considered, although this was on the basis of the 1997 Dun and Bradstreet Linkages data. To control for self-citation we defined a dummy variable (*Same_MNE*) that is 1 if the cited and citing patents are owned by the same MNE, and 0 otherwise.

The second sub-sample includes patents owned by 2197 publicly listed European companies as described in Thoma and Torrisi (2007). The consolidation of patents under the names of the parent companies was based on information contained in Bureau Van Dijk's Amadeus dataset from 1997 to 2005, and obtained using an approximate matching algorithm, rather than the time consuming manual procedure. As before, we defined a variable *Same EU firm* which is set to 1 if the citing and cited patents are owned by the same company. Because in both sub-samples patent consolidation was undertaken at the end of our sample period, in the regressions where we control for self-citations we include only citing patents applied for between 1993 and 2000.

The dataset assembled by Thoma and Torrisi contains a larger number of firms, mostly smaller sized, than the sample created by Verspagen and Schoenmakers, but it is biased towards European companies. Thus, we believe that by estimating our models for both sub-samples we can test for the impact of self-citation on the realisation of knowledge flows.

We also account for strategic motivations influencing inventor citations (e.g., Sampat, 2004). The first of these is related to the citation categories presented in Table 1. We constructed three mutually exclusive dummy variables capturing the most frequent classes (A, Y and X) other than D, which defines our dependent variable. The remaining categories account for a minor fraction of the patents in our sample (see Table 1), and hence we dropped the citations classified under these categories from the sample. Categories X and Y pose a serious threat to the novelty of the patent and, hence, as observed above, we expect that inventors will be less likely to add citations in these categories, even if a knowledge flow in fact occurred. Thus, by including the X and Y citation types as independent variables, we correct for this potential bias in the dependent variable.

We also controlled for whether the EPO patent was applied for through the World Intellectual Property Organization (WIPO) by including a dummy variable (WO), which is equal to 1 if the EPO patent was filed under the Patent Cooperation Treaty (PCT). International patent applications through WIPO have to abide by the rules set by the PCT, which establish (rule 5) that the applicant should 'indicate the background art which, as far as known to the applicant, can be regarded as useful for the understanding, searching and examination of the invention, and, *preferably*, cite the documents reflecting such art' (emphasis added). Thus, although WIPO is less strict than USPTO in terms of imposing an obligation on the applicant to disclose the prior art, its requirements are more stringent than EPO terms. Equally, WIPO sets rigorous requirements for patent examiners with regard to the extent of the documentation searched to establish relevant prior art (see rule 33 of the PCT). Therefore, citations in EPO patents applied for through WIPO might have a higher percentage of inventor citations than other EPO patent applications not subject to an international applications procedure. Finally we included 30 technological dummies to account for the different citation patterns across technological fields.

Our estimation method is a logit model. Because citation behaviour may be influenced by the personal characteristics of the applicant and/or examiner as well as the specific technology involved in the citing patent, we can expect the error terms in our econometric equation to be correlated between citation pairs involving the same citing patent. In order to deal with the correlated error terms, we follow Alcacer and Gittleman (2004) and apply a random effects panel model, in which the random effects refers to the citing patent, and what normally is the 'time' dimension is represented by the various citations in a given citing patent.[13]

[11] We did not attempt to control for self-citation at inventor level, because of the phenomenal difficulties involved in matching names in such a large dataset. Also, we do not have information on the names of examiners and, hence, cannot control for examiner 'self-citations'.

[12] Verspagen and Schoenmakers (2004) provides more details of the consolidation process, and the names of the companies.

[13] We experimented with alternative logit models, such as a model with clustered errors on citing patents and a complementary log–log model. Based on information criteria (AIC or BIC), we concluded that the random effects logit model was most appropriate for our data. We also ran a logit with fixed effects, and a linear probability model with fixed effect on the citing patent, and obtained qualitatively similar results to those reported here.

Table 3
Descriptive statistics

Variable	Within Europe sample					Within the US sample				
	Inventor citations (n = 50,106)		Examiner citations (n = 219,583)		T-test	Inventor citations (n = 21,677)		Examiner citations (n = 134,908)		T-test
	Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.	
Geographic distance										
Distance (km)	1.422	2.076	2.833	2.714	109.30***	3.456	6.187	7.146	7.891	65.67***
Technology similarity										
Same technology	0.710	0.454	0.681	0.466	−12.75***	0.655	0.475	0.620	0.485	−9.78***
Temporal distance										
Citation lag	5.218	3.524	5.573	3.909	18.68***	4.453	3.054	4.594	3.374	5.78***
Self-citation										
Same MNE[a,b]	0.611	0.488	0.304	0.460	−57.02***	0.688	0.463	0.364	0.481	−41.84***
Same EU Firm[c,d]	0.447	0.497	0.229	0.420	−53.10***	0.539	0.498	0.276	0.447	−28.82***
Citations categories										
Class A	0.718	0.450	0.647	0.478	−30.40***	0.652	0.476	0.550	0.497	−28.12***
Class X	0.108	0.310	0.212	0.409	53.84***	0.157	0.363	0.257	0.437	32.15***
Class Y	0.174	0.379	0.141	0.348	−19.08***	0.191	0.393	0.192	0.394	0.40
Other controls										
Wo	0.113	0.317	0.179	0.383	35.46***	0.145	0.353	0.394	0.489	71.94***

***Significant at 1%.
[a] For the within Europe sample, $N_{inventor}$ = 9268, $N_{examiner}$ = 37,784.
[b] For the within the US sample, $N_{inventor}$ = 4444, $N_{examiner}$ = 28,339.
[c] For the within Europe sample, $N_{inventor}$ = 13,951, $N_{examiner}$ = 58,654.
[d] For the within the US sample, $N_{inventor}$ = 2944, $N_{examiner}$ = 16,301.

5. Results

Table 3 reports the descriptive statistics for the variables in the regression for the sample of within-EPO citations and the sample of within-US citations. In both these samples inventor citations are more co-localised (Distance (km) is smaller) than examiner citations: on average, inventors are more likely to cite local patents and the difference in means is quite large – especially for the within-US sample – and statistically significant. Inventors are also more likely to include citations to patents in the same 4-digit IPC class than examiners, but the difference in means, though statistically significant, is not very large: for the within Europe (US) sample 71% (65%) of inventor citations are to patents in the same technology class, while the corresponding proportion for examiner citations is 68% (62%). This is in line with our expectations about cognitive distance. Also consistent with our expectations we find that inventors show a slightly higher tendency to cite more recent patents than examiners: for the sample of EU (US) patents the citations lag for inventor citations is on average 5.2 (4.4) years while for examiner citations it is 5.5 (4.6) years.

Our findings for the self-citation variables indicate that inventors rather than examiners tend to cite prior patents applied for by their own firm. Table 3 shows that for both samples of within Europe and within the US citations, the *Same MNE* and *Same EU Firm* variables are, on average, twice as large for inventor citations relative to examiner citations. Finally, we find that examiners on average have a higher tendency to include citations to patents that anticipate the claims listed in the patent application (Class X citations), while patents that describe the prior state-of-the-art (Class A) are more likely to be cited by the inventor.

Surprisingly, we also find that citations questioning the inventive steps claimed in the patent when combined with another document (Class Y) are more frequent for inventor citations than for examiner citations. We asked EPO patent examiners to comment on this result, and they indicated that it might be the result of a ‘trigger’ effect: an inventor might cite a patent because it signals a problem (which is also addressed in the citing patent), which prompts the examiner to look for patents that have provided a solution to this problem. When such citations are identified, they are added, along with the original inventor citation, as a Y type. This clearly indicates non-rational expectations on the part of the inventor who added the ‘triggering’ citation, but it does seem to explain the tendency for Class Y to be higher for inventor citations.[14]

To conclude, the descriptive evidence confirms many of our theoretical expectations about citations as a measure of knowledge flows, but it needs to be tested in a multivariate analysis.

We first estimate a number of models for the ‘within Europe’ citations shown in Table 4. We estimate a model for the entire sample (columns 1–4), and for the two sub-samples, which enables us to control for self-citations (columns 5–8 and 9–12). Within each of these three basic samples, we differentiate between four sub-samples: starting from the complete (sub) sample, we exclude citations with citing patents where all citations were added by the examiner, by the inventor, or by both examiner or inventor

[14] By their nature, Y citations do not occur as single citations. Thus, we can identify the ‘trigger’ effect by examining the Y citations where at least one of them is made by the inventor, and at least one is made by the examiner. At a descriptive level, we compared the relative frequency of these ‘trigger’ Y citations to non-Y citations that combine inventor and examiner citations within a single citing source patent. We indeed found that ‘trigger’ combinations are relatively more frequent, which suggests that the trigger effect is real.

Table 4
Odds ratios from random effects logit regressions within-Europe citations

	Entire sample				Sample of MNEs' patents				Sample of EU firms' patents			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Distance (km)	0.733	0.831	0.773	0.831	0.862	0.919	0.888	0.926	0.829	0.893	0.860	0.898
	(90.07)***	(52.46)***	(62.42)***	(41.02)***	(16.84)***	(8.75)***	(11.28)***	(6.41)***	(28.71)***	(16.31)***	(19.78)***	(12.51)***
Citation lag	0.996	0.999	1.005	1.006	1.017	1.019	1.020	1.020	1.006	1.008	1.005	1.005
	(2.26)**	(0.56)	(2.61)***	(2.28)**	(4.48)***	(4.07)***	(4.49)***	(3.88)***	(1.84)*	(2.27)**	(1.44)	(1.28)
Same technology	1.124	1.030	1.100	1.046	1.064	1.024	1.016	1.014	1.110	1.040	1.102	1.060
	(8.21)***	(1.74)*	(5.64)***	(2.22)**	(1.87)*	(0.62)	(0.41)	(0.30)	(3.89)***	(1.28)	(3.18)***	(1.60)
Class X	0.377	0.578	0.483	0.640	0.410	0.562	0.548	0.657	0.380	0.548	0.494	0.620
	(48.94)***	(24.26)***	(31.11)***	(16.37)***	(19.78)***	(11.93)***	(12.10)***	(7.40)***	(26.10)***	(14.94)***	(17.10)***	(10.07)***
Class Y	1.119	1.161	1.194	1.289	1.008	0.965	1.169	1.149	1.066	1.127	1.148	1.253
	(6.30)***	(6.84)***	(8.67)***	(10.00)***	(0.17)	(0.72)	(3.27)***	(2.44)**	(1.80)*	(2.87)***	(3.53)***	(4.74)***
WO	0.556	1.074	0.605	1.038	0.727	1.235	0.717	1.107	0.778	1.190	0.736	1.049
	(28.42)***	(2.90)***	(21.01)***	(1.28)	(7.79)***	(4.66)***	(7.40)***	(1.90)*	(7.80)***	(4.82)***	(8.56)***	(1.12)
Same MNE					3.065	2.159	2.440	2.110				
					(28.81)***	(18.46)***	(20.53)***	(15.02)***				
Same EU firm									2.346	1.854	1.951	1.778
									(29.16)***	(18.74)***	(20.59)***	(15.00)***
Observations	269,689	78,928	242,074	51,313	47,052	15,111	42,322	10,381	72,605	23,136	65,608	16,139
Number of groups	170,443	42,499	146,804	18,860	28,411	7,633	24,505	3,727	43,328	11,583	37,499	5,754
Log-likelihood	−117290.9	−49716.81	−69400.9	−34103.47	−20465.55	−9541.39	−12950.96	−6830.63	−31952.83	−14775.91	−20415.31	−10644.79
Avg cited per citing	1.58	1.86	1.65	2.72	1.66	1.98	1.73	2.79	1.68	2	1.75	2.8
Max cited per citing	23	23	23	23	23	23	23	23	23	23	23	23
Wald Chi 2	12935.46	3933.13	7674.95	2174.01	2966.93	1009.93	2060.71	581.82	3933.29	1428.84	2634.91	803.31
Degree of freedom	35	35	35	35	36	36	36	36	36	36	36	36
Rho	0.33	0.01	0.13	0.01	0.33	0.01	0.12	0.01	0.31	0.01	0.12	0.01
Chi bar 2	3801.1	63.07	580.7	258.47	757.85	8.99	109.14	45.26	1071.55	20.53	153.72	74.54

Models 1,5,9 include all citing patents in the respective samples. Model 2,6,10 exclude citing patents with all citations added by the examiner, Models 3,7, 11 exclude citing patents with all citations added by the inventor, Models 4,8, 12 exclude those citing patents with all citations added by the inventor and those with all citations added by the examiner, absolute value of z statistics in brackets, *significant at 10%; **significant at 5%; ***significant at 1%.

(in the last case, we have only patents where examiners and inventors added citations). We present the results for these sub-samples in order to check for robustness with regard to potential fixed effects related to individual examiners or inventors.

In all 12 equations, we find that a smaller geographical distance increases the probability of inventor citations or, in other words, that geographical closeness increases the probability of a knowledge flow occurring. This is shown by the odds-ratios for the variable Distance (km), which is always smaller than 1, and significantly so. Excluding citing patents with a single source (examiner or inventor) of citations leads to somewhat higher odds ratios on Distance (km), but these are still clearly below 1.

Our expectation with regard to the time lapse between cited and citing patents is mostly, but not always confirmed. 8 out of 12 of the equations in Table 4 show an odds ratio for this variable that is significant and larger than 1 (as expected). This indicates that, ceteris paribus, time increases the likelihood of a spillover occurring. However, the odds-ratios of this variable (*Citation lag*) is very close to 1, indicating that the time effect is small. Column 1 shows an odds ratio smaller than 1 and significant, i.e., contrary to our expectations.

Next, we test the hypothesis that inventors cite more within the technology class (*Same technology*), i.e., that cognitive distance has a negative effect on knowledge flows. The odds ratio for this variable is larger than 1 and significant in all four equations for the total sample, although we observe that this effect is somewhat smaller if we exclude citing patents with a single citation source (examiner or inventor). For the MNE sample, the effect is much weaker, although all odds ratios are larger than 1. Only the equation for the complete MNE sample is (weakly) significant. In the EU firms sample, again, all odds ratios are larger than 1, but only two of the four equations are significant.

With regard to strategic factors, the results confirm that examiners are more likely to add the 'dangerous' citation type X than the 'common' citation type A, which is the reference category. This effect is sizeable, as the deviation from 1 in the odds ratio indicates. For example, in equation 1 in Table 4, the effect of an X citation is comparable to an increase of 400 km distance.[15] As the descriptive evidence suggests, however, inventors are generally more likely to add citations type Y than citations type A. Only two equations in the MNE sample are not significant for this variable. This suggests that the trigger effect is the main explanatory factor in the Y type citations.

We found mixed evidence for the international application dummy (WO). In half of the equations this has an odds ratio greater than 1 (our a priori expectation); in the other half, the odds ratio is smaller than 1, and often sizeably so (in equation 1 it is comparable to the effect of a distance of some 285 km). Interestingly, odds ratios greater than 1 are found only in cases where we exclude citing patents, and where all citations are added by the examiner. This shows the importance of controlling for this factor; but the WO effect does not seem greatly to influence the other parameter estimates.

With regard to self-citations, the two sub-samples (equations 5–8 and 9–12) confirm that many of the general findings for the total sample (columns 1–4) also hold for the smaller subsets of firms, especially for geographic effect, although there are obviously some differences with regard to the numeric values of the estimated parameters. This indicates that the mechanisms governing spillovers do not differ qualitatively between the different types of firms in these samples (large MNEs, small and large firms). The cognitive distance variable is the most obvious exception to this. It also indicates that whether or not we control for self-citations does not influence the results in a qualitative way. With regard to the self-citation variable, the results in Table 4 confirm our expectation that these are much more likely for inventor citations than for examiner citations (i.e., inventors are more likely to cite their own firm's patents). The self-citation effect is somewhat stronger in the MNE sample compared to the more general EU firms sample.

The results in Table 4 refer to citing and cited patents invented in Europe. We want to test whether these results, particularly for geography, hold for the US space. Consequently, we present estimations for the 'within-US' sample of EPO citation pairs in Table 5. In this case, our geographical units are US states (we exclude Alaska and Hawaii, which are geographic outliers), which tend to be larger than European regions, and this may affect our results. In particular, we might expect that the effect of distance would be smaller, since we are using bigger minimum distances.

Table 5 presents the estimations for the same 12 samples as in Table 4. The first row confirms the geography effect: all odds ratios are smaller than 1 and significant. However, as expected, the geographic effect is somewhat smaller than for the European space (the odds ratios are closer to 1, in some cases only a couple of percentage points difference). This is consistent with previous studies of USPTO patents (Alcacer and Gittelman, 2006; Thompson, 2006). The results in Table 5 also confirm the (small) effect of time: all odds ratios for the citation lag variable are larger than 1 and significant.

On the other hand, the effects for cognitive similarity (*Same technology*) are mostly insignificant, and where they are significant, the odds ratios are smaller than 1, which is contrary to our expectations. This is particularly the case in the last three columns, i.e., for the EU firms sample where we are able to control for self-citations. This suggests that the negative findings for the *Same Technology* variable are related to citations made by US-affiliates of European firms.

The results for the Class X variable and the WO and self-citations variables are again as expected. For the X citations in equation 1, the effect is comparable to an increased distance of about 1200 km, which also confirms the smaller geographic effect in the US compared to Europe. The Class Y variable is less often significant than in Table 4, but when it is, it maintains its sign (>1). The odds ratios for the WO variable are more often below 1 (contrary to our expectations).

[15] To calculate this effect, it is necessary to know that 1 minus the odds ratio for Distance (km) corresponds to the effect of 173 km (the unit of Distance (km)). 1 minus the odds ratio of a dummy variable, e.g. *ClassX*, corresponds to the effect of the dummy variable taking the value 1.

Table 5
Odds ratios from random effects logit regressions within-US citations

	Entire sample				Sample of MNEs' patents				Sample of EU firms' patents			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Distance (km)	0.923	0.949	0.936	0.936	0.957	0.974	0.971	0.971	0.945	0.972	0.949	0.949
	(49.26)***	(30.17)***	(32.97)***	(32.97)***	(10.56)***	(5.94)***	(6.13)***	(6.13)***	(10.99)***	(5.42)***	(8.52)***	(8.52)***
Citation lag	1.01	1.023	1.011	1.011	1.03	1.04	1.024	1.024	1.026	1.034	1.017	1.017
	(3.35)***	(5.93)***	(3.18)***	(3.18)***	(4.40)***	(4.74)***	(3.12)***	(3.12)***	(3.34)***	(3.67)***	(1.96)**	(1.96)**
Same technology	1.015	0.97	0.971	0.971	1.006	1.019	0.971	0.971	0.908	0.891	0.866	0.866
	(0.72)	(1.21)	(1.21)	(1.21)	(0.12)	(0.34)	(0.52)	(0.52)	(1.62)	(1.73)*	(2.18)**	(2.18)**
Class X	0.475	0.64	0.6	0.6	0.369	0.559	0.509	0.509	0.42	0.636	0.574	0.574
	(27.92)***	(14.58)***	(16.40)***	(16.40)***	(15.51)***	(8.35)***	(9.34)***	(9.34)***	(11.42)***	(5.55)***	(6.74)***	(6.74)***
Class Y	0.975	1.09	1.018	1.018	0.907	1.051	0.983	0.983	1.071	1.121	1.229	1.229
	(0.95)	(2.71)***	(0.59)	(0.59)	(1.53)	(0.68)	(0.23)	(0.23)	(0.90)	(1.33)	(2.53)**	(2.53)**
WO	0.208	0.926	0.268	0.268	0.463	1.135	0.408	0.408	0.411	1.009	0.446	0.446
	(57.15)***	(2.33)**	(41.14)***	(41.14)***	(11.56)***	(1.77)*	(11.70)***	(11.70)***	(11.92)***	(0.11)	(10.13)***	(10.13)***
Same MNE					4.041	2.31	3.073	3.073				
					(24.12)***	(13.45)***	(17.06)***	(17.06)***				
Same EU firm									2.974	2.156	2.283	2.283
									(16.88)***	(11.09)***	(11.87)***	(11.87)***
Observations	156,585	34,055	144,435	144,435	32,788	6,992	30,323	30,323	19,245	4,795	17,701	17,701
Number of groups	96,089	18,207	85,845	85,845	20,030	3,655	18,002	18,002	11,296	2,387	10,059	10,059
Log-likelihood	−55292.64	−21476.39	−32063.95	−32063.95	−11101.01	−4320.75	−6501.72	−6501.72	−7215.18	−3034.61	−4447.65	−4447.65
Avg cited per citing	1.63	1.87	1.68	1.68	1.64	1.91	1.68	1.68	1.7	2.01	1.76	1.76
Max cited per citing	17	17	17	17	14	14	14	14	14	14	14	14
Wald Chi 2	7773.92	1602.09	3941.79	3941.79	1743.7	493.96	994.33	994.33	928.47	297.45	622.13	622.13
Degree of freedom	35	35	35	35	36	36	36	36	36	36	36	36
Rho	0.38	0.01	0.17	0.17	0.46	0.01	0.24	0.24	0.43	0.01	0.18	0.18
Chi bar 2	2772.67	17.98	527.11	527.11	828.49	0.55	175.15	175.15	491.12	1.51	82.73	82.73

Models 1, 5, 9 include all citing patents in the respective samples, Models 2, 6, 10 exclude citing patents with all citations added by the examiner, Models 3, 7, 11 exclude citing patents with all citations added by the inventor, Models 4, 8, 12 exclude those citing patents with all citations added by the inventor and those with all citations added by the examiner, absolute value of *z* statistics in brackets, *significant at 10%; **significant at 5%; ***significant at 1%.

Table 6
Odds ratios from random effects logit model for the sample of within the US patents owned by EU firms

	Sample of self-citations				Sample with self-citations excluded			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance (km)	0.954	0.983	0.964	0.964	0.945	0.970	0.947	0.947
	(4.23)***	(1.40)	(3.21)***	(3.21)***	(9.76)***	(5.04)***	(7.52)***	(7.52)***
Citation lag	1.062	1.060	1.037	1.037	1.011	1.023	1.006	1.006
	(4.20)***	(3.36)***	(2.58)***	(2.58)***	(1.13)	(2.00)**	(0.50)	(0.50)
Same technology	0.700	0.639	0.762	0.762	1.078	1.106	0.959	0.959
	(3.63)***	(3.88)***	(2.79)***	(2.79)***	(0.96)	(1.17)	(0.46)	(0.46)
Class X	0.443	0.640	0.615	0.615	0.403	0.632	0.559	0.559
	(6.80)***	(3.49)***	(4.11)***	(4.11)***	(8.94)***	(4.18)***	(5.02)***	(5.02)***
Class Y	1.118	1.227	1.270	1.270	0.979	1.012	1.188	1.188
	(0.92)	(1.46)	(2.05)**	(2.05)**	(0.21)	(0.11)	(1.49)	(1.49)
WO	0.387	0.988	0.470	0.470	0.424	1.049	0.416	0.416
	(7.81)***	(0.09)	(6.42)***	(6.42)***	(9.27)***	(0.47)	(8.07)***	(8.07)***
Observations	6089	2211	5248	5248	13156	2584	12453	12453
Number of groups	4412	1466	3724	3724	8663	1548	8051	8051
Log-likelihood	−3218.16	−1256.86	−2028.62	−2028.62	−3944.47	−1730.55	−2358.55	−2358.55
Avg cited per citing	1.38	1.51	1.41	1.41	1.52	1.67	1.55	1.55
Max cited per citing	12	12	10	10	11	10	11	11
Wald Chi 2	197.18	102.49	141.04	141.04	452.52	103.43	285.63	285.63
Degree of freedom	35	35	35	35	35	35	35	35
Rho	0.49	0.04	0.18	0.18	0.40	0.01	0.19	0.19
Chi bar 2	215.05	1.55	35.4	35.4	213.55	0.53	47.75	47.75

Models 1 and 5 include all citing patents in the respective samples, Model 2 and 6 exclude citing patents with all citations added by the examiner, Models 3 and 7 exclude citing patents with all citations added by the inventor, Models 4 and 8 exclude those citing patents with all citations added by the inventor and those with all citations added by the examiner, absolute value of *z* statistics in brackets, *significant at 10%; **significant at 5%; ***significant at 1%.

Given the fact that the estimations in Table 5 do not confirm our prior expectations about cognitive similarity (*Same technology* variable), and that this appears to be driven largely by the sub-sample of US-affiliates of EU firms, we investigate this sub-sample (i.e., the Thoma and Torrisi sample) in somewhat more detail (Table 6). In particular, we split the sample into self-citations and non-self-citations samples. This allows us to investigate whether the self-citations effect is purely additive, or is also partly multiplicative with respect to the other variables.

The results in Table 6 clearly show that the negative finding for cognitive similarity is driven mostly by self-citation among the firms. The odds ratios for *Same technology* are all smaller than 1 and significant for the self-citations sub-sample, while there are no significant estimates for *Same technology* in the non-self citations part of this sample. This may be an indication of so-called asset-seeking behaviour, i.e., the US-affiliates of EU firms are active in fields that are different from rather than similar to the activities of the European home-base (see e.g. Kuemmerle, 1996; Criscuolo et al., 2005). Future research should investigate the role of this effect in inventor vs. examiner citations, in more detail.

The results in Table 6 also confirm that the distance effect is relevant even at the intra-firm level, i.e., between US-affiliates of EU firms. In columns 1–4 of Table 6, i.e., for self-citations only, we find a significantly negative effect of Distance (km). Although we do present these results here for reasons of space, they hold for the within-EU citations in the Thoma and Torrisi sample and the Verspagen and Schoenmakers MNE sample (for both the within-EU and within-US samples).

Our estimations so far have not addressed the effect of long distances, i.e., we have excluded overseas citations (i.e., EU- or US-based inventors citing overseas patents). Our final set of estimations makes use of the *KMclEU* and *KMclUS* variables to test for these long distances. For reasons of space, Table 7 shows the estimations for the sample that includes only citing patents with both inventor and examiner added citations (the results for the other samples were qualitatively similar for the geographic distance variable).

It should be remembered that the *KMcl* = 0 class is the reference, i.e., if geographical distance is an impediment to inventors' citing a patent, we would expect the odds ratios for the other *KMcl* classes would be <1. This is indeed always the case, for both US space and European space, and for the total sample as well as for the MNE and EU firms sub-samples. In particular, the *KMcl* = 4 dummy is always significant, which means that overseas inventors are, *ceteris paribus*, less likely to be cited than proximate inventors. As before, MNEs appear to be less influenced by distance, i.e., the odds ratios are higher (but < 1) in the MNE sample. The EU firms sample has odds ratios between those of the MNE and the total samples.

If longer distances are a stronger impediment to knowledge flows, we would expect the dummies associated with higher *KMcl* values to be smaller. In the EU-citing inventor sample, this holds very clearly up to and including the *KMclEU* = 3 dummy. But the difference between the odds ratio of *KMclEU* = 3 and *KMclEU* = 4 is very small in this case, which indicates that an overseas distance has about the same effect as a large distance within Europe. In the US-citing inventor sample, the *KMclUS* = 4 dummy is much lower than any of the others, which indicates that overseas distances from the US matter more (in a relative sense) compared to Europe. But, here, we find less clear differences between the *KMclUS* = 2 and *KMclUS* = 3 dummies, especially so for the MNE and EU firms sample.

Table 7
Odds ratios from random effects logit model with discrete distance variables

	Sample of EU-based citing inventors			Sample US-based citing inventors		
	Total	MNEs	EU firms	(5)	(6)	(7)
KMcl = 1	0.532	0.811	0.666	0.481	0.716	0.622
	(22.61)***	(3.39)***	(7.98)***	(18.06)***	(3.41)***	(4.39)***
KMcl = 2	0.414	0.717	0.605	0.416	0.654	0.597
	(35.20)***	(5.26)***	(10.53)***	(22.20)***	(4.31)***	(4.85)***
KMcl = 3	0.31	0.617	0.47	0.373	0.747	0.592
	(51.57)***	(7.72)***	(16.40)***	(27.44)***	(3.24)***	(4.77)***
KMcl = 4	0.286	0.613	0.44	0.227	0.424	0.399
	(61.50)***	(8.53)***	(19.10)***	(53.53)***	(11.28)***	(11.16)***
Citation lag	1.015	1.02	1.013	1.048	1.05	1.045
	(7.58)***	(4.86)***	(3.96)***	(14.63)***	(7.44)***	(5.98)***
Same technology	1.053	1.126	1.07	0.983	1.019	0.875
	(3.19)***	(3.41)***	(2.32)**	(0.75)	(0.37)	(2.28)**
Class X	0.66	0.679	0.631	0.649	0.618	0.687
	(19.10)***	(8.60)***	(11.94)***	(14.83)***	(7.47)***	(5.12)***
Class Y	1.239	1.186	1.233	1.074	1.042	1.103
	(10.52)***	(3.84)***	(5.49)***	(2.47)**	(0.65)	(1.30)
WO	1.067	1.055	1.072	0.899	1.026	0.981
	(2.73)***	(1.27)	(2.00)**	(3.62)***	(0.40)	(0.28)
Same MNE		2.21			2.519	
		(16.87)***			(14.29)***	
Same EU firm			1.832			2.149
			(16.06)***			(10.59)***
Observations	82,570	17,537	25,681	40,353	8,708	6,153
Number of groups	29,201	6,077	8,933	13,836	2,954	2,050
Log-likelihood	−53114.03	−11196.67	−16405.33	−25349.52	−5354.1	−3849.78
Avg cited per citing	2.83	2.89	2.87	2.92	2.95	3
Max cited per citing	21	21	21	20	16	14
Wald Chi 2	5160.48	1157.81	1626.01	3235.86	894.35	493.02
Degree of freedom	38	39	39	38	39	39
Rho	0.01	0.01	0.01	0.01	0.01	0.01
Chi bar 2	326.94	61.47	95.58	136.71	25.71	18.29

All models exclude those citing patents with all citations added by the inventor and those with all citations added by the examiner, absolute value of *z* statistics in brackets, *significant at 10%; **significant at 5%; ***significant at 1%.

Overall, and contrary to the findings in Alcacer and Gittelman (2006), the results for our discrete distance variable show that the results in Tables 4 and 5 are completely robust to the inclusion of overseas distances, as well as the ordinal measurement scale. This means that inventor citations in the EPO system are more sensitive to distance than examiner citations, suggesting that geographical distance is a strong factor working against knowledge flows.[16]

6. A closer look at the effect of distance

In Tables 4–6, we (implicitly) assumed that the effect of distance is linear, but it might be the case that the relation between knowledge flows and distance is non-linear. In particular, and in line with some of the results in Table 7, we would expect that at small distances, an increase in distance by one unit (1 km) would lead to a stronger effect of the likelihood of an inventor citation, than the same increase over a longer distance. In order to test for this, we employ a non-parametric method that starts by eliminating the effect of variables other than distance from the likelihood of an inventor citation (Hausman and Newey, 1995). To this end, we first estimate a random effects linear probability regression model, with our usual dependent variable and using the same independent variables as in Table 4. We calculated a residual from this regression as $r_i = e_i - \hat{e}_i$, where e is our dependent variable, and $\hat{e}_i = \hat{c} + \hat{\beta}X_i + \delta_i$. Here c and β are the parameters in our linear model, X is the vector of the independent variables with the exception of Distance (km), δ is the random effect associated with the citing patent, and hats indicate estimated values. Note that the regressions from which $\hat{c}$ and $\hat{\beta}$ were derived included Distance (km) as an independent variable, but we do not include this variable in the calculation of the residual r. Hence r 'partials out' all variables except distance from the dependent variable (inventor citations).

Next, we run a locally weighted regression (*lowess*) of r on Distance (km) (we use a bandwidth of 0.8). This regression yields a smooth curve on which each point corresponds to the 'local' (for the value of Distance (km)) effect of distance on the likelihood of an examiner citation. Fig. 5 depicts the results of this procedure for within-Europe.[17] Instead of the version of Distance (km) that is standard-

[16] Following Alcacer and Gittelman (2006), we also experimented with dummy variables for whether or not citing and cited inventors are in the same EU-region or US-state, the same country, and the same continent. For reasons of space, we do not document these results, but the findings entirely confirmed those reported here.

[17] We applied other methods to assess the potential non-linear nature of the distance relationship, e.g., a step-function for Distance (km), a linear

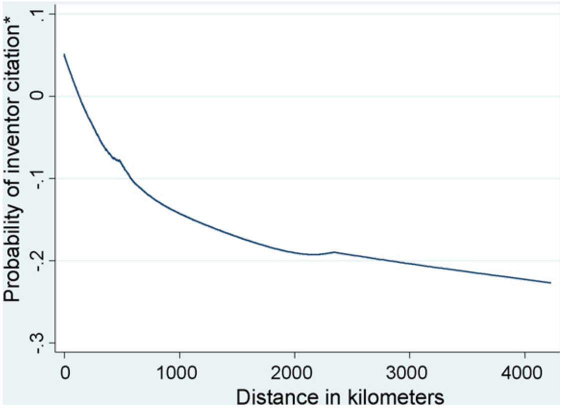

Fig. 5. The relationship between distance and the likelihood of an inventor citation, within-Europe sample.

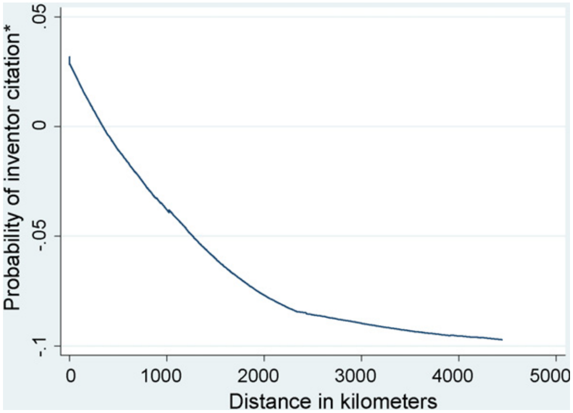

Fig. 6. The relationship between distance and the likelihood of an inventor citation within the US sample.

ised into units of 173 km, on the horizontal axis we take a distance variable in units of 1 km.

Fig. 5 confirms that the effect of distance is non-linear. At short distances between the cited and citing patent, the likelihood of a knowledge flow (inventor citation) decreases rapidly with distance, but this effect reduces at larger distances. Beyond 1000 km (which is the distance, say, between the Brussels and Vienna regions), the marginal effect of distance on the likelihood of a knowledge flow is very low. This non-linear effect of distance is consistent with the results in Bottazzi and Peri (2003).

Fig. 6 depicts the results for the within-US sample. For the US space, the maximum distance is somewhat larger (the horizontal scale extends to 5000 instead of the 4000 in Fig. 5, and while this is a real outlier in Europe, it is not in the US). Despite this, the curve for the US is also clearly non-linear, although the rate of decline is more modest than in Europe. This finding is in line with the evidence reported in Alcacer and Gittelman (2006), who found also that the localisation effect is stronger over shorter distances.

spline function for Distance (km), and kernel regression instead of locally weighted regression in the above procedure. The results of these methods were generally in the same direction as the results documented here.

7. Conclusions

The European patent database allows identification of whether citations are added by the applicant/inventor (inventor citations) or the patent examiner for all of its patents. Moreover, since the legal requirements for citing prior art differ between the European and US systems, we expect the indication of knowledge spillovers based on the patent citations in our database to be different from the USPTO citation studies that dominate the literature. On the basis of the EPO database, we asked about what factors have an influence on the probability that a potential knowledge flow or spillover actually takes place. We found evidence that geographical distance has a negative impact. Time since the date of the cited patent has a small positive impact. Cognitive distance between knowledge receiver and knowledge sender also has a negative impact, except for the part of our sample that refers to the US space. This latter result is because of the effect of US-affiliates of EU-firms, and we suggest it is related to asset-seeking behaviour by European firms. We also found that examiner citations more often involve citations that may compromise novelty, which shows that inventors may have a tendency to ignore citations that may endanger their patent claims.

These results can be interpreted as supporting the hypothesis that inventor citations and examiner citations (in the EPO system) are different, i.e., we find that examiner citations generally do not track inventor citations. Contrary to some of the evidence based on USPTO patent citations, in Alcacer and Gittelman (2006), we find that inventor citations are more localised than examiner citations. This hypothesis has generally been tested using patent citation data, which must be considered a very noisy indicator of knowledge flows. Whereas Alcacer and Gittelman (2006) issue a clear warning against extrapolating these 'noisy' results on geography in a wide variety of spatial contexts, our results based on European patents clearly support the importance of distance in the broadest way possible.

But our results also indicate that geographic distance is not the only variable that impacts on spillovers: cognitive distance and time are also important. These factors have been under-explored in the econometric literature on knowledge flows, and our results suggest that more applied and theoretical work on these factors could be very useful.

In summary, our results clearly indicate that European patent and US patent citations are quite different. These differences stem from the institutional differences between the two systems, discussed in Section 2, and particularly the less stringent requirements in Europe on describing the state of the art in the patent application, which may lead European inventors to cite only those patents that they actually know about ex ante. But it may also be the case that these differences are due to different behaviour and incentives on behalf of the patent examiners in the two systems. While additional research on the background to the different citation practices between the EPO and the USPTO would be useful, our results clearly suggest that it is inventor citations rather than the total set of citations, that should be taken as indicators of knowledge flows.

Acknowledgements

The authors would like to thank three anonymous referees, Gustavo Crespi, Aldo Geuna, Colin Webb and Joe Hilbe for helpful comments and Grid Thoma and Salvatore Torrisi for letting us use their dataset. Paola Criscuolo acknowledges financial support from the Engineering and Physical Sciences Research Council's IMRC at Imperial College London. The usual disclaimer applies.

References

Akers, N., 1999. The European Patent System: an introduction for patent searchers. World Patent Information 21, 135–163.
Akers, N., 2000. The referencing of prior art documents in European patents and applications. World Patent Information 22, 309–315.
Alcacer, J., Gittleman, M., 2004. How do I know what you know? The role of inventors and examiners in the generation of patent citations, Boston.
Alcacer, J., Gittelman, M., 2006. Patent citations as a measure of knowledge flows: the influence of examiner citations. Review of Economics and Statistics 88, 774–779.
Allison, J.R., Lemley, M.A., Moore, K.A., Trunkey, R.D., 2004. Valuable Patents. Georgetown Law Journal 92, 435–480.
Bottazzi, L., Peri, G., 2003. Innovation and spillovers in regions: evidence from European patent data. European Economic Review 47, 687–710.
Breschi, S., Lissoni, F., 2001. Knowledge spillovers and local innovation systems: a critical survey. Industrial and Corporate Change 10, 975–1005.
Breschi, S., Lissoni, F., 2004. Knowledge Networks from Patent Data: Methodological Issues and Research Targets. CESPRI Working papers, Milan.
Collins, P., Wyatt, S., 1988. Citations in patents to the basic research literature. Research Policy 17, 65–74.
Criscuolo, P., Narula, R., Verspagen, B., 2005. The role of home and host country innovation systems in R&D internationalisation: a patent citation analysis. Economic of Innovation and New Technologies 14, 417–433.
Duguet, E., MacGarvie, M., 2005. How well do patent citations measure flows of technology? Evidence from French innovation surveys. Economics of Innovation and New Technologies 14, 375–394.
Griliches, Z., 1992. The search for R&D spillovers. Scandinavian Journal of Economics 94, S29–S47.
Hausman, J.A., Newey, W.K., 1995. Nonparametric estimation of exact consumers surplus and deadweight loss. Econometrica 63, 1445–1476.
Hedge, D., Sampat, B.N., 2005. Examiner citations, applicant citations and the private value of patents. Mimeo Columbia University.
Jaffe, A.B., Trajtenberg, M., 1996. Flows of knowledge from universities and federal labs: modelling the flow of patent citations over time and across institutional and geographical boundaries.
Jaffe, A.B., Trajtenberg, M., 1999. International knowledge flows: evidence from patent citations. Economics of Innovation and New Technologies 8, 105–136.
Jaffe, A., Trajtenberg, M., 2002. Patents Citations and Innovations: A Window on the Knowledge Economy. MIT Press, Cambridge, MA.
Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. The Quarterly Journal of Economics 108, 577–598.
Jaffe, A.B., Fogarty, M.S., Banks, B.A., 1998. Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation. Journal of Industrial Economics 46, 183–205.
Jaffe, A., Trajtenberg, M., Fogarty, M., 2000. The meaning of patent citations: report on the NBER/Case-Western Reserve survey of patentees. American Economic Review 90, 215–218.
Kuemmerle, W., 1996. Home Base and Foreign Direct Investment in R&D. Harvard Business School.
Maurseth, P., Verspagen, B., 2002. Knowledge spillovers in Europe. A patent citations analysis. Scandinavian Journal of Economics 104, 531–545.
Meyer, M., 2000. What is special about patent citations? Differences between scientific and patent citations. Scientometrics 49, 93–123.
Michel, J., Bettels, B., 2001. Patent citation analysis. A closer look at the basic input data from patent search reports. Scientometrics 51, 795–816.
Narin, F., Rosen, M., Olivastro, D., 1988. Patent citation analysis: new validation studies and linkages statistics. In: van Raan, A.F.J., Nederhof, A.J., Moed, H.F. (Eds.), Science Indicators: Their Use in Science Policy and Their Role in Science Studies. DSWO Press, The Netherlands.
OST, 2002. Science et Technologie: Indicateurs Economica. OST, Paris.
Sampat, B.N., 2004, Examining patent examination: an analysis of examiner and applicant generated prior art', Paper presented at NBER conference. Boston.
Sampat, B.N., 2005. Determinants of patent quality: an empirical analysis. Mimeo Columbia University.
Schmoch, U., 1993. Tracing the knowledge transfer from science to technology as reflected in patent indicators. Scientometrics 26, 193–211.
Thoma, G., Torrisi, S., 2007. Creating Powerful Indicators for Innovation Studies with Approximate Matching Algorithms. A Test Based on PATSTAT and Amadeus databases. Bocconi University, Milan.
Thompson, P., 2006. Patent citations and the geography of knowledge spillovers: evidence from inventor – and examiner – added citations. Review of Economics and Statistics 88, 383–388.
Verspagen, B., Schoenmakers, W., 2004. The spatial dimension of patenting by multinational firms in Europe. Journal of Economic Geography 4 (1), 23–42.
Webb, C., Dernis, H., Harhoff, D., Hoisl, K., 2004. A first set of epo patent database building blocks for analysing European and international patent citations.