Firm strategic behavior and the measurement of knowledge flows with patent citations

Published:
DOI: http://doi.org/10.1002/smj.3016

Corsino, M., Mariani, M., & Torrisi, S. (2019). Firm strategic behavior and the measurement of knowledge flows with patent citations. Strategic Management Journal, 40(7), 1040-1069.

https://doi.org/10.1002/smj.3016

Firm strategic behavior and the measurement of knowledge flows with patent citations

Marco Corsino, **Myriam Mariani**, **Salvatore Torrisi**

***Research Summary*:** This research addresses firms' use of external knowledge sources to develop patented inven-tions and explores the validity of patent citations as an indicator of interfirm knowledge flows. By comparing pat-ent citations with primary data reported by the inventors, we uncover systematic measurement errors in patent cita-tions and show that they depend on the firms' patent strate-gies (e.g., to reduce the risk of imitation or litigation), the source of knowledge employed (e.g., competitors, users), the technology of the underlying invention, and the insti-tutional characteristics of the patent system. Our findings about the role of these factors in external knowledge sourcing and citing propensity highlight the importance of firms' strategic behavior and offer novel insights for the use of patent citations as an indicator of knowledge flows. ***Managerial Summary*:** Firms' open innovation strategies rely on the sourcing of knowledge from other organiza-tions. Tracing these knowledge flows is difficult, such that the empirical research on this matter typically uses cita-tions that patents make to prior art in order to track them. However, patent citations might be added also for reasons other than the actual transfer of knowledge. We use pri-mary information from a large survey of inventors to assess the accuracy of patent citations to measure knowl-edge flows, and we find evidence of measurement errors that depend on the applicants' patent strategies, the type of knowledge sources used, the filing jurisdiction, and the technology of the underlying invention. We offer insights to evaluate the settings in which patent citations are a reli-able measure of knowledge flows.

KEYWORDS

external search strategy, knowledge flows, measurement error, patent citations, patent strategy

1 INTRODUCTION

Firms' openness to external interactions and the ability to select, integrate, and profit from external knowledge sources are key features to exploit innovation opportunities and achieve market success (Chesbrough, 2003). As such, the idea that firms can benefit from knowledge flows from external parties has permeated the innovation and strategy literature, in general, and the studies on firms' open innovation models, in particular (Chatterji & Fabrizio, 2014; Laursen & Salter, 2006; Leiponen & Helfat, 2010).

Building on Jaffe, Trajtenberg, and Henderson (1993), the most widely used indicator of knowledge flows at a fine-grained level of detail relies on patent citations (e.g., Agarwal, Ganco, & Ziedonis, 2009; Alcácer & Gittelman, 2006; Thompson & Fox-Kean, 2005). However, not only do we "know surprisingly little about how well citations measure knowledge flows" (Roach & Cohen, 2013, p. 504), but extant research also supports the view that they are a noisy indicator of knowledge flows and suffer from measurement errors (Duguet & MacGarvie, 2005; Jaffe, Trajtenberg, & Fogarty, 2000; Roach & Cohen, 2013). Analyzing the sources and magnitude of such errors has gained renewed interest for research, in particular because patent citation patterns are affected by the increasing adoption of patent-based strategies in several industries (Sampat, 2010), such as those aimed at reducing the risk of patent litigation and invalidation (Cotropia, Lemley, & Sampat, 2013; Lampe, 2012; Steensma, Chari, & Heidl, 2015).

Our research addresses the issue of the validity of patent citations as a measure of interfirm knowledge flows, and the role played by firms' strategic behavior in this relationship. We employ data on 12,619 research projects that led to patented inventions in business organizations and compare direct survey information about knowledge sourcing that occurred during the inventive process with prior art references listed in the resulting patents. We focus on knowledge flows between business organizations, which account for most patented inventions and represent the most important external sources of knowledge in several industries (Arora, Cohen, & Walsh, 2016; Cohen, Goto, Nagata, Nelson, & Walsh, 2002). Our data allow us to investigate the type of knowledge sources used during the inventive activity (e.g., customers, suppliers, and competitors), the associated firms' patent strategy (e.g., commercial exploitation, preempting competing inventions, avoiding litigation), and the decision to disclose prior art.

Our findings reveal that patent citations reflect several dimensions of the actual transfer of knowledge. Nevertheless, they also show both errors of omission and errors of commission, and the magnitude of these errors depends on the institutional setting in which the patent is filed, the technology characteristics of the invention, the applicant's strategic intent, and the heterogeneity in the use of different types of external sources of knowledge.

With this study, we contribute to three streams of research. First, we extend the literature on the validation of patent citations as measures of knowledge flows, with a focus on knowledge interactions between firms. Despite their importance, earlier studies either focus on knowledge flows from public research (e.g., Roach & Cohen, 2013) or do not distinguish between public and private sources of knowledge (e.g., Duguet & MacGarvie, 2005), thus neglecting the potentially different patterns and (strategic) reasons driving knowledge transfer in the private sector compared with the "open science" culture of the public sector.

Second, we contribute to a related body of research that investigates the reasons for the existence of systematic measurement errors in patent citations, such as the characteristics of the patent system and the applicant patenting strategies. Legal requirements to delimit the scope of a patent's claims create an incentive for the applicant to include references that highlight the novelty of the claims and to exclude those that may undermine the granting process (Hegde & Sampat, 2009). “Legal” or “strategic” citations are included to decrease the likelihood of postgrant litigation and invalidation (Harhoff & Reitzig, 2004; Lampe, 2012). Institutional differences between patent systems also create different incentives to disclose prior art, such that the U.S. inequitable conduct doctrine, for example, may make the USPTO (United States Patent and Trademark Office) patent citations a noisier indicator of knowledge flows than EPO (European Patent Office) citations (Criscuolo & Verspagen, 2008). Finally, a large share of citations are added by the patent examiners in both the USPTO and the EPO, making it difficult to establish whether inventors are aware of the cited patents at the time of invention (e.g., Alcácer & Gittelman, 2006; Jaffe et al., 2000; Steensma et al., 2015).

Third, we bridge the stream of research on patent strategies (e.g., Cohen, Goto, et al., 2002; Somaya, 2012) with the open innovation literature and complement the latter by showing that the use of specific knowledge sources covaries with the firm's patent strategies. This association is reflected in the decision to disclose prior art, and therefore in the reliability of patent citations as indicators of knowledge flows. Thus, for example, patents filed to preempt other patents typically rely on competitors as a source of knowledge. In turn, knowledge sourced from competitors leaves a trail in patent citations. Differently, knowledge sourced from users is not reflected in backward citations.

2 BACKGROUND LITERATURE

Our work draws on three streams of the literature. The first one comprises studies on the validity of patent citations to measure knowledge flows. Contributions in the second body of research uncover confounding factors that dampen the suitability of patent citations as proxies for knowledge flows and assess the role of the applicant's strategic considerations in adding or withholding patent citations. The third set of studies in the open innovation literature concerns the use that firms make of, and the benefits that they draw from, external knowledge sources in the innovation process.

2.1 Validation studies

Extant contributions that assess the validity of patent citations as a measure of knowledge flows rely on the key informants about the use of external knowledge during the inventive process, that is, inventors and R&D managers (Jaffe et al., 2000; Mattes, Stacey, & Marinova, 2006; Tijssen, 2002). Jaffe et al. (2000) surveyed a sample of U.S. inventors and find that only one-quarter of the respondents recognize the correspondence between patent citations and knowledge spillovers. Half the inventors report a low degree of familiarity with the inventions cited in their patents, and about one-third indicate that they did not know about the cited invention before the interview. Tijssen (2002) finds similar results for citations to nonpatent literature.

Other studies use the firm or the R&D laboratory as the unit of analysis. With data from the French Community Innovation Survey, Duguet and MacGarvie (2005) find a positive correlation between patent citations and firms' acquisition of new technologies, whereas they find no correlation with the channels of knowledge flows from open science. Nelson (2009) shows that over 82% of the organizations that developed and licensed new products based on Stanford University's patents on recombinant DNA technology never cite them in their inventions. More recently, for patents assigned to public research organizations (PROs), Roach and Cohen (2013) provide evidence of a shared variation between survey-based measures of knowledge flows and backward citations. These citations, however, correlate with firms' appropriability and citing strategies rather than capturing open science and contract-based knowledge exchanges.

A feature common to all these studies is that they do not explicitly address the validity of patent citations as a measure of knowledge flows between firms. Firms are instead the source of external knowledge most relevant to other firms' R&D projects (Arora et al., 2016; Cohen, Goto, et al., 2002) and those for which the applicants' strategic use of patents as competitive weapons (Somaya, 2012) may play a key role in affecting the citing behavior. These considerations lead to our first research question: *How accurate are citations as a measure of knowledge flows between firms?*

2.2 Firms' strategic behavior and the measurement errors in patent citations

Two factors can produce errors in patent citations as measures of knowledge flows. The first one is that most citations are added by the patent examiners rather than by the applicants or the inventors. Alcácer, Gittelman, and Sampat (2009) and Sampat (2010) for USPTO patents and Criscuolo and Verspagen (2008) for EPO patents find that examiners add the majority of citations, and therefore the inventors may not be aware of them at the time of the invention. Besides, examiner-added citations may depend on examiners' cognitive biases (Alcácer & Gittelman, 2006; Cotropia et al., 2013) and therefore could not be randomly distributed (Alcácer et al., 2009).

Second, the applicants' decision to search for and disclose prior art is affected by their patent strategies (Atal & Bar, 2010; Langinier & Marcoul, 2016; Sampat, 2010). On one hand, applicants may omit prior art in order to obtain a broader patent scope or to reduce the likelihood that patent applications are rejected (Cotropia et al., 2013; Lampe, 2012). These omissions are more likely for applicants with large patent portfolios because the latter reduce the probability of lawsuit attacks (Lampe, 2012). They also depend on the accumulated experience with the patenting process that nurtures the inventors' ability to estimate the risk of invalidation (Alcácer et al., 2009; Steensma et al., 2015). The fact that examiners suffer from resource constraints and must comply with productivity targets (Jaffe & Lerner, 2004) may further reduce the inclusion of references. On the other hand, applicants have an incentive to disclose more prior art in the case of high-expected-value patents, for which the cost of invalidation would be large relative to the loss of breadth of the property right (Lampe, 2012; Sampat, 2010), or if they want to reduce the probability of invalidation in the case of postgrant litigation (Allison, Lemley, Moore, & Trunkey, 2004), or, still, when they file patent applications to preempt the granting of competing patents (Guellec, Martinez, & Zuniga, 2012). Patent attorneys may also contribute to inflating the count of citations to early patents to establish the patentability of an invention (Moser, Ohmstedt, & Rhode, 2018).

The unfolding of these contrasting forces and their influence on the strategic disclosure of prior art varies across technological fields (Jaffe & de Rassenfosse, 2017). Applicants in discrete technologies (e.g., chemicals and drugs), in which a single patent can define and protect an invention, are more active in searching for prior art than applicants in complex-product technologies (e.g., computers, communications, and electronics), in which the boundaries between patents are blurred and patents are more important as bargaining chips in cross-licensing deals and infringement suits than for appropriating the returns from R&D (Cohen, Goto, et al., 2002). In these technologies the possibility that a patent does not survive a validity challenge is less critical and, hence, withholding citations is more frequent (Lampe, 2012; Sampat, 2010).

Finally, institutional differences between patent systems and regulations also create contrasting applicants' incentives to disclose prior art. In the U.S. patent system, the duty of candor and the legal doctrine of inequitable conduct legally bind applicants to disclose known prior art and prompt them to overcomply with the submission of references (Cotropia, 2009) because the costs of undercompliance (i.e., the rejection of a patent application and the unenforceability of a granted patent) are disproportionally larger than the costs of overcompliance (i.e., a possible reduction of the scope of protection). However, the rule that applicants must disclose prior art that they already know may weaken the incentive to search for new references (Atal & Bar, 2010; Langinier & Marcoul, 2016). Finally, postgrant opposition is cheaper at the EPO because of the limited attractiveness of the reexamination procedure in the USPTO as an early-stage litigation mechanism (Harhoff & Reitzig, 2004) and the U.S. courts' reluctance to overturn a decision made by the patent office after a patent is granted (Allison & Lemley, 1998; Sampat, 2010). This may increase the EPO applicant's incentive to search for and disclose prior art to avoid invalidation. These considerations lead to our second research question: *How much does a patent applicant's strategic behavior account for the mismatch between knowledge flows and citations flows?*

2.3 Knowledge flows and patent strategies: A window on open innovation models

The idea that external knowledge sourcing yields productivity benefits has gained popularity in the innovation and strategy research (e.g., Baldwin & von Hippel, 2011; Chesbrough, 2003). The open innovation literature, in particular, offers contributions that describe the extent and implications of firms' use of specific sources of knowledge, or the different degrees of breadth and depth of external search strategies (e.g., Chatterji & Fabrizio, 2014; Laursen & Salter, 2006; Leiponen & Helfat, 2010), or, still, the trade-off between the openness to external knowledge sourcing and the ability to protect a firm's own knowledge (Laursen & Salter, 2014), known as the 'paradox of openness.' In this latter context, intellectual property rights play a twofold role, as they may foster open innovation postures (e.g., Chesbrough, 2003; Zobel, Balsmeier, & Chesbrough, 2016) or they can instead hamper external knowledge sourcing in open, free innovation settings (Baldwin & von Hippel, 2011; Wadhwa, Bodas Freitas, & Sarkar, 2017).

Appropriability concerns, in turn, depend on the type of knowledge source used. They are particularly severe when firms use knowledge drawn from competitors (Laursen & Salter, 2014). In these cases patents can serve offensive purposes, such as that of blocking other innovations, and, at the same time, they may also cite more prior art to reduce the risk of infringement and litigation. Knowledge sourcing from customers and users, on the other hand, involves a lower risk of imitation and patent litigation, which likely materializes in a lower likelihood of citing prior art for either offensive or defensive purposes. Our study will investigate this aspect of the "paradox of openness" by digging into the heterogeneity of patent strategies and the citation behavior associated with the use of different knowledge sources. This is the aim of our third research question: *How do patent strategies, patent citations, and the use of different external sources of knowledge relate to each other?*

3 METHOD, DATA, AND VARIABLES

3.1 Method

We apply Roach and Cohen's (2013) methodology to investigate the reliability of patent citations as a measure of knowledge flows between firms. The estimation approach distinguishes between two sources of measurement errors: errors of omission, or the failure of citations to capture important dimensions of knowledge flows; and errors of commission, or the possibility that citations correlate with factors other than knowledge flows. To uncover the two sources of errors, we estimate the following reduced-form regression models:

$$k_c = \alpha_1 X_1 + \alpha_2 X_2 + \gamma_c P + \varepsilon_c \tag{1}$$

$$k_s = \theta_1 X_1 + \theta_2 X_2 + \gamma_s P + \varepsilon_s \tag{2}$$

The dependent variable in model (1), k_c, is a measure of knowledge flows based on references to patented prior art (or backward citations). The dependent variable in model (2), k_s, is an indicator of the "true" knowledge flows that exploits a survey-based measure of the importance of knowledge sources during the invention process. The set of covariates in both equations are X_1, a vector of correlates of knowledge flows reflected by patent citations; X_2, a vector of correlates of knowledge flows that are not reflected by patent citations; and P, a vector of factors correlated with patent citations that do not reflect knowledge flows. X_2 and P are crucial in identifying the sources of errors of omission and commission, respectively.

3.2 Data

This research employs data from a comprehensive data set that combines survey and archival data at the level of the patent application. Primary data about the knowledge sources used during the inventive projects come from the InnoS&T survey that collects information on 23,044 randomly selected patent applications filed at the EPO, with priority dates between 2003 and 2005 (see Torrisi et al., 2016, for details on the sampling and data collection procedures).[1] For the purpose of this study, we focus on 20,825 patent applications filed by firms.

We matched 20,499 surveyed patents with secondary data from the CRIOS-PatStat database (Coffano & Tarasconi, 2014) that contains standardized information on patents, applicants, and the Documentation Database (DOCDB) patent families of EPO patent applications, that is, the collection of patent documents with the same priority date (Martínez, 2011). We also retrieved citation data for the 20,499 patents at the DOCDB patent–family level. After eliminating duplicates in citing–cited pairs within the same DOCDB family, cited patents without EPO equivalents, self-citations, and observations with missing values in key covariates, we ended up with a working sample of 12,619 patents that comprises 4,356 firms (the applicants), with an average number of patents per firm equal to 2.9 (the median is 1). Section 2 of the Appendix S1, Supporting Information shows that the working sample of 12,619 observations and the sample of 7,880 observations not used in this study do not differ systematically in some key patent-related characteristics.

3.3 Variables

3.3.1 Dependent variables

For each surveyed patent, we build two types of dependent variables: citations- and survey-based indicators of knowledge flows from other firms. The variable Back cites counts the total number of backward citations to patents whose applicant is a business organization. To account for the backward citations added by the examiners, App back cites counts only applicant-added citations, which, in our sample, represent 39.34% of all citations. Section 3 of Appendix S1 shows the distribution of backward citations along several dimensions.

The InnoS&T survey asked inventors to rate on a 6-point Likert scale (0 = not used; 5 = very important) the importance of five types of external sources of knowledge during the inventive process, that is, customers, users, suppliers, competitors, and consulting or contract R&D. The variable KFs survey is constructed by summing the scores assigned to each of these sources; therefore, it ranges between 0 (none of the source was used, 12.98% of the cases) and 25 (all sources were very important, 0.38% of the cases).

[1]Section 1 of Appendix S1 reports the wording of the questions in the InnoS&T survey.

3.3.2 Searching for errors of omission: Correlates of knowledge flows

We consider two sets of factors as potential sources of errors of omission: (a) the channels through which knowledge flows to the recipient inventor, that is, open innovation, formal collaborations, and employee mobility and (b) the type of R&D activity leading to the focal patent.

Open innovation can materialize through the consultation of patent documents (PATENTS), the participation in technical conferences and workshops (TECHNICAL CONFERENCES), and the exploitation of informal interactions in the form of discussions, meetings, and exchange of ideas with people in other organizations (INFORMAL INTERACTIONS). The surveyed inventors rate the importance of these three sources of information on a 5-point Likert scale (1 = not important/not used; 5 = very important). We expect all three sources of information to be positively associated with the survey-based measure of knowledge flows. Citation-based metrics, especially examiner-added citations, may fail to reflect the transfer of knowledge through the participation in technical conferences and informal interactions that involve the use of "invisible colleges" (Cotropia, 2009, p. 754; Kesan, 2002). In contrast, because examiners can easily retrieve patented prior art, knowledge embodied in patents is likely reflected in the citation-based indicator, especially in legal systems that compel applicants to disclose known prior art.

We measure the transfer of knowledge through formal collaborations (Cohen, Nelson, & Walsh, 2002) with a dichotomous variable (FORMAL COLLABORATIONS) that equals 1 if the organization engaged in collaborations involving written contracts with other firms during the inventive process, and 0 otherwise. Finally, knowledge inflows may occur through the hiring of employees (Hoisl, 2007; Marx, Strumsky, & Fleming, 2009), which we measure with the variable INVENTOR MOBILITY that equals 1 if the inventor was formerly employed by a different business organization and moved to the applicant in the 5 years before the invention, and 0 otherwise.

As far as the type of R&D is concerned, compared with applied research, science-based research is more likely to draw on external knowledge from specialized firms or public research. We measure the scientific orientation of the research project by employing information on whether the results of the project are published in scientific journals. The variable PUBLISHED OUTPUT equals 1 if the results related to the focal invention have been published in scientific journals, and 0 otherwise.

3.3.3 Searching for errors of commission: Correlates of patenting and citing behavior

Firms' appropriability and citing strategies can affect the number of patent citations for reasons unrelated to knowledge flows (Roach & Cohen, 2013). For example, the effectiveness of patents as protection mechanisms of the inventions increases the firms' propensity to patent and, as a result, to cite prior art. The variable PATENT EFFECTIVENESS therefore measures on a 5-point Likert scale the importance of obtaining patent rights to exploit the invention economically. We also account for the firm's citing and patenting behavior with the following variables: CITING PROPENSITY, computed as the ratio between the applicant's stock of backward citations (excluding self-citations) and the stock of patents before the priority year of the focal patent; three indicators of the patent value, that is, FAMILY SIZE, or the number of equivalents in the DOCDB family, FORWARD CITATIONS, or the number of citations received by the patent in the 5 years after the EPO application (Hall, Jaffe, & Trajtenberg, 2005; Sampat, 2010; Steensma et al., 2015), and CLAIMS, or the number of claims in the focal patent that defines the scope of the patent protection (Allison et al., 2004).

3.3.4 Exploring the effect of patent strategies

Extant research on strategic citations (e.g., Akers, 2000) suggests the existence of a link between patent citations and applicants' expectations about the use of the patents at the time of application, which is the moment when they add the citations (Cotropia et al., 2013) and also the moment that the survey refers to about the motives for patenting.[2] We draw from the literature on patent strategies (e.g., Allison et al., 2004; Cohen, Goto, et al., 2002; Lampe, 2012; Sampat, 2010; Somaya, 2012) and define three patent uses that firms can pursue and that can influence citing behavior. The first use is to protect from imitation the inventions that a firm uses in the final market; the second is to preempt rivals' inventions; and the third is to protect the applicant from the risk of litigation (Torrisi et al., 2016).

Three dichotomous variables reflect these patent strategies. The variable USED PATENT equals 1 if the inventor claims that the patent was used commercially in a product, a service, a process, or if it was used to found a new firm, or was sold or licensed to third parties; it takes the value 0 if the patent was not used for any of these purposes. To identify patents used to preempt rivals' inventions we concentrate on the group of patents that were not used for commercial reason (i.e., conditional on USED PATENT = 0), and we distinguish between strategically unused patents and sleeping patents. To this end we employ the survey information on the importance of "preventing other firms from patenting similar inventions as a reason for filing the patent application" (BLOCKING, 5-point Likert scale). The variable STRATEGIC NON-USE$_{\text{BLOCKING}}$ takes the value 1 if the score reported for BLOCKING is equal or greater than 4; the variable SLEEPING PATENT$_{\text{BLOCKING}}$, which is our baseline category, reflects patents unused for non-strategic reasons and takes the value 1 if the score for BLOCKING is less than 4. Typically, the disclosure of prior art strengthens the patentee's bargaining power with respect to postgrant oppositions filed by third parties or to validity challenges in national courts, and to licensees' questioning of the validity of the patent (Akers, 2000). Thus, an applicant who expects to use a patent commercially is likely to be particularly concerned about the risk of future validity challenges compared with the case of an unused (sleeping) patent, because the damage s(he) would suffer in the event of invalidation would be much larger. This concern increases the likelihood of adding backward citations in commercially used patents compared with unused, sleeping patents. Applicants of patents filed to preempt other inventions have contrasting incentives to disclose prior art. On the one hand, patent invalidation due to prior art withholding does not directly harm the firm's market position. In addition, when patenting to preempt competing inventions, applicants aim to establish an exclusive patent leadership in a technology, and therefore try to enforce their patents aggressively against possible infringement (Somaya, 2012). These factors reduce the incentive to cite prior art. On the other hand, the incentives to cite prior art by preempting patents could increase when they draw on knowledge from competitors. In this case, the risk of patent invalidation is higher than when drawing on users' or customers' knowledge. We estimate the net effect of these contrasting forces on the incentive to cite prior art.

We also consider a second strategic reason for patenting, that is, the "intention to build a credible threat to countersue firms" (PREVENT SUITS, 5-point Likert scale). Conditional on the patent not being used for commercial reasons (USED PATENT = 0), STRATEGIC NON-USE$_{\text{PREVENT SUITS}}$ takes the value 1 if the score reported for PREVENT SUITS is equal or greater than 4; SLEEPING PATENT$_{\text{PREVENT SUITS}}$ takes the value 1 if the score reported for PREVENT SUITS is less than 4. Similarly to patents used for commercial reasons, the applicant is likely to disclose more prior art in patents filed to prevent litigation compared with sleeping patents, to reduce the risk of third parties' questioning of the validity of the patent, although this could reduce the patent scope (Allison et al., 2004).

3.3.5 Control variables

Firms may accumulate patents as a defensive weapon (Hall & Ziedonis, 2001), and because a large patent portfolio deters third parties from challenging the validity of a firm's patent, defensive citations

[2]Our interview with a patent attorney confirms the existence of this link.

become less likely (Lampe, 2012). We therefore take into account the applicant's patent stock before the priority date of the focal patent (PATENT STOCK). We also control for the presence of a U.S. equivalent in the patent family that may increase the citation propensity because of the duty of candor. The variable US EQUIVALENT is equal to 1 if there is a U.S. equivalent in the DOCDB family of the patent, and 0 otherwise. To further control for the different patent-related regulations, we include eight dummy variables identifying the patent office of the initial filing (PRIORITY OFFICE), and four additional dummy variables accounting for the patent's legal status (LEGAL STATUS).

Three dummy variables account for size differences across applicant organizations, and 30 dummy variables control for the technological class of the focal patent as defined in the Observatoire des Sciences et des Techniques (OST) classification (Schmoch, 2008). Table 1 describes the variables employed in the empirical study and Table 2 shows their descriptive statistics and the pairwise correlation matrix.

4 ECONOMETRIC ANALYSIS

Because of the non-negative, integer values of the dependent variables, we estimate Equations (1) and (2) with a negative binomial regression model and report the average marginal effects for the covariates. However, the different natures and scales of the citation- and survey-based dependent variables do not allow for a straight comparison of the estimated results from the two equations. To this end, we compute the relative magnitudes of the estimated effects with respect to the average value of the dependent variables and present them in the discussion of the results. Only some model specifications use an ordered logit regression model because of the ordinal, categorical nature of the dependent variable.

This section reports four sets of results. The first set shows the existence of measurement errors in the citation-based indicator of knowledge flow. The second set explores the role of patent strategies on the applicant's citing behavior. The third set focuses on the shared variation between the survey- and citation-based indicators, and it documents heterogeneous results according to the type of knowledge source employed. The fourth set reports the robustness checks of our results, including a replication of the Roach and Cohen (2013) study for knowledge flows from PROs.

4.1 Baseline model

Results in columns 1 through 3 of Table 3 address our first research question and point to the existence of errors of omission and errors of commission in citation-based indicators. All channels of knowledge flows are positively correlated with the survey-based measure (Column 1). Thus, for example, a 1-*SD* rise in PATENTS implies a 1.35-point increase in KFs SURVEY, which represents a 14.06% shift in the average value of this variable. Inventions that rely on formal collaborations entail an increase of 2.20 points in the value of KFs SURVEY, and inventors' mobility is associated with a 0.40-point increase in the same variable.

The citation-based measure, instead, reflects only knowledge flows occurring through PATENTS, INFORMAL INTERACTIONS, and INVENTOR MOBILITY. It does not capture knowledge acquired through technical conferences and formal collaborations, pointing to the existence of errors of omission. A 1-*SD* increase in the importance of patents, informal interactions, and inventors' mobility is associated with a rise of 0.14, 0.11, and 0.20 of APP BACK CITES, respectively (Column 3). The largest effect is that of INVENTOR MOBILITY, which represents a 10.25% shift in the average number of APP BACK CITES. Note that the size of the correlations decreases after the inclusion of the examiner-added citations in the

TABLE 1 Description of variables

Name	Source	Measure
Measures of knowledge flows		
KFs survey	InnoS&T survey	Variable measuring how important other firms (customers, users; suppliers, competitors, consulting) were as sources of knowledge during the invention process. The variable reports the summation of the scores assigned to each source and ranges from 0 (no source was used) to 25 (all sources were used and were important)
KFs customers	InnoS&T survey	Importance of customers as knowledge source, 6-point Likert scale (0 = not used; 5 = very important)
KFs users	InnoS&T survey	Importance of users as knowledge source, 6-point Likert scale (0 = not used; 5 = very important)
KFs suppliers	InnoS&T survey	Importance of suppliers as knowledge source, 6-point Likert scale (0 = not used; 5 = very important)
KFs competitors	InnoS&T survey	Importance of competitors as knowledge source, 6-point Likert scale (0 = not used; 5 = very important)
KFs consulting	InnoS&T survey	Importance of consulting or contract R&D firms as knowledge source, 6-point Likert scale (0 = not used; 5 = very important)
Back cites	CRIOS-PatStat	Number of backward citations to other firms' patents (no self-citations)
App back cites	CRIOS-PatStat	Number of applicant's backward citations to other firms' patents (no self-citations)
Channels of knowledge flows		
Patents	InnoS&T survey	Importance of patent documents as sources of information during the invention process, 5-point Likert scale (1 = not important/not used, 5 = very important)
Technical conferences	InnoS&T survey	Importance of technical conferences/workshops as sources of information during the invention process, 5-point Likert scale (1 = not important/not used, 5 = very important)
Informal interactions	InnoS&T survey	Importance of informal interactions with people in other organizations during the invention process, 5-point Likert scale (1 = not important/not used, 5 = very important)
Formal collaborations	InnoS&T survey	Dummy that equals 1 if there were collaborations involving written contracts with other firms during the invention process, 0 otherwise
Inventor mobility	InnoS&T survey	Dummy that equals 1 if the inventor changed employer in the 5 years before developing the invention, and previous employer was a firm, 0 otherwise
Type of R&D		
Published output	InnoS&T survey	Dummy that equals 1 if results related to the invention have been published in scientific journals, 0 otherwise
Patenting and citing behavior		
Patent effectiveness	InnoS&T survey	Importance of commercial exploitation (i.e., obtain exclusive rights to exploit the invention economically) as a reason for patenting the invention, 5-point Likert scale (1 = not important, 5 = very important)
Family size	CRIOS-PatStat	Log of the number of equivalents in the DOCDB family of the focal patent
Claims	CRIOS-PatStat	Log of the number of claims in the focal patent
Forward citations	PATSTAT	Log of the number of forward citations received by the focal patent in the 5 years after the application to the EPO
Citing propensity	CRIOS-PatStat	Log of the of the applicant's average number of backward citations per patent
Patent strategy		
Used patent	InnoS&T survey	Dummy that equals 1 if the patent has been used commercially (i.e., in a product, service, or process), or sold, or licensed, or used to found a new firm, 0 otherwise
Blocking	InnoS&T survey	Importance of blocking (i.e., avoid that others patent similar inventions, complements or substitutes) as a reason for patenting the invention, 5-point Likert scale (1 = not important, 5 = very important)

TABLE 1 (Continued)

Name	Source	Measure
Prevent suits	InnoS&T survey	Importance of preventing infringements suits (build a credible threat such that your organization can sue others if they sue your organization) as a reason for patenting the invention, 5-point Likert scale (1 = not important, 5 = very important)
Strategic Non-use $_{\text{Blocking}}$	InnoS&T survey	Dummy that equals 1 if Used patent==0 & Blocking ≥ 4, 0 otherwise
Sleeping patent $_{\text{Blocking}}$	InnoS&T survey	Dummy that equals 1 if Used patent==0 & Blocking<4, 0 otherwise
Strategic Non-use $_{\text{Prevent suits}}$	InnoS&T survey	Dummy that equals 1 if Used patent==0 & Prevent suits ≥ 4, 0 otherwise
Sleeping patent $_{\text{Prevent suits}}$	InnoS&T survey	Dummy that equals 1 if Used patent==0 & Prevent suits<4, 0 otherwise
Controls		
Patent stock	CRIOS-PatStat	Log of the applicant's patent stock
US equivalent	CRIOS-PatStat	Dummy that equals 1 if the focal patent has a U.S. equivalent, 0 otherwise
Priority office	CRIOS-PatStat	8 dummy variables identifying the patent office of the first filing of the application (Europe, United States, Japan, Germany, France, United Kingdom, Italy, Other national authorities): reference category is United States)
Legal status	CRIOS-PatStat	4 dummy variables identifying the legal status of the patent as of December 2014 (1 = Granted; 2 = Pending; 3 = Refusal or Revocation; 4 = Withdrawn)
Technological class	CRIOS-PatStat	30 OST classification dummy variables: reference category is Telecommunications
Firm size	InnoS&T survey	Three dummy variables measuring the size of the applicant (Firm size$_{\text{SME}}$ = 1 if applicant has fewer than 1,000 employees; Firm size$_{\text{LARGE}}$ = 1 if applicant has 1,000 to 4,999 employees; Firm size$_{\text{VERY LARGE}}$ = 1 if applicant has 5,000 or more employees)

dependent variable (Back cites in Column 2). The variable Published output is not significantly associated with the citation-based measure App back cites. It is negatively correlated with KFs survey and Back cites.

The average marginal effects of the variables Family size, Claims, Forward citations, and Citing propensity indicate that errors of commission plague citations as measures of knowledge flows. While none of them is associated with the survey-based indicator, they all correlate with the citation-based measures: a 1-*SD* increase in Family size, Claim, Forward citations, and Citing propensity leads to an increase in the number of applicant-added citations of 0.53, 0.41, 0.39, and 0.62, respectively. Citing propensity carries the largest effect, with a 31.66% increase in the average value of App back cites. Patent effectiveness, instead, is positively associated with both the survey-based and the citation-based measures of knowledge flows, with a 1-*SD* increase in Patent effectiveness producing a 0.35-point increase in KFs survey and a 0.21 increase in App back cites (0.18 in the case of Back cites). Thus, with the exception of Patent effectiveness, all patenting variables covary with backward citations, but do not covary with KFs survey.

We address the second research question in columns 4 through 6 in Table 3 that add the patent strategy variables Used patent and Strategic non-use$_{\text{Blocking}}$ to the estimated models. Sleeping patent$_{\text{BLOCKING}}$ is the excluded reference strategy. Columns 7 through 9 focus on patent strategies to prevent infringement suits (Strategic non-use$_{\text{Prevent suits}}$), with Sleeping patent$_{\text{PREVENT SUITS}}$ being the excluded baseline strategy. The estimated results show that patents filed for commercial purposes (Used patent) correlate positively with both the survey- and citation-based measures of knowledge flows, and they entail an increase of 1.19 points (Column 4) and 0.21 units (Column 6) respectively, compared with sleeping patents, similar to the estimated effects in columns 7 and 9. Differently, patents filed for strategic reasons rely on knowledge from other firms more than sleeping patents do, as

TABLE 2 Descriptive statistics and correlation matrix

		Mean	***SD***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
1	KFs survey	9.575	6.121																	
2	Back cites	5.011	7.324	−0.01																
3	App back cites	1.971	5.865	−0.02	0.89															
4	KFs customers	2.475	1.966	0.77	0.01	0.00														
5	KFs users	2.343	1.953	0.78	−0.06	−0.06	0.61													
6	KFs suppliers	1.854	1.757	0.67	−0.01	−0.01	0.35	0.37												
7	KFs competitors	1.931	1.796	0.67	0.00	−0.01	0.36	0.38	0.30											
8	KFs consulting	0.972	1.313	0.52	0.05	0.04	0.23	0.23	0.31	0.24										
9	Patents	3.356	1.612	0.24	0.05	0.04	0.14	0.12	0.15	0.29	0.12									
10	Technical conferences	2.239	1.379	0.25	0.03	0.02	0.17	0.14	0.16	0.22	0.20	0.24								
11	Informal interactions	1.700	1.287	0.26	0.01	0.02	0.18	0.17	0.27	0.08	0.24	0.08	0.14							
12	Formal collaborations	0.245	0.430	0.24	0.01	0.01	0.16	0.14	0.28	0.08	0.18	0.08	0.10	0.23						
13	Inventor mobility	0.233	0.423	0.04	0.08	0.07	0.03	0.01	0.04	0.01	0.06	−0.03	−0.02	0.05	0.03					
14	Published output	0.157	0.364	0.02	0.03	0.04	−0.01	−0.01	0.00	0.03	0.09	0.08	0.21	0.10	0.06	−0.03				
15	Patent effectiveness	4.306	1.110	0.09	0.07	0.07	0.10	0.06	0.06	0.06	0.03	0.13	0.04	0.04	0.04	0.02	0.02			
16	Family size	1.855	0.504	−0.03	0.31	0.25	−0.03	−0.06	−0.02	−0.01	0.05	0.08	0.03	0.00	0.03	0.00	0.08	0.09		
17	Claims	2.681	0.589	−0.03	0.22	0.20	−0.02	−0.06	−0.03	−0.05	0.07	0.01	0.04	0.04	0.00	0.08	0.07	0.02	0.14	
18	Forward citations	0.534	0.632	0.01	0.23	0.20	−0.01	−0.03	−0.01	0.05	0.06	0.09	0.07	0.01	0.02	0.01	0.08	0.05	0.19	0.17
19	Citing propensity	1.776	0.591	−0.07	0.26	0.22	−0.03	−0.12	−0.09	0.01	−0.01	0.09	0.06	−0.08	−0.03	−0.04	0.01	−0.01	0.18	0.14
20	Used patent	0.564	0.496	0.11	0.02	0.03	0.12	0.13	0.09	0.00	0.02	−0.05	−0.04	0.05	0.07	0.05	0.04	0.13	0.04	−0.01
21	Strategic non-use$_{\text{Blocking}}$	0.279	0.449	−0.05	0.00	0.00	−0.07	−0.07	−0.05	0.03	−0.02	0.08	0.03	−0.05	−0.05	−0.04	−0.06	−0.02	−0.02	0.00
22	Sleeping patent$_{\text{Blocking}}$	0.157	0.364	−0.08	−0.04	−0.04	−0.08	−0.08	−0.06	−0.04	−0.01	−0.04	0.02	−0.02	−0.04	−0.02	0.01	−0.16	−0.04	0.02
23	Strategic non-use$_{\text{Prevent suits}}$	0.183	0.387	−0.01	0.01	0.00	−0.03	−0.03	−0.02	0.04	0.02	0.06	0.07	−0.03	−0.04	−0.02	−0.04	0.01	−0.02	0.01
24	Sleeping patent$_{\text{Prevent suits}}$	0.253	0.435	−0.12	−0.03	−0.03	−0.11	−0.12	−0.08	−0.04	−0.05	0.00	−0.01	−0.04	−0.05	−0.03	−0.02	−0.15	−0.03	0.01
25	US equivalent	0.711	0.454	−0.05	0.26	0.19	−0.01	−0.08	−0.07	0.00	0.01	0.07	0.06	−0.05	−0.01	−0.01	0.05	0.02	0.41	0.12
26	Granted	0.498	0.500	0.02	0.01	0.01	0.01	0.02	0.03	0.01	−0.02	−0.01	−0.04	−0.01	0.02	−0.04	−0.03	0.01	0.25	−0.09
27	Pending	0.146	0.353	−0.03	0.10	0.08	−0.02	−0.02	−0.03	−0.01	−0.01	−0.01	0.02	−0.01	−0.03	0.00	0.02	0.01	0.00	0.09
28	Refusal	0.028	0.164	0.01	0.02	0.01	0.01	0.01	0.00	−0.01	0.01	−0.02	0.00	0.02	0.00	0.01	−0.01	−0.01	−0.03	0.02
29	Withdrawn	0.328	0.470	0.00	−0.09	−0.07	0.00	−0.01	−0.01	−0.01	0.02	0.03	0.03	0.02	0.00	0.04	0.03	−0.02	−0.26	0.02

TABLE 2 (Continued)

		Mean	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
30	Patent stock	3.984	2.544	−0.08	0.01	0.00	−0.07	−0.10	−0.09	0.03	−0.05	0.02	0.05	−0.12	−0.05	−0.13	−0.05	−0.08	0.00	−0.08
31	Firm size$_{\text{SME}}$	0.278	0.448	0.05	−0.01	0.00	0.06	0.07	0.04	−0.04	0.06	−0.04	−0.05	0.11	0.05	0.15	0.04	0.08	−0.02	0.08
32	Firm size$_{\text{LARGE}}$	0.138	0.345	0.03	−0.01	−0.01	0.03	0.03	0.00	0.03	−0.01	0.04	0.02	−0.02	0.00	−0.02	0.00	0.01	0.01	−0.04
33	Firm size$_{\text{VERY LARGE}}$	0.584	0.493	−0.07	0.02	0.00	−0.07	−0.09	−0.04	0.01	−0.05	0.01	0.03	−0.09	−0.05	−0.13	−0.04	−0.08	0.01	−0.05
		18	19	20	21	22	23	24	25	26	27	28	29	30	31	32				
19	Citing propensity	0.15																		
20	Used patent	−0.03	−0.13																	
21	Strategic non-use$_{\text{BLOCKING}}$	0.03	0.11	−0.71																
22	Sleeping patent$_{\text{BLOCKING}}$	0.00	0.04	−0.49	−0.27															
23	Strategic non-use$_{\text{PREVENT SUITS}}$	0.02	0.08	−0.54	0.58	0.02														
24	Sleeping patent$_{\text{PREVENT SUITS}}$	0.01	0.08	−0.66	0.29	0.54	−0.28													
25	US equivalent	0.19	0.25	−0.01	0.02	−0.01	0.03	−0.01												
26	Granted	0.00	−0.05	0.06	−0.04	−0.03	−0.02	−0.04	0.02											
27	Pending	0.08	0.07	0.03	−0.01	−0.03	−0.01	−0.03	0.11	−0.41										
28	Refusal	0.01	0.02	0.01	−0.01	0.01	0.00	−0.01	−0.01	−0.17	−0.07									
29	Withdrawn	−0.06	−0.01	−0.09	0.06	0.05	0.03	0.07	−0.10	−0.70	−0.29	−0.12								
30	Patent stock	0.06	0.57	−0.20	0.15	0.08	0.14	0.11	0.14	−0.01	0.02	0.04	−0.02							
31	Firm size $_{\text{SME}}$	−0.04	−0.37	0.17	−0.13	−0.08	−0.10	−0.11	−0.12	−0.01	−0.04	−0.02	0.05	−0.58						
32	Firm size $_{\text{LARGE}}$	0.00	0.01	0.02	−0.01	−0.01	−0.02	0.00	0.01	0.03	0.00	−0.01	−0.02	−0.10	−0.25					
32	Firm size $_{\text{VERY LARGE}}$	0.04	0.33	−0.17	0.12	0.08	0.11	0.10	0.10	−0.01	0.03	0.03	−0.03	0.60	−0.74	−0.47				

TABLE 3 Knowledge flows from other firms: Baseline model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	KFs Survey	**Back cites**	**App Back cites**	**KFs Survey**	**Back cites**	**App Back cites**	**KFs Survey**	**Back cites**	**App Back cites**
Channels of KF									
Patents	1.346	0.146	0.137	1.364	0.144	0.137	1.361	0.145	0.137
	(0.000)	(0.004)	(0.011)	(0.000)	(0.005)	(0.011)	(0.000)	(0.005)	(0.011)
Technical conferences	1.376	0.014	−0.009	1.398	0.015	−0.008	1.385	0.007	−0.017
	(0.000)	(0.766)	(0.829)	(0.000)	(0.755)	(0.853)	(0.000)	(0.883)	(0.693)
Informal interactions	1.028	0.083	0.107	1.027	0.083	0.108	1.026	0.084	0.112
	(0.000)	(0.096)	(0.018)	(0.000)	(0.095)	(0.018)	(0.000)	(0.092)	(0.015)
Formal collaborations	2.203	−0.027	−0.218	2.135	−0.031	−0.226	2.137	−0.032	−0.230
	(0.000)	(0.797)	(0.014)	(0.000)	(0.769)	(0.011)	(0.000)	(0.756)	(0.009)
Inventor mobility	0.395	0.458	0.202	0.381	0.454	0.199	0.382	0.452	0.191
	(0.003)	(0.000)	(0.036)	(0.004)	(0.000)	(0.038)	(0.003)	(0.000)	(0.046)
Type of R&D									
Published output	−0.560	−0.330	−0.060	−0.633	−0.323	−0.060	−0.632	−0.329	−0.068
	(0.000)	(0.017)	(0.645)	(0.000)	(0.021)	(0.643)	(0.000)	(0.018)	(0.591)
Patenting and citing behavior									
Patent effectiveness	0.350	0.177	0.210	0.280	0.161	0.196	0.267	0.158	0.188
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.002)	(0.000)
Family size	−0.082	0.952	0.529	−0.103	0.950	0.521	−0.103	0.949	0.518
	(0.234)	(0.000)	(0.000)	(0.134)	(0.000)	(0.000)	(0.133)	(0.000)	(0.000)
Claims	−0.176	0.485	0.410	−0.162	0.486	0.410	−0.166	0.485	0.409
	(0.004)	(0.000)	(0.000)	(0.007)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)
Forward citations	0.095	0.573	0.390	0.102	0.571	0.389	0.101	0.574	0.394
	(0.098)	(0.000)	(0.000)	(0.077)	(0.000)	(0.000)	(0.077)	(0.000)	(0.000)
Citing propensity	−0.087	0.945	0.624	−0.081	0.943	0.625	−0.076	0.946	0.625
	(0.251)	(0.000)	(0.000)	(0.285)	(0.000)	(0.000)	(0.317)	(0.000)	(0.000)

TABLE 3 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	KFs Survey	**Back cites**	**App Back cites**	**KFs Survey**	**Back cites**	**App Back cites**	**KFs Survey**	**Back cites**	**App Back cites**
Patent Strategy									
Used patent				1.187	0.279	0.208	1.283	0.235	0.270
				(0.000)	(0.015)	(0.076)	(0.000)	(0.027)	(0.008)
Strategic non-use$_{\text{Blocking}}$				0.342	0.344	0.170			
				(0.046)	(0.012)	(0.204)			
Strategic non-use$_{\text{Prevent suits}}$							0.749	0.418	0.399
							(0.000)	(0.005)	(0.006)
Controls									
Patent stock	−0.255	−0.383	−0.257	−0.181	−0.381	−0.253	−0.186	−0.383	−0.254
	(0.004)	(0.000)	(0.000)	(0.038)	(0.000)	(0.000)	(0.034)	(0.000)	(0.000)
US equivalent	−0.045	2.597	1.680	−0.045	2.596	1.679	−0.047	2.595	1.680
	(0.767)	(0.000)	(0.000)	(0.765)	(0.000)	(0.000)	(0.751)	(0.000)	(0.000)
Observations	12,619	12,619	12,619	12,619	12,619	12,619	12,619	12,619	12,619
Log likelihood	−40,141	−31,294	−17,227	−40,113	−31,290	−17,226	−40,106	−31,288	−17,221

Note. Values shown in the table are the average marginal effects in the importance of other firms as knowledge sources (KFs Survey), and the number of backward citations (Back Cites, App Back Cites) for a 1-*SD* increase in the continuous independent variable, or a change from 0 to 1 for discrete independent variables. A negative binomial model is used to compute the values reported in the table. *p values* are reported in parenthesis. All models include controls for priority office, legal status of the patent, and technological class. Standard are errors clustered at the applicant level.

shown by the positive association with KFs SURVEY, but the extent to which this knowledge flow also translates into prior art citations differs according to the strategic intent. In the case of STRATEGIC NON-USE$_{\text{BLOCKING}}$, they do not give rise to more applicant-added citations compared with sleeping patents (Column 6). Contrarily, in the case of STRATEGIC NON-USE$_{\text{PREVENT SUITS}}$, the number of applicant-added citations increases by 0.40 (Column 9). The average marginal effect estimated for this variable is 47.77% higher than the one associated with USED PATENT, and it entails a 20.24% increase in the average value of APP BACK CITES. These results suggest that firms have weak incentives to cite relevant prior art if the applications are filed to strategically preempt others from patenting similar inventions; the incentive to cite prior art is instead higher if they want to reduce the risk of litigation.

Among the control variables, we note that US EQUIVALENT is not correlated with the survey-based measure of knowledge flows, whereas it is strongly and positively associated with the citation-based measure of knowledge flows. In particular, the presence of a U.S. equivalent in the focal DOCDB family increases the number of applicant-added citations by 1.68, or 85.23%, with respect to the mean value of APP BACK CITES. These results suggest that the U.S. inequitable conduct doctrine triggers the inclusion of an overwhelming number of backward citations.[3] The applicant's PATENT STOCK, instead, negatively correlates with the number of applicant-added citations (-12.89%), which confirms earlier results about the role of the applicant's experience with the patenting process (Steensma et al., 2015). Finally, technology-based differences bear meaningful consequences for the strategic disclosure of known prior art (Jaffe & de Rassenfosse, 2017). The results reported in Section 5 of Appendix S1 reveal large differences across technologies. Specifically, discrete technologies show a negative correlation with the survey-based measure of knowledge flows (e.g., BIOTECHNOLOGY; PHARMACEUTICALS; ORGANIC CHEMISTRY) and a positive association with the number of applicant-added citations, suggesting that characteristics such as the degree of cumulativeness and complexity of the technology (Cohen, Goto, et al., 2002; von Graevenitz, Wagner, & Harhoff, 2013) may explain part of the incentive to cite prior art. Thus, for example, in discrete technological fields like chemistry, in which a single patent may be important to protect an entire invention (e.g., a protein), applicants pay particular attention to citing prior art to reduce the risk of patent invalidation.

4.2 Firms' patent strategies and the disclosure of prior art

To further explore our second research question and to understand whether the incentive to disclose known prior art differs according to the applicant's patent strategy, we estimate the same baseline regressions as those in Table 3 (columns 1–3) for the three separate samples of USED PATENTS, STRATEGIC NON-USE$_{\text{BLOCKING}}$ patents, and SLEEPING PATENTS$_{\text{BLOCKING}}$.

The results in Table 4 show a positive association between PATENTS and the survey-based measure of knowledge for the three subsamples; a 1-*SD* increase in PATENTS implies an increase of 1.45 points in KFs SURVEY for used patents, 1.39 points for strategically unused patents, and 1 point for sleeping patents. Thus, knowledge available in patented prior art is essential to develop new inventions, regardless of the patent use. However, the knowledge acquired from patents leaves paper trails in backward citations only for patents used commercially. A 1-*SD* increase in PATENTS is associated with a 0.21 surge in APP BACK CITES (Column 3). Instead, the same covariate does not correlate with backward citations in the case of STRATEGIC NON-USE$_{\text{BLOCKING}}$ patents as well as SLEEPING PATENTS$_{\text{BLOCKING}}$ (columns 6 and 9).

[3]We conduct additional analyses for the two groups of patents with and without a U.S. equivalent, and, separately, for those with and without a U.S. priority. Section 4 of Appendix S1 shows the estimated results that fully confirm the role of institutional factors in adding patent citations.

Results in Table 4 also reveal that, for the sample of Strategic Non-use$_{\text{Blocking}}$ patents, the average marginal effects of Family size (0.44) and Claims (0.24) in Column 6 are lower than those reported for the sample of Used Patents in Column 3 (0.70 and 0.54, respectively), while the magnitude of Citing propensity rises from 0.59, in the case of Used Patents, to 0.78, for Strategic Non-use$_{\text{Blocking}}$ patents. These heterogeneous effects of the variables across the three samples of patents suggest that the errors of omission and the errors of commission in backward citations are affected by the firm's patent strategies. In this respect, a noticeable result is that Strategic Non-use$_{\text{Blocking}}$ patents do not cite prior art even if patents are an important source of information for the invention.[4]

4.3 Unpacking knowledge sourcing

This part of the investigation explores the existence and magnitude of a common component of variation between the citation-based measure and the survey measure of knowledge flows, and it answers our third research question by unpacking the analysis according to the type of knowledge source used in the inventive process.

Columns 1–10 of Table 5 report the estimated results of five sets of regressions that employ the importance of each of the five types of external knowledge sources (i.e., customers, users, suppliers, competitors, and consulting R&D) as dependent variables, and the same covariates as in columns 4 and 7 of Table 3.

A few results concerning specific channels of knowledge flows are worth noting. Earlier Patents represent an important channel of information for the applicant, irrespective of the source. Especially in the case of KFs competitors (columns 7 and 8), a 1-*SD* increase in the importance of patents is reflected in a 57.80% increase in the odds of reporting a higher importance of competitors as a knowledge source, and this association is more than twice as large as the association between Patents and any of the other sources of knowledge.

Formal and informal collaborations with third parties are positively associated with all five dependent variables. The highest correlation (135.71% for formal collaborations and 44.01% for informal interactions) is with KFs suppliers (Column 6), while the lowest one (14.38 and 8.13%, respectively) involves the dependent variable KFs competitors (Column 8).

These findings suggest that, while knowledge deriving from suppliers is transmitted through both publicly available channels like patents and collaborative links (which often entail tacit or secret information), the knowledge arising from competitors is primarily transmitted by codified, publicly available information.

The variables Used patent and Strategic non-used$_{\text{Prevent suits}}$ are positively associated with each of the five sources of knowledge, although their magnitude varies across sources. When the dependent variable is KFs customers, KFs users, or KFs suppliers (columns 1–6), the percentage-change coefficient associated with Used patent is more than twice as large as the coefficient estimated for the covariate Strategic non-used$_{\text{Prevent suits}}$. Differently, for KFs competitors, the magnitude of the strategic variable is similar to that of the commercial-use variable (14.39 vs. 13.63%). Finally, for KFs consulting, the effect of Strategic non-used$_{\text{Prevent suits}}$ (23.82%) is much larger than the effect of Used patent (11.27%). The estimated effect for the variable Strategic non-used$_{\text{Blocking}}$, instead, shows no correlation with any of the knowledge sources, with the only important exception of KFs competitors (Column 7), suggesting that, relative to the sleeping baseline, patents filed to prevent

[4] We obtain similar results (available from the authors upon request) when we use the samples of Strategic non-use$_{\text{Prevent suits}}$ patents in place of Strategic non-use$_{\text{Blocking}}$.

TABLE 4 Knowledge flows from other firms: Analysis by type of patent strategy

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Used patent			Strategic Non-use$_{Blocking}$			Sleeping patent$_{Blocking}$		
	KFs Survey	**Back cites**	**App Back cites**	**KFs Survey**	**Back cites**	**App Back cites**	**KFs Survey**	**Back cites**	**App Back cites**
Channels of KF									
Patents	1.450	0.186	0.206	1.388	0.009	0.045	0.996	0.200	0.132
	(0.000)	(0.005)	(0.005)	(0.000)	(0.919)	(0.626)	(0.000)	(0.111)	(0.153)
Technical conferences	1.324	0.045	0.014	1.522	−0.013	−0.038	1.394	−0.026	−0.050
	(0.000)	(0.499)	(0.810)	(0.000)	(0.880)	(0.644)	(0.000)	(0.789)	(0.525)
Informal interactions	1.015	0.008	0.068	1.015	0.215	0.131	1.092	0.100	0.182
	(0.000)	(0.903)	(0.225)	(0.000)	(0.019)	(0.182)	(0.000)	(0.306)	(0.022)
Formal collaborations	2.272	0.022	−0.248	2.082	−0.094	−0.015	1.950	−0.076	−0.210
	(0.000)	(0.870)	(0.046)	(0.000)	(0.656)	(0.941)	(0.000)	(0.723)	(0.199)
Inventor mobility	0.322	0.586	0.234	0.361	0.302	0.082	0.665	0.149	0.147
	(0.047)	(0.000)	(0.081)	(0.181)	(0.143)	(0.674)	(0.050)	(0.501)	(0.454)
Type of R&D									
Published output	−0.508	−0.326	−0.071	−0.443	−0.404	−0.115	−1.271	−0.138	−0.056
	(0.006)	(0.046)	(0.640)	(0.138)	(0.185)	(0.665)	(0.001)	(0.587)	(0.778)
Patenting and citing behavior									
Patent effectiveness	0.254	0.143	0.197	0.262	0.240	0.289	0.412	0.094	0.102
	(0.001)	(0.029)	(0.004)	(0.015)	(0.008)	(0.002)	(0.010)	(0.313)	(0.271)
Family size	−0.111	1.099	0.704	−0.097	0.761	0.439	−0.133	0.712	0.180
	(0.202)	(0.000)	(0.000)	(0.484)	(0.000)	(0.000)	(0.436)	(0.000)	(0.071)
Claims	−0.073	0.609	0.541	−0.228	0.319	0.241	−0.326	0.291	0.216
	(0.370)	(0.000)	(0.000)	(0.047)	(0.002)	(0.008)	(0.013)	(0.007)	(0.018)
Forward citations	0.141	0.603	0.440	−0.185	0.643	0.455	0.421	0.314	0.175
	(0.047)	(0.000)	(0.000)	(0.092)	(0.000)	(0.000)	(0.007)	(0.000)	(0.015)
Citing propensity	−0.072	0.912	0.588	−0.231	1.067	0.784	0.299	0.872	0.684
	(0.464)	(0.000)	(0.000)	(0.074)	(0.000)	(0.000)	(0.118)	(0.000)	(0.000)

TABLE 4 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Used patent			Strategic Non-use$_{Blocking}$			Sleeping patent$_{Blocking}$		
	KFs Survey	Back cites	App Back cites	KFs Survey	Back cites	App Back cites	KFs Survey	Back cites	App Back cites
Controls									
Patent stock	−0.331	−0.318	−0.203	0.161	−0.386	−0.230	−0.360	−0.528	−0.406
	(0.000)	(0.000)	(0.003)	(0.250)	(0.000)	(0.003)	(0.036)	(0.000)	(0.000)
US equivalent	−0.111	2.729	1.743	0.134	2.604	1.797	0.068	2.080	1.326
	(0.563)	(0.000)	(0.000)	(0.628)	(0.000)	(0.000)	(0.844)	(0.000)	(0.000)
Observations	7,116	7,116	7,116	3,526	3,526	3,526	1,977	1,977	1,977
Log likelihood	−22,760	−17,695	−9,909	−11,133	−8,804	−4,792	−6,102	−4,709	−2,456

Note. Values shown in the table are the average marginal effects in the importance of other firms as knowledge sources (KFS SURVEY), and the number of backward citations (BACK CITES, APP BACK CITES) for a 1-*SD* increase in the continuous independent variable, or a change from 0 to 1 for discrete independent variables. A negative binomial model is used to compute the values reported in the table. All models include controls for priority office, legal status of the patent, and technological class. *p*-Values are reported in parenthesis. Standard errors are clustered at the applicant level.

TABLE 5 Knowledge flows from other firms: Analysis for separate sources of knowledge

	(1) KFs customers	(2)	(3) KFs users	(4)	(5) KFs suppliers	(6)	(7) KFs competitors	(8)	(9) KFs consulting	(10)	(11) App back cites	(12)
Channels of KF												
Patents	25.799	25.757	21.992	21.896	20.840	20.782	57.855	57.843	13.916	13.675	0.119	0.120
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.028)	(0.028)
Technical conferences	32.273	31.991	29.686	29.341	29.231	29.006	36.006	35.723	32.727	32.331	−0.017	−0.025
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.691)	(0.555)
Informal interactions	25.045	25.015	21.734	21.691	44.020	44.012	8.184	8.128	33.374	33.308	0.110	0.114
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.019)	(0.015)
Formal collaborations	60.234	60.306	39.797	39.939	135.597	135.714	14.344	14.382	51.554	51.768	−0.231	−0.233
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.011)	(0.010)
Inventor mobility	−1.643	−1.715	0.580	0.465	11.450	11.385	15.871	15.801	17.131	16.933	0.181	0.173
	(0.659)	(0.644)	(0.877)	(0.900)	(0.006)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.060)	(0.072)
Type of R&D												
Published output	−16.657	−16.625	−16.681	−16.582	−16.923	−16.897	−9.475	−9.563	13.485	14.049	−0.064	−0.073
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.026)	(0.025)	(0.014)	(0.010)	(0.624)	(0.569)
Patenting and citing behavior												
Patent effectiveness	13.137	12.906	5.287	4.894	2.477	2.284	3.326	3.246	−1.437	−2.167	0.202	0.194
	(0.000)	(0.000)	(0.002)	(0.005)	(0.152)	(0.186)	(0.057)	(0.063)	(0.408)	(0.215)	(0.000)	(0.000)
Family size	−3.185	−3.189	−4.159	−4.145	0.388	0.386	−4.019	−4.039	2.012	2.028	0.514	0.511
	(0.134)	(0.134)	(0.038)	(0.038)	(0.855)	(0.856)	(0.051)	(0.051)	(0.364)	(0.361)	(0.000)	(0.000)
Claims	−6.630	−6.698	−4.722	−4.821	−1.496	−1.551	−4.965	−5.012	3.484	3.434	0.407	0.405
	(0.000)	(0.000)	(0.007)	(0.006)	(0.422)	(0.406)	(0.008)	(0.007)	(0.086)	(0.090)	(0.000)	(0.000)
Forward citations	−0.012	0.000	−0.113	−0.092	0.454	0.477	7.412	7.446	5.876	5.919	0.381	0.386
	(0.994)	(1.000)	(0.945)	(0.955)	(0.787)	(0.776)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Citing propensity	3.059	3.172	−3.628	−3.536	−4.987	−4.951	3.299	3.395	−2.298	−2.261	0.636	0.637
	(0.273)	(0.256)	(0.104)	(0.114)	(0.025)	(0.027)	(0.150)	(0.140)	(0.319)	(0.327)	(0.000)	(0.000)
Patent strategy												
Used patent	44.128	46.182	46.563	52.351	25.466	27.590	14.719	13.632	0.281	11.274	0.232	0.299
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.002)	(0.955)	(0.010)	(0.048)	(0.003)

TABLE 5 (Continued)

	(1) KFs customers	(2)	(3) KFs users	(4)	(5) KFs suppliers	(6)	(7) KFs competitors	(8)	(9) KFs consulting	(10)	(11) App back cites	(12)
Strategic non-used$_{\text{Blocking}}$	7.090		6.208		4.437		10.850		−2.276		0.169	
	(0.172)		(0.234)		(0.412)		(0.043)		(0.665)		(0.202)	
Strategic non-used$_{\text{Prevent suits}}$		14.844		20.314		11.242		14.391		23.823		0.407
		(0.008)		(0.000)		(0.045)		(0.011)		(0.000)		(0.005)
Knowledge source												
KFs customers											−0.043	−0.046
											(0.423)	(0.388)
KFs users											−0.127	−0.131
											(0.014)	(0.011)
KFs suppliers											0.056	0.056
											(0.239)	(0.235)
KFs competitors											0.122	0.123
											(0.028)	(0.026)
KFs consulting											0.030	0.030
											(0.467)	(0.483)

Note. Values shown in columns 1 to 10 are percentage changes in the odds of reporting a higher importance of each knowledge source for a 1-*SD* increase in the continuous independent variable, or a change from 0 to 1 for discrete independent variables: an ordered logit model is used to compute values reported in these specifications. Values shown in columns 11 and 12 are average marginal effects in KFs Survey and App Back Cites: a negative binomial model is used to compute values reported in these specifications. All models include controls for priority office, legal status of the patent, technological class, applicant's patent stock, and presence of a U.S. equivalent in the DOCDB family. *p* Values are in parentheses. Standard errors are clustered at the applicant level. The regression sample comprises 12,619 observations.

others from developing similar inventions are more likely to draw on knowledge developed by competitors.

At the same time, knowledge drawn from competitors leaves a paper trial. Columns 11 and 12 in Table 5 show the association between the importance of the five knowledge sources and the variable App back cites. We note that the inclusion of these five covariates does not change the sign, magnitude, and statistical significance of the other covariates compared with those reported in Table 3. Moreover, only the use of two knowledge sources translates into the disclosure of known prior art, albeit in opposite directions. Specifically, a 1-*SD* increase in KFs users decreases the number of backward citations by 0.13 units, while a 1-*SD* increase in KFs competitors raises the number of backward citations by 0.12 units**.**

Section 6 of Appendix S1 confirms these associations when the five sources of knowledge are included separately in the regression model (Table S6a), and it explores the origin of measurement errors according to the type and importance of the different sources of knowledge used in the inventive process (Table S6b). The estimated results show that the errors of omission vary across different types of knowledge sources, whereas the errors of commission are pervasive across all five sources of knowledge. At the same time, these findings offer compelling evidence that backward citations reflect the exploitation of knowledge when this is drawn from competitors, possibly because this attenuates the risk of litigation.

4.4 Robustness checks

We conducted four robustness tests. First, we estimate the specifications in Table 3 with a different survey-based dependent variable that we built to measure the intensity of the use of external knowledge sources. Second, we employ a random-coefficient regression model to estimate the equations. The results of these two tests are reported in sections 7 and 8 of the Appendix S1 and are fully consistent with those presented in Table 3. Third, drawing on earlier studies suggesting that organizational characteristics provide differential incentives to disclose known prior art (e.g., Steensma et al., 2015), we perform our regressions on three separate subsamples of patents applied for by small, medium, and large firms, respectively. The estimated results in Section 9 of the Appendix S1 reveal a significant difference between large and small firms, with small firms showing a more marked nondisclosure behavior of prior art than large firms. Finally, we quasi-replicate the study of knowledge flows from PROs by Roach and Cohen (2013). In the remainder of this section we reproduce the analysis in Roach and Cohen (2013) on the validity of patent citations as measures of knowledge flows from PROs. This exercise, which exploits a different data set and different, albeit comparable, measures (Bettis, Helfat, & Shaver, 2016), allows us to ascertain that our findings do not depend on the peculiarities of our data or research setting.

The InnoS&T survey asks inventors to rate the importance of two sources of public knowledge on a 6-point Likert scale (0 = not used; 5 = very important): universities or other education institutions, and public research institutions. We construct the variable KFs $\text{survey}_{\text{PRO}}$ as the maximum score between the two sources. The two citation-based measures of knowledge flows are Back $\text{cites}_{\text{PRO}}$, which counts all backward citations to PROs' patents, and App back $\text{cites}_{\text{PRO}}$, which counts only applicant-added, backward citations to PROs' patents.

In line with Roach and Cohen (2013), we account for the open science feature of public research with three variables: Scientific publications, Conferences, and Informal interactions, all three measured on a 5-point Likert scale (1 = not important/not used; 5 = very important).[5] To account for the importance of interactions stemming from contract-based relationships with PROs, we employ the

[5]To make the analysis of knowledge flows from PROs comparable with Roach and Cohen (2013), we used Scientific publications instead of Patents and Industrial scientist instead of Inventor mobility.

dichotomous variable Formal collaborations$_{PRO}$, which may not be readily reflected in patent citations because they often produce documents that are not publicly disclosed, and they typically involve face-to-face and confidential communication (Cohen, Nelson, & Walsh, 2002). Roach and Cohen (2013) hypothesize that public research that is used to complete existing projects is less likely to be traced by patent citations than public research that triggers their initiation. We construct the variable Completion of invention, which measures on a 5-point Likert scale the importance of reading scientific articles for the completion of the invention. To identify the role of the errors of commission, we maintain the same set of covariates used in the analysis of knowledge flows from other firms.[6]

Table 6 reports the results of this analysis. Columns 1 through 6 involve factors that can spur errors of omission in citation-based indicators of knowledge flows. Columns 7 through 9 explore the existence of errors of commission, and columns 10 through 12 report the estimated results for the full model, which we comment on in this section.

All three dimensions of open science reflect knowledge flows from PROs (Column 10). Scientific publications play a prominent role: a 1-*SD* increase in Scientific publications implies a 53.40% rise in the odds of reporting a higher value of KFs survey$_{PRO}$. However, only the variable Scientific publications is significantly correlated with the citation-based measure of knowledge flows; the estimated increase in the expected number of backward citations is 23.20% (Column 12). Formal collaborations are an important bridge to public research: a 1-*SD* increase in Formal collaborations$_{PRO}$ entails a 270.37% upsurge in KFs survey$_{PRO}$ and a 42.11% increase in App back cites$_{PRO}$. Finally, both dependent variables reflect knowledge inflows through highly educated scientists: KFs survey$_{PRO}$ increases by 29.97%, and App back cites$_{PRO}$ increases by 49.62% when the inventor holds a PhD. The results in Table 6 suggest that there are also errors of commission in citation-based indicators of knowledge flows from PROs: the covariates gauging the patenting behavior of the applicant are not significantly associated with KFs survey$_{PRO}$, but they are positively related to App back cites$_{PRO}$.

Thus, overall, our results support Roach and Cohen's (2013) findings and highlight additional, interesting correlations. First, backward citations capture some dimensions of open science, that is, knowledge drawn from scientific publications. Second, formal collaborations are associated with both the survey-based and the citation-based indicators, thus suggesting that they entail the transfer of codified knowledge that PROs protect through patents. Third, our survey measure of basic research conducted by the firm is correlated with patent citations to public research, which suggests that PROs are important as a source of knowledge for firms' basic research, even when the latter leads to patentable results.

5 DISCUSSION AND CONCLUSIONS

This paper investigates the extent to which patent citations reflect knowledge flows in a context where firms open their innovation activities and draw from external knowledge sources. Based on Roach and Cohen's (2013) methodology, we compare patent citations with inventors' accounts of knowledge sourcing during the inventive process and find evidence of both errors of omission and errors of commission in patent citations.

Our findings indicate that patent citations capture knowledge flows that occur through the reading of other firms' patents, informal interactions with employees of other organizations, and inventors'

[6]The correlation matrix and descriptive statistics for the variables used in this part of the paper are in Section 10 of the online appendix. Notice that, unlike Roach and Cohen (2013), we could not retrieve the backward citations made to nonpatent literature.

TABLE 6 Knowledge flows from public research organizations (PROs)

	(1) KFs $SURVEY_{PRO}$	(2) Back $CITES_{PRO}$	(3) App back $CITES_{PRO}$	(4) KFs $SURVEY_{PRO}$	(5) Back $CITES_{PRO}$	(6) App back $CITES_{PRO}$	(7) KFs $SURVEY_{PRO}$	(8) Back $CITES_{PRO}$	(9) App back $CITES_{PRO}$	(10) KFs $SURVEY_{PRO}$	(11) Back $CITES_{PRO}$	(12) App back $CITES_{PRO}$
Channels of KF												
Scientific publications	95.747	48.362	49.886							53.402	25.620	23.196
	(0.000)	(0.000)	(0.000)							(0.000)	(0.000)	(0.005)
Technical conferences	31.651	−0.874	1.349							26.021	−3.502	0.132
	(0.000)	(0.787)	(0.791)							(0.000)	(0.293)	(0.978)
Informal interactions	23.556	1.341	−4.042							23.154	−0.069	−3.461
	(0.000)	(0.626)	(0.395)							(0.000)	(0.980)	(0.366)
Formal $COLLABORATIONS_{PRO}$	296.089	60.971	44.960							270.365	71.425	42.110
	(0.000)	(0.000)	(0.002)							(0.000)	(0.000)	(0.005)
Industrial scientist	41.673	54.979	61.105							29.967	53.589	49.615
	(0.000)	(0.000)	(0.000)							(0.000)	(0.000)	(0.000)
Type and use of R&D												
Published output				86.519	57.321	42.354				28.867	37.361	28.839
				(0.000)	(0.000)	(0.002)				(0.000)	(0.000)	(0.031)
Completion of invention				123.077	47.132	52.746				45.429	14.756	16.429
				(0.000)	(0.000)	(0.000)				(0.000)	(0.004)	(0.035)
Patenting and citing behavior												
Patent effectiveness							2.707	−5.028	8.687	−1.983	−6.017	14.491
							(0.117)	(0.116)	(0.186)	(0.252)	(0.069)	(0.031)
Family size							3.192	30.117	44.627	1.726	30.822	45.654
							(0.163)	(0.000)	(0.000)	(0.413)	(0.000)	(0.000)
Claims							5.126	21.334	28.317	−1.005	20.084	25.709
							(0.007)	(0.000)	(0.000)	(0.580)	(0.000)	(0.000)
Forward citations							8.937	19.431	26.262	3.769	16.716	26.737
							(0.000)	(0.000)	(0.000)	(0.042)	(0.000)	(0.000)

TABLE 6 (Continued)

	(1) KFs SURVEY_{PRO}	(2) Back CITES_{PRO}	(3) App back CITES_{PRO}	(4) KFs SURVEY_{PRO}	(5) Back CITES_{PRO}	(6) App back CITES_{PRO}	(7) KFs SURVEY_{PRO}	(8) Back CITES_{PRO}	(9) App back CITES_{PRO}	(10) KFs SURVEY_{PRO}	(11) Back CITES_{PRO}	(12) App back CITES_{PRO}
Citing propensity							−4.690	20.003	36.296	−5.343	16.425	27.141
							(0.039)	(0.000)	(0.000)	(0.018)	(0.000)	(0.000)
Controls												
Patent stock	−9.173	−13.055	−20.824	−8.256	−13.720	−22.332	−5.134	−22.858	−34.045	−6.472	−21.663	−29.883
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.020)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)
US equivalent	2.599	194.537	1,127.266	3.622	196.029	1,138.337	6.114	118.013	748.852	−0.301	85.353	459.100
	(0.522)	(0.000)	(0.000)	(0.359)	(0.000)	(0.000)	(0.173)	(0.000)	(0.000)	(0.946)	(0.000)	(0.000)
Observations	12,423	12,423	12,423	12,423	12,423	12,423	12,423	12,423	12,423	12,423	12,423	12,423
Pseudo R2	0.0994	0.150	0.181	0.0702	0.145	0.177	0.0160	0.144	0.191	0.106	0.169	0.207

Note: An ordered logit model is used in regressions where the dependent variable is KFs SURVEY_{PRO}. A negative binomial model is used in regressions where the dependent variable is Back Cites_{PRO} and App Back Cites_{PRO}. The values shown in the table are the percentage change in the odds of reporting a higher importance of other firms as knowledge sources (KFs SURVEY_{PRO}), and in the expected number of backward citations (Back Cites_{PRO}, App Back Cites_{PRO}) for a 1-*SD* change in the continuous independent variable, or a change from 0 to 1 for discrete independent variables. *p* values are reported in parentheses. All models include controls for priority office, legal status of the patent, and technological class. Standard errors are clustered at the applicant level.

mobility. However, they do not reflect knowledge acquired through technical conferences and formal collaborations. In addition, they are strongly associated with several other dimensions that do not involve any transfer of knowledge. We also explore the reasons for these measurement errors, whether they vary with the type of knowledge source (e.g., customers, competitors), and the circumstances in which researchers can safely use citations to measure interfirm knowledge flows.

The origins of these measurement errors are rooted in the firms' contrasting incentives to cite prior art (i.e., broad patent scope vs. low risk of invalidation), which, in turn, depend on some key factors. A first important factor is the patenting system and regulations, such as the U.S. inequitable conduct doctrine that triggers the inclusion of a large number of backward citations. Likewise, the technological field contributes to shaping the citation incentives: applicants in discrete technologies (e.g., biotechnology) disclose more prior art than those in complex technologies (e.g., telecommunications). This is because of the lower cost of not surviving a validity challenge for a patent in complex technologies, where several patented interconnected inventions enter a single product, compared with discrete technologies, in which the invalidation of a single patent due to incomplete disclosure of prior art can be much more damaging. Larger firms are also less likely to withhold relevant prior art because of the high cost they would suffer from a precautionary injunction to stop production or to invalidate a patent used in commerce. Finally, firms' patent strategies play a key role. Recent contributions that address the strategic origin of backward citations (Cotropia et al., 2013; Lampe, 2012; Somaya, 2012) acknowledge that a better understanding of the predictors of backward citations "holds significant implications for firms competing through their patent rights and limiting those of others" (Steensma et al., 2015, p. 1187). Our study contributes to advance this understanding, and it shows that the applicant's citing behavior is closely associated with the pursuit of specific patent strategies. Patents that are filed to preempt others from patenting similar inventions are less likely to disclose prior art, even if this prior art was an important source of information to develop an invention. In contrast, the propensity to disclose prior art rises in the case of patents used commercially or filed to prevent infringement suits.

We also show that firms' patent strategies are closely linked to the use of specific knowledge sources, and, in turn, to the propensity to cite prior art. The establishment of this relationship helps bridge the patent strategy research (e.g., Lampe, 2012; Sampat, 2010; Somaya, 2012; Steensma et al., 2015) with the open innovation perspective (e.g., Chatterji & Fabrizio, 2014; Chesbrough, 2003; Laursen & Salter, 2006; Leiponen & Helfat, 2010; Rosenkopf & Nerkar, 2001), and it highlights that applicants draw on a limited set of external sources of knowledge (i.e., competitors) when pursuing a patent strategy to preempt other patents, whereas they engage with a wider range of external sources when adopting a patent strategy to mitigate the risk of litigation. Preempting other patents reveals a protection strategy focused on preventing knowledge spillovers, which may discourage investment in external knowledge screening and sourcing (Wadhwa et al., 2017).

It also suggests that, under specific conditions, citations can opportunely track knowledge flows. Thus, for example, compared with knowledge sources from users, customers, and suppliers, knowledge flows from competitors are likely to be traced by backward citations.

These findings have implications for research on the relationship between firms' open innovation strategies and patent protection mechanisms, as they show that there is large heterogeneity in the relationship between the different patent strategies that firms pursue and the breadth and depth of external knowledge search.

Our study can help future empirical research on open innovation to distinguish contexts where citations accurately measure knowledge flows (as in the case of knowledge flows from competitors) from settings where their validity is arguable (as in the case of knowledge flows from users). It also offers insights on the use of citations as a measure of knowledge flows. First, exploiting only applicant-added citations to build citation-based indicators of knowledge flows is recommended because it neutralizes the confounding effect surrounding examiner-added citations and washes out some of the noise that would otherwise hide or diminish the effect of several channels through which knowledge flows between firms. Second, accounting for technological differences across patents is important. Our analysis highlights that technology-related dummies differentially influence the survey- and citation-based indicators of knowledge flows and influence the marginal effects of other covariates especially when the dependent variable is the citation-based indicator. Moreover, in some technological fields (e.g., chemistry or electrical machinery, for opposite reasons) researchers must be even more cautious in using backward citations as a measure of knowledge flows. Last, researchers may want to use information at the level of the patent family to explore how institutional differences between different patent systems affect the incentive to cite prior art.

These recommendations can be implemented easily in empirical studies that use the patent family or the research project as the unit of analysis. However, in more aggregate research settings, in which the unit of analysis is the business unit or the whole firm and their portfolio of patents, the insights for the use of patent citations as a measure of knowledge flows are more nuanced. We show that the errors of omission and commission are strictly related to specific patent characteristics, such as the expected use of the patent, the filing jurisdiction, and the specific technological class of the underlying invention. The immediate implication of this finding is that, at the aggregate firm-level, one should be aware of these patent-specific factors and take them into account, for example, by analyzing the composition of the firm's patent portfolio in terms of technologies, filing jurisdictions, and share of patents that translate to commercial applications. Another, subtler implication for future empirical work that takes this heterogeneity into account concerns the consequences of averaging-out indicators that, especially in the case of large, technologically diversified companies, may produce contrasting effects and cancel each other out (e.g., patent portfolios containing patents filed with both the EPO and USPTO, or inventions in both discrete and complex technologies), making the use of patent citations a highly noisy indicator of knowledge flows. This is confirmed by the results of a separate analysis (available from the authors) that estimates our baseline model using the firm-average values of the variables factored into the models to compute the between-firm effects. Although the results of this aggregate analysis are consistent with those presented in Table 3, the magnitude and statistical significance of some of the covariates are much lower than those for the patent-level estimates (i.e., PATENTS, INFORMAL INTERACTION with respect to the citation-based indicator of knowledge flows, and PATENT EFFECTIVENESS with respect to the survey-based measure).

Our work is subject to some limitations. One is that survey-based data as a measure of knowledge flows does also have limitations. Although inventors are key informants, we cannot completely rule out the possibility of biases due to the subjectivity of some items, retrospective biases, and common method bias. To moderate the effects of these potential biases, we constructed our variables drawing on items employed in previous studies (Torrisi et al., 2016). In addition, several variables are based on data from various secondary sources, and all survey items were tested during the pilot stage of the survey. More precisely, we conducted three pretests aiming to check, respectively, the face validity of the final version of the questionnaire (e.g., the inventor's experience with the questionnaire), the different response rates between “paper and pencil” and online survey instruments, and the reasons for not responding to the survey. Despite some reasons for not answering the questionnaire, such as time restrictions, confidentiality issues, and general suspicion of surveys, the inventors generally described the questionnaire as clear and easy to handle. Finally, respondents did not know the research questions at the time of the survey (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). In addition, because of the cross-sectional nature of our data, we cannot rule out the possibility that unobserved heterogeneity at the applicant level may confound the results discussed in the paper, although the multilevel, mixed-effect regression model addresses some of these concerns and corroborates the findings of our inquiry for the subset of applicants with repeated observations. Nonetheless, we believe that our investigation uncovers meaningful correlations between the applicant's patent strategies and the disclosure of prior art.

Finally, this paper does not directly observe the strategic interactions of the firm with other firms. However, our findings on the association between particular external knowledge sources (e.g., competitors, users), patent strategies, and backward citations are suggestive of the strategic interaction of the firm in the technology and the product markets. Future studies could build upon our results to match patent and citation data with finer-grained information on firms' interactions in the technology and product markets.

ACKNOWLEDGEMENTS

We are grateful to Editor Constance Helfat and two anonymous reviewers for their feedback and suggestions during the review process. We thank the participants at the 2017 Annual Meeting of the Academy of Management, the 2016 Annual Conference of the European Policy for Intellectual Property Association and a 2016 Bocconi University workshop. We also thank Gianluca Tarasconi and ICRIOS-Bocconi University for providing us with data on EPO patents. Finally, we acknowledge the European Commission [Contract FP7-SSH-2007-1- grant agreement No. 217299] for supporting the creation of the dataset, and the Italian Ministry of Education, University and Research [Projects #2010H37KAW and CUP B41J12000160008]. All errors are ours.

REFERENCES

Agarwal, R., Ganco, M., & Ziedonis, R. (2009). Reputations for toughness in patent enforcement: Implications for knowledge spillovers via inventor mobility.*Strategic Management Journal*, *30*(13), 1349–1374.

Akers, N. (2000). The referencing of prior art documents in European patents and applications.*World Patent Information*, *22*(4), 309–315.

Alcácer, J., & Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations.*The Review of Economics and Statistics*, *88*(4), 774–779.

Alcácer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in U.S. patents: An overview and analysis.*Research Policy*, *38*(2), 415–427.

Allison, J. R., & Lemley, M. A. (1998). Empirical evidence on the validity of litigated patents.*AIPLA Quarterly Journal*, *26*(3), 185–275.

Allison, J. R., Lemley, M. A., Moore, K. A., & Trunkey, R. D. (2004). Valuable patents.*The Georgetown Law Journal*, *92*(3), 435–479.

Arora, A., Cohen, W. M., & Walsh, J. P. (2016). The acquisition and commercialization of invention in American manufacturing: Incidence and impact.*Research Policy*, *45*(6), 1113–1128.

Atal, V., & Bar, T. (2010). Prior art: To search or not to search.*International Journal of Industrial Organization*, *28*(5), 507–521.

Baldwin, C., & von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation.*Organization Science*, *22*(6), 1399–1417.

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication.*Strategic Management Journal*, *37* (11), 2193–2203.

Chatterji, A. K., & Fabrizio, K. R. (2014). Using users: When does external knowledge enhance corporate product innovation?*Strategic Management Journal*, *35*(10), 1427–1445.

Chesbrough, H. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Boston, MA: Harvard Business School Press.

Coffano M, Tarasconi G. 2014. *CRIOS–Patstat database: Sources, contents and access rules* (CRIOS Working Paper No. 1). Milan, Italy: Center for Research on Innovation, Organization and Strategy, Bocconi University. doi:https://doi.org/10.2139/ssrn.2404344.

Cohen, W. M., Goto, A., Nagata, A., Nelson, R. R., & Walsh, J. P. (2002). R&D spillovers, patents and the incentives to innovate in Japan and the United States.*Research Policy*, *31*(8–9), 1349–1367.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). Links and impacts: The influence of public research on industrial R&D.*Management Science*, *48*(1), 1–23.

Cotropia, C. A. (2009). Modernizing patent law's inequitable conduct doctrine.*Berkeley Technology Law Journal*, *24*(2), 723–783.

Cotropia, C. A., Lemley, M. A., & Sampat, B. (2013). Do applicant patent citations matter?*Research Policy*, *42*(4), 844–854.

Criscuolo, P., & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents.*Research Policy*, *37*(10), 1892–1908.

Duguet, E., & MacGarvie, M. (2005). How well do patent citations measure flows of technology? Evidence from French innovation surveys.*Economics of Innovation and New Technology*, *14*(5), 375–393.

Guellec, D., Martinez, C., & Zuniga, P. (2012). Pre-emptive patenting: Securing market exclusion and freedom of operation.*Economics of Innovation and New Technology*, *21*(1), 1–29.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations.*The RAND Journal of Economics*, *36*(1), 16–38.

Hall, B. H., & Ziedonis, R. H. (2001). The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979–1995.*The RAND Journal of Economics*, *32*(1), 101–128.

Harhoff, D., & Reitzig, M. (2004). Determinants of opposition against EPO patent grants–The case of biotechnology and pharmaceuticals.*International Journal of Industrial Organization*, *22*(4), 443–480.

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents.*Economics Letters*, *105*(3), 287–289.

Hoisl, K. (2007). Tracing mobile inventors – The causality link between inventor mobility and inventor productivity.*Research Policy*, *36*(5), 619–636.

Jaffe, A. B., & de Rassenfosse, G. (2017). Patent citation data in social science research: Overview and best practices.*Journal of the Association for Information Science and Technology*, *68*(6), 1360–1374.

Jaffe, A. B., & Lerner, J. (2004). *Innovation and its discontents: How our broken patent system is endangering innovation and progress, and what to do about it*. Princeton, NJ: Princeton University Press.

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors.*The American Economic Review*, *90*(2), 215–218.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations.*The Quarterly Journal of Economics*, *108*(3), 577–598.

Kesan, J. P. (2002). Carrots and sticks to create a better patent system.*Berkeley Technology Law Journal*, *17*(2), 763–797.

Lampe, R. (2012). Strategic citation.*The Review of Economics and Statistics*, *94*(1), 320–333.

Langinier, C., & Marcoul, P. (2016). The search of prior art and the revelation of information by patent applicants.*Review of Industrial Organization*, *49*(3), 399–427.

Laursen, K., & Salter, A. (2006). Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms.*Strategic Management Journal*, *27*(2), 131–150.

Laursen, K., & Salter, A. (2014). The paradox of openness: Appropriability, external search and collaboration.*Research Policy*, *43*(5), 867–878.

Leiponen, A., & Helfat, C. E. (2010). Innovation objectives, knowledge sources, and the benefit of breadth.*Strategic Management Journal*, *31*(2), 224–236.

Martínez, C. (2011). Patent families: When do different definitions really matter?*Scientometrics*, *86*(1), 39–63.

Marx, M., Strumsky, D., & Fleming, L. (2009). Mobility, skills, and the Michigan non-compete experiment.*Management Science*, *55* (6), 875–889.

Mattes, E., Stacey, M. C., & Marinova, D. (2006). Surveying inventors listed on patents to investigate determinants of innovation.*Scientometrics*, *69*(3), 475–498.

Moser, P., Ohmstedt, J., & Rhode, P. W. (2018). Patent citations – an analysis of quality differences and citing practices in hybrid corn.*Management Science*, *64*(4), 1926–1940.

Nelson, A. J. (2009). Measuring knowledge spillovers: What patents, licenses and publications reveal about innovation diffusion. *Research Policy*, *38*(6), 994–1005.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies.*The Journal of Applied Psychology*, *88*(5), 879–903.

Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research.*Management Science*, *59*(2), 504–525.

Rosenkopf, L., & Nerkar, A. (2001). Beyond local search: Boundary-spanning, exploration, and the impact in the optical disk drive industry.*Strategic Management Journal*, *22*(4), 287–306.

Sampat, B. N. (2010). When do applicants search for prior art?*The Journal of Law and Economics*, *53*(2), 399–416.

Schmoch, U. (2008). *Concept of a technology classification for country comparisons. Final Report to the World Intellectual Property Organisation (WIPO)*. Geneva, Switzerland: WIPO Retrieved from: http://publica.fraunhofer.de/documents/N-114385.html

Somaya, D. (2012). Patent strategy and management.*Journal of Management*, *38*(4), 1084–1114.

Steensma, H. K., Chari, M., & Heidl, R. (2015). The quest for expansive intellectual property rights and the failure to disclose known relevant prior art.*Strategic Management Journal*, *36*(8), 1186–1204.

Thompson, P., & Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment.*The American Economic Review*, *95*(1), 450–460.

Tijssen, R. J. W. (2002). Science dependence of technologies: Evidence from inventions and their inventors.*Research Policy*, *31*(4), 509–526.

Torrisi, S., Gambardella, A., Giuri, P., Harhoff, D., Hoisl, K., & Mariani, M. (2016). Used, blocking and sleeping patents: Empirical evidence from a large-scale inventor survey.*Research Policy*, *45*(7), 1374–1385.

von Graevenitz, G., Wagner, S., & Harhoff, D. (2013). Incidence and growth of patent thickets: The impact of technological opportunities and complexity.*The Journal of Industrial Economics*, *61*(3), 521–563.

Wadhwa, A., Bodas Freitas, I. M., & Sarkar, M. B. (2017). The paradox of openness and value protection strategies: Effect of extramural R&D on innovative performance.*Organization Science*, *28*(5), 873–893.

Zobel, A. K., Balsmeier, B., & Chesbrough, H. (2016). Does patenting help or hinder open innovation? Evidence from new entrants in the solar industry.*Industrial and Corporate Change*, *25*(2), 307–331.