Management Science

Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research

Michael Roach, Wesley M. Cohen,

Please scroll down for article—it is on subsequent pages

Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research

Michael Roach

Fuqua School of Business, Duke University, Durham, North Carolina 27708, michael.roach@duke.edu

Wesley M. Cohen

Fuqua School of Business, Duke University, Durham, North Carolina 27708; and
National Bureau of Economic Research, Cambridge, Massachusetts 02138, wcohen@duke.edu

This paper assesses the validity and accuracy of firms' backward patent citations as a measure of knowledge flows from public research by employing a newly constructed data set that matches patents to survey data at the level of the research and development lab. Using survey-based measures of the dimensions of knowledge flows, we identify sources of systematic measurement error associated with backward citations to both patent and nonpatent references. We find that patent citations reflect the codified knowledge flows from public research, but they appear to miss knowledge flows that are more private and contract based in nature, as well as those used in firm basic research. We also find that firms' patenting and citing strategies affect patent citations, making citations less indicative of knowledge flows. In addition, an illustrative analysis examining the magnitude and direction of measurement error bias suggests that measuring knowledge flows with patent citations can lead to substantial underestimation of the effect of public research on firms' innovative performance. Throughout our analyses we find that nonpatent references (e.g., journals, conferences, etc.), not the more commonly used patent references, are a better measure of knowledge originating from public research.

Key words: patent citations; knowledge flows; academic research; innovation; measurement
History: Received October 10, 2010; accepted December 24, 2011, by Lee Fleming, entrepreneurship and innovation. Published online in *Articles in Advance* November 9, 2012, and updated January 15, 2013.

1. Introduction

Knowledge flows have long been thought to be a critical determinant of firm innovation. Notwithstanding the difficulty of observing knowledge flows (Jaffe et al. 1993), patent citations have been used extensively to measure knowledge flows from universities (Jaffe et al. 1993, Narin et al. 1997, Henderson et al. 1998, Gittelman and Kogut 2003, Sorenson and Fleming 2004, Branstetter and Ogura 2005); within and between firms (Almeida and Kogut 1999, Rosenkopf and Nerkar 2001, Rosenkopf and Almeida 2003, Singh and Agrawal 2011); and across geographic boundaries (Duguet and MacGarvie 2005, Peri 2005, Singh 2005, MacGarvie 2006). Indeed, patent citations are the most widely employed measure of knowledge flows in the economics, management, and policy literatures. This widespread use reflects the attractive features of patent citations as measures, notably their comprehensive coverage across industries and firms, and over time. Moreover, they are readily available from public sources such as the U.S. Patent and Trademark Office (USPTO) and the National Bureau of Economic Research (NBER) patent database (Hall et al. 2001), as well as from private sources.

To date, however, we know surprisingly little about how well citations measure knowledge flows, though a number of studies provide grounds for skepticism (Jaffe et al. 1998, Agrawal and Henderson 2002, Jaffe et al. 2002, Duguet and MacGarvie 2005, Alcacer and Gittelman 2006, Alcacer et al. 2009, Lampe 2012). Some of these studies note that patent citations may be noisy measures of knowledge flows (Jaffe et al. 1998, Agrawal and Henderson 2002, Jaffe et al. 2002), and others highlight that the large share of citations contributed by patent examiners may not accurately reflect the knowledge used by the patenting firm (Alcacer and Gittelman 2006, Alcacer et al. 2009). Although these studies call into question how well patent citations may serve as indicators of knowledge flows, the absence of alternative, comparable measures of knowledge flows has limited our ability to assess whether citations are simply noisy measures, as widely assumed, or whether they are subject to sources of systematic measurement error. Such error may not only compromise the use of citations as an indicator of knowledge flows from public research, but may also compromise their role as a dependent variable in regression analyses of knowledge flows or

as a regressor in analyses of innovation and productivity growth.

In this paper, we use a newly developed data set that matches patent citations to contemporaneous survey reports from research and development (R&D) lab managers to assess citations as a measure of one particular type of knowledge flows—those originating from public research institutions (e.g., universities and federal labs). Provided by the Carnegie Mellon Survey of the Nature and Determinants of Industrial R&D (Cohen et al. 2002), these managers' reports on the nature and impact of public research allow us to identify dimensions of knowledge flows that citations should capture and do, as well as to search for two sources of measurement error: dimensions of knowledge flows that citations should capture but do not, which we term "errors of omission," and factors that citations reflect but are not informative of knowledge flows, which we call "errors of commission."

In brief, although patent citations appear to reflect selected aspects of knowledge flows, they also exhibit strong evidence of errors of omission and errors of commission. For example, we find that citations reflect knowledge flows through the channels of open science, such as scientific publications, as well as the contribution of public research in stimulating new R&D projects. At the same time, illustrating errors of omission, we find that citations fail to reflect knowledge flows through contract-based relationships with academic scientists (e.g., consulting or cooperative R&D ventures), possibly because of the private, less codified way in which knowledge moves through this channel. Citations similarly do not reflect the contribution of public research to firms' basic research, perhaps because the outputs of basic research are less likely to be patented relative to those of applied research and development. Nor do citations fully reflect how firms use public research to solve technical challenges encountered in the conduct of R&D. Indicative of errors of commission, patent citations are influenced by firms' appropriability strategies and strategic citing practices in ways that are not revealing of knowledge flows. In all of our analyses, we find that citations to *nonpatent* references, such as scientific journal articles, correspond more closely to managers' reports of the use of public research than do the more commonly employed citations to *patent* references. Finally, an illustrative regression analysis suggests that using patent citations as a measure of knowledge flows from public research can lead to a substantial underestimate of the impact of public research on firms' innovative performance.

2. Background and Approach

Given that patent citations are among the most widely used measures of knowledge flows, an understanding of what they reflect about knowledge flows should inform our interpretation of results that rely upon them, as well as guide efforts to improve their utility. There are compelling reasons to believe that citations may not accurately reflect the contribution of knowledge flows to industrial R&D. First, not all innovations are patented (Scherer 1983, Griliches 1990, Cohen et al. 2000). Second, not all knowledge flows are cited or even citable (Griliches 1990, Pavitt 1991). Third, one may question how well patent citations reflect knowledge flows because their purpose, unlike citations in academic publications, is not to identify the antecedent knowledge upon which a given invention or discovery is built, but rather to delimit the scope of the patented invention (e.g., Jaffe et al. 1993). Moreover, what is cited is influenced not only by the inventor, but also by firms' citing strategies (Lampe 2012), by patent attorneys, and by patent examiners (Alcacer and Gittelman 2006, Alcacer et al. 2009).

Why should we care about the accuracy of citations as a measure of the contribution of public research to industrial innovation? First, citation-based measures have been used to characterize the importance of public research to industrial innovation (e.g., Narin et al. 1997) and, in turn, to justify federal support of public research (National Science Board 2012). Second, they have been used as dependent variables in regression analyses examining either knowledge outflows from public research to industry (Jaffe et al. 1993) or the importance of university patents (Henderson et al. 1998, Mowery et al. 2002, Mowery and Ziedonis 2002). Third, they have been used as regressors in models explaining the influence of public science on invention (Fleming and Sorenson 2004, Sorenson and Fleming 2004) and industrial R&D productivity (Gittelman and Kogut 2003).

Where citations are used as dependent variables, biased estimates will result if the associated measurement error is correlated with any independent variables. Even if such error is orthogonal to the independent variables, the predictive power of the model may be compromised. When citation-based measures are employed as independent variables, the consequences are more complicated. At best, we can expect attenuation of the estimated coefficients, assuming the measurement error is classical. If the measurement error is correlated with other independent variables, however, the direction and magnitude of bias may extend beyond the mismeasured variable to affect other variables of interest (Bound et al. 2001). In any event, we can expect bias and inconsistency in coefficient estimates.

Patent citations may be associated with two types of measurement error. First, what we call "errors of omission" may occur when knowledge originating from public research is not reflected in patent

citations. For example, if we take the common definition of basic research seriously—that it is conducted to achieve "...fuller knowledge or understanding of the fundamental aspects of phenomena and of observable facts without specific applications towards processes or products in mind...."[1]—then basic research is less likely to generate patentable inventions because the output of such research is less likely to pass the utility hurdle required for patentability. Therefore, to the degree that firms use knowledge flows from public research in their own basic research, we are less likely to observe a patent, and in turn, citations. This would understate the contribution of public research to a firm's innovation. A second possible class of measurement error in patent citations, termed "errors of commission," occurs when firms' citations to public research reflect something other than knowledge flows. This may occur when firms cite public research for reasons having to do with their patenting and citing strategies that are not directly indicative of knowledge flows. For example, firms concerned with the validity of their patents may include more citations of all sorts in an attempt to make their patents less vulnerable to validity challenges in the courts (Allison and Lemley 1998).

To look for either source of measurement error, we first assume that knowledge flows can be expressed as a function of various dimensions, some of which are reflected by citations and some of which are not. Accordingly, we assume that knowledge flows are a linear additive function of two sets of variables such that

$$k = \boldsymbol{\beta}_1 \mathbf{X}_1 + \boldsymbol{\beta}_2 \mathbf{X}_2, \qquad (1)$$

where k is true knowledge flows, $\mathbf{X}_1$ includes dimensions of knowledge flows reflected by patent citations, and $\mathbf{X}_2$ includes dimensions of knowledge flows not reflected by patent citations. Thus, we can express patent citations as a measure of knowledge flows, denoted as k_c, as a function of $\mathbf{X}_1$ but not of $\mathbf{X}_2$. Furthermore, we assume that there is a set of variables, $\mathbf{P}$, that, although *not* correlated with true knowledge flows, is correlated with patent citations and thus contains information on the measure k_c above and beyond that shared with true knowledge flows k. Thus, we express k_c as follows:

$$k_c = \boldsymbol{\alpha}_1 \mathbf{X}_1 + \boldsymbol{\gamma}_c \mathbf{P} + v_c, \qquad (2)$$

where v_c is an error term.

To express k_c as a function of true knowledge flows and the sources of measurement error, we subtract

Equation (2) from (1) and assume for simplicity that β_1 is equal to α_1, to yield

$$k_c = k + \mu_c, \qquad (3)$$

where μ_c is

$$\mu_c = \boldsymbol{\gamma}_c \mathbf{P} - \boldsymbol{\beta}_2 \mathbf{X}_2 + v_c. \qquad (4)$$

Although many studies that use patent citations are careful to recognize the limitations associated with these measures, the pervasive assumption is that the associated measurement error is "classical" (Jaffe et al. 1993, 1998, 2002), meaning that it is uncorrelated with the true measure k such that $\mathrm{Cov}(\mu_c, k) = 0$ and $\mathrm{E}(\mu_c) = 0$. Such classical measurement error in a dependent variable will neither bias nor reduce the consistency of the coefficient estimates; it will simply reduce their efficiency. When classical measurement error characterizes an independent variable, however, it may result in an attenuated, or downward-biased, and inconsistent coefficient estimate. When either of these assumptions is violated (i.e., either μ_c is correlated with k or $\mathrm{E}(\mu_c) \neq 0$), then "nonclassical" measurement error may lead to biased and inconsistent estimates (Bound et al. 2001, Carroll et al. 2006).

Equations (3) and (4) allow us to consider the nature of the measurement error in k_c. When using patent citations as a measure of knowledge flows, any unobserved elements of $\mathbf{X}_2$ in the composite error term μ_c will be related to k per Equation (1), violating the requirement for classical measurement error that $\mathrm{Cov}(\mu_c, k) = 0$. Furthermore, if k_c is correlated with $\mathbf{P}$, we would expect $\boldsymbol{\gamma}_c \mathbf{P} \neq 0$, thereby violating the second assumption of classical measurement error that $\mathrm{E}(\mu_c) = 0$. As a consequence, patent citations contain two possible sources of nonclassical measurement error that may impact the accuracy of estimates of knowledge flows and their impact on dependent variables, and bias coefficient estimates of other independent variables of interest that are themselves correlated with $\mathbf{X}_2$ or $\mathbf{P}$.

Equations (2)–(4) suggest that, to identify sources of measurement error, we need to determine whether there are correlates of knowledge flows that patent citations reflect (i.e., $\mathbf{X}_1$), correlates of knowledge flows that patent citations do not reflect (i.e., $\mathbf{X}_2$), and correlates of patent citations that do not reflect knowledge flows ($\mathbf{P}$). To address these questions, and in turn determine whether k_c is subject to the sources of nonclassical measurement error, we estimate an empirical specification that regresses k_c against observable elements of $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{P}$, as follows:

$$k_c = \boldsymbol{\alpha}_1 \mathbf{X}_1 + \boldsymbol{\alpha}_2 \mathbf{X}_2 + \boldsymbol{\gamma}_c \mathbf{P} + \varepsilon_c, \qquad (5)$$

where ε is a random disturbance term. As implied by the absence of $\mathbf{X}_2$ in Equation (2), we expect

---

[1] U.S. Federal Government, Office of Management and Budget Circular A-11, available at http://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/s84.pdf (accessed October 16, 2012).

that observable elements of $\mathbf{X}_2$ are uncorrelated with k_c and the coefficient vector $\boldsymbol{\alpha}_2$ to be insignificant. In contrast, we assume that patent citations are correlated with elements of $\mathbf{P}$, and thus $\boldsymbol{\gamma}_c$ to be significant.

We rely principally upon the Carnegie Mellon Survey (CMS) to provide measures for elements of $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{P}$. Later, we offer arguments grounded in the literature suggesting which measurable correlates of knowledge flows may be reflected by citations, which correlates of knowledge flows may not be reflected by citations, and which correlates should not be indicative of knowledge flows but are reflected by patent citations. We group these correlates into four sets of variables that reflect, respectively, the channels of knowledge flows, the ways in which firms use public research, the composition of a firm's R&D activity, and a firm's appropriability and citing strategies.

The CMS also offers an opportunity to validate our priors about what factors may be correlated with knowledge flows (or not) by providing an alternative, survey-based measure of knowledge flows from public research to firms. Accordingly, in parallel with our estimation of Equation (5), we regress the identical right-hand-side variables from Equation (5) against a survey measure of knowledge flows from public research, k_s, where the subscript s denotes the survey as our data source:

$$k_s = \boldsymbol{\theta}_1 \mathbf{X}_1 + \boldsymbol{\theta}_2 \mathbf{X}_2 + \boldsymbol{\gamma}_s \mathbf{P} + \varepsilon_s. \tag{6}$$

Our initial goal is to identify (1) variables in $\mathbf{X}_1$ that are significantly associated with both measures of knowledge flows, k_c and k_s; and (2) variables in $\mathbf{X}_2$ that are significantly associated with the survey measure k_s but not with citations k_c. We argue that variables that fall into the latter category are plausible sources of errors of omission. We also attempt to identify variables reflecting elements of $\mathbf{P}$ that are correlated with k_c but are not informative of knowledge flows from public research nor, we conjecture, of k_s. We suggest that such variables are plausible sources of errors of commission. Finally, our analysis assumes that although both the citation and survey measures almost certainly suffer from their own unique sources of measurement error, they are likely to do so in ways that are independent of one another. Thus, we anticipate that a comparison of the estimated coefficients for Equations (5) and (6) will shed light on possible sources of measurement error in patent citations.

3. Data

We employ a novel data set that combines survey and patent data at the level of a firm's R&D lab. The data set begins with the CMS, which contains survey responses from R&D managers regarding their labs'

R&D and patenting activities. We matched patents to the CMS at the R&D lab level, rather than at the firm level (as is common practice), to ensure greater precision in our data. To do this, we began with patents from the NBER patent database (Hall et al. 2001) applied for between 1991 and 1993. These years correspond to the time frame in the CMS, which was administered in 1994 and asked respondents about their respective labs' R&D activities throughout the previous three years. We paired patents to each CMS lab by matching patent assignee names to the company names drawn from the CMS. A challenge to matching patents to specific labs is posed by the fact that firms may have more than one lab, and that the R&D labs in the CMS may not be collocated with the assignee address provided on the patent. To more accurately match patents to their source R&D labs, we used the lab address provided by the CMS and the inventors' residential addresses provided on a given patent to calculate the geographic proximity of each inventor to the corresponding CMS lab linked to the patent's assignee. We matched a patent to a specific lab if the patent listed at least one inventor living within 35 miles of the CMS lab. For the matched patents, 88% of all listed inventors resided within this radius.

In our consideration of patent citations below, we distinguish between citations to patent references versus citations to nonpatent references, where the latter include scientific publications, conference proceedings, and other published documents. Obtaining information on the institutional sources of cited patents is straightforward given that each patent reference lists the assignee name, which can easily be classified by source type. Obtaining comparable information for nonpatent references, however, requires examination of the original source document to identify the institutional affiliation of the authors. To access nonpatent references, which are not included in the NBER patent data, we first retrieved the full record for each patent from Delphion, a commercial patent database provided by Thomson Scientific. To extract the institutional affiliation of the authors of the nonpatent references, we first developed software to extract, for each reference, the author name(s), document title, and journal name when the reference is a published article. Next, to identify the institutional affiliations of the authors of journal articles covered in the *Science Citation Index* (SCI) database, we manually retrieved the full record for each publication, including the list of authors and their affiliations. Nonpatent references not covered in SCI were examined for information on author affiliation and coded where possible.

We confined our observations to patenting labs to allow for the most direct comparison possible

between the two measures of knowledge flows.[2] Our final data set of 676 matched labs provides a comprehensive set of patent and nonpatent references linked to public research institutions. The number of patents granted to these labs annually ranged from 1 to 700, with an average of 10.9 patents per lab. The mean annual R&D lab budget was $22 million and ranged from as little as $30,000 to more than $1 billion.

3.1. Measures of Knowledge Flows

As noted earlier, our data set includes two different and independent measures of knowledge flows from public research: (1) backward citations to patent and nonpatent references, and (2) survey reports from R&D lab managers on the use of public research in their R&D projects.

3.1.1. Citation-Based Measures of Knowledge Flows. As previously noted, our patent-based data distinguish between citations to patent references and citations to nonpatent references. The latter are the most widely employed measure of knowledge flows largely because of the ease with which these data may be obtained. To reflect knowledge flows from public research, however, citations to nonpatent references are arguably better suited, because publications, conference proceedings, and the like are the primary form of public research output (Narin et al. 1997, Agrawal and Henderson 2002). Reinforcing the claim that nonpatent references are a better measure than patent references, recent research has also found that more than 40% of references to patents are introduced by patent examiners (Alcacer and Gittelman 2006, Alcacer et al. 2009), while less than 10% of nonpatent references are inserted by examiners (Lemley and Sampat 2012).

We designated patent references as citations to public research where the patent assignee was listed as a U.S. university, government lab, research institute, or hospital. We designated nonpatent references as citations to public research where at least one author of the cited work was affiliated with a U.S. university, government lab, research institute, or hospital.[3] For our matched data, almost 80% of all citations to

public research were made to nonpatent references, including scientific journal articles (49.8%); conference proceedings, working papers, and reports (8.0%); and edited academic volumes or textbooks (21.5%). For patent references and nonpatent references taken together, the majority of citations to public research were to universities (83%).

3.1.2. Survey-Based Measure of Knowledge Flows. Our survey-based measure of knowledge flows reflects an R&D manager's estimate of the fraction of their R&D unit's projects that used public research, reported on a five-point categorical scale (less than 10%, 10%–40%, 41%–60%, 61%–90%, or greater than 90%). Because this measure reflects the views of R&D lab managers, it ideally conveys an informed understanding of the lab's sources of knowledge as well as the full array of information channels and uses. In addition, external parties such as patent attorneys or examiners do not influence the survey measure.

Although offering advantages, this survey measure also has limitations. First, the fact that our response is a five-point scale rather than a continuous measure of knowledge flows reduces the precision of the measure. Second, unlike some patent-based measures, particularly citation counts, our survey measure does not provide information on the intensity of use of public research in each project, but rather its breadth of use across projects.[4] Third, respondents may misreport the use of public research because of, for example, inaccurate recall, lack of familiarity with actual knowledge flows, or a social desirability bias. Thus, we do not claim that our survey provides a more accurate measure of knowledge flows than citations. Rather, it provides an alternative measure that, when juxtaposed to our analyses with patent citations, should advance our understanding of patent citations as a measure of knowledge flows.

3.2. Comparison of Industry-Level Measures

Table 1 compares the industry-average reported fraction of R&D projects that use public research to the industry average fraction of patents that cite at least one public research reference. Figure 1 graphs this relationship. For purposes of this comparison, we used the midpoint from each of the survey response categories. In aggregate, the average share of R&D projects that reported using public research (20.2%) is lower than the share of patents citing public research (30.4%), and this is true for most industries. Consistent with prior research (Narin et al. 1997), biotechnology exhibits the highest use of public research, followed by pharmaceuticals, medical

---

[2] For the full sample of 1,246 R&D labs, 51.7% of labs reported using public research in their R&D projects. Of these labs, approximately 23.6% did not patent and thus did not cite public research. Yet there was no difference in the reported use of public research between those firms that patented (36.3%) and those that did not (36.9%). Furthermore, in a probit analysis reported in the appendix, we find that there was no significant relationship between the use of public research and the likelihood of a firm patenting, suggesting that sample selection bias associated with patenting versus nonpatenting labs is unlikely. Nonetheless, confining our sample to patenting labs paints patent citations in a more favorable light.

[3] The observed frequency of citations to papers coauthored with industrial scientists was less than 5%.

[4] As reported in §4.3, we tested the robustness of our analysis to this potential limitation and found qualitatively identical results.

**Table 1   Mean Comparison of Measures of Knowledge Flows**

Industry	Obs.	% of R&D projects that use public research	% of patents that cite public research
Biotechnology	18	51.1	86.9
Pharmaceuticals	24	30.2	58.7
Food	34	27.5	38.1
Aerospace	31	26.0	47.2
Computers	23	24.8	30.1
Seminconductors	25	24.2	31.1
Search/navigation equipment	19	23.7	42.4
Telecommunications	22	23.0	27.7
Medical devices	57	22.7	37.4
Miscellaneous chemicals	15	22.7	27.3
Chemicals	46	20.9	33.9
Concrete, cement, glass, etc.	14	19.6	28.7
Plastics, resins, etc.	15	19.0	20.4
Agriculture, mining, etc.	37	18.2	17.8
Precision instruments	22	17.5	37.7
Metal, steel, etc.	13	17.3	25.6
Rubber, plastic, etc.	17	16.5	16.9
Basic chemicals	28	16.3	32.1
Special purpose machinery	53	16.0	16.1
General purpose machinery	46	14.7	16.1
Metal products	22	13.9	26.0
Automobiles	23	12.8	9.3
General manufacturing	45	12.6	18.7
Electrical equipment	28	10.5	32.7
Total	677	20.2	30.4

Note. Reported are the survey responses of the average percentage of R&D projects that use public research and the average percentage of patents that cite at least one reference to public research.

devices, semiconductors, and computers. Given that the survey-based measure reflects midpoints of categorical ranges, we cannot claim that the absolute differences are meaningful. Of greater interest are the correlations between these two different types of measures.

Table 2 presents correlations between industry averages of the survey measure on the one hand, and four citation-based measures on the other. The different industry-level citation-based measures are highly correlated with the survey-reported use of public research, ranging from 0.87 for the industry average percentage of patents that cite public research (shown in Figure 1) to 0.51 for the industry average number of patent references. At the firm level, however, the correlations with the survey measure drop dramatically (ranging from 0.23 to 0.14), and further still in partial correlations that control for industry effects, ranging from a high of 0.16 for the number of nonpatent references to a low of 0.12 for the number of patent references. Although these correlations suggest that the survey and citation measures likely reflect a common underlying latent variable of knowledge flows, the relatively weak relationships when controlling for industry effects suggest that the two types

Figure 1   Mean Comparison of Citation and Survey Measures by Industry
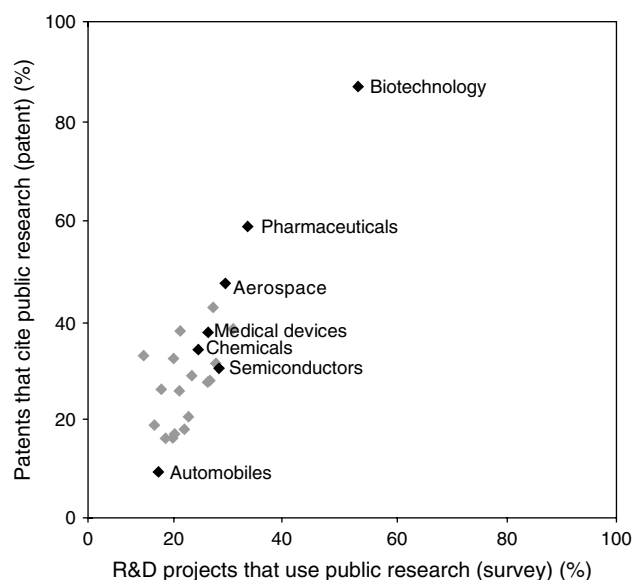
of measures differ considerably across firms within industries. We explore possible sources of these differences in the regression analyses that follow.

4.   Analysis of Measures of Knowledge Flows

We analyze sources of measurement error by identifying dimensions of knowledge flows that patent citations reflect (i.e., elements of X_1), dimensions that they miss (i.e., elements of X_2), and other extraneous factors that they reflect but should not (i.e., elements of P). We begin this analysis by estimating Equations (5) and (6) for four sets, or "blocks," of right-hand-side (RHS) variables. The first three of these blocks correspond to the following correlates of knowledge flows: (1) channels of knowledge flows, (2) uses of public research, and (3) the composition of firm R&D activities. The data for these correlates of knowledge flows are drawn from the CMS. The fourth block focuses on features of firms' appropriability and citing strategies, but not features of knowledge flows. For this latter set of correlates, we rely upon both patent data and the CMS.

We use ordered logit regression to estimate equations where the survey measure is the dependent variable.[5] For the patent citation equations, we measured knowledge flows as the number of citations to patent references and nonpatent references, respectively. We used the number of citations rather than the

---

[5] In a robustness test discussed in §4.3, we also employed fractional logistic estimation where the survey measure is converted to a share measure (i.e., bound between 0% and 100% employing category midpoints) with identical results.

Table 2   **Correlations with the Survey Measure of Knowledge Flows**

	Obs.	(1) % of patents that cite	(2) No. of total references	(3) No. of patent references	(4) No. of nonpatent references
Industry-level correlation[a]	24	0.87*	0.71*	0.51*	0.77*
Firm-level correlation	677	0.23*	0.20*	0.14*	0.21*
Firm-level partial correlation[b]	677	0.13*	0.15*	0.12*	0.16*

Note. Correlations are between the variables listed in the header and the survey measure (percentage of R&D projects that use public research).

[a]Industry averages; 24 International Standard Industrial Classification (ISIC) industries represented.

[b]Partial correlations controlling for 24 ISIC industries.

share of citations for comparability to the literature, which predominately uses citation counts as dependent or independent variables (see, e.g., Henderson et al. 1998, Branstetter and Ogura 2005, Duguet and MacGarvie 2005). Given the presence of overdispersion in the count measure, and that the mean number of patent citations to public research is small relative to the maximum, we employ negative binomial regression because it assigns less weight to larger values when adjusting for overdispersion relative to Poisson quasi-maximum likelihood (QML).[6] For all four block analyses and our subsequent analyses, we control for industry fixed effects, as well as for each firm's R&D and patenting activities. The construction of each variable is described below.

Although one might be concerned with potential endogeneity for a number of the right-hand-side variables in our block analyses, recall that our exercise is diagnostic. As such, we are careful to interpret our results as suggesting associations between variables and do not attempt to infer causality. An additional concern when both the RHS and left-hand-side (LHS) variables are drawn or constructed from the same source (i.e., the CMS) is common methods bias, which could conceivably magnify the correlations across our survey variables. First, as shown below, there are numerous survey variables that are not correlated with one another; they are often intended to measure different phenomena, use different response scales, and, whereas some are based on Likert scales, others

---

[6] Poisson QML assumes that the variance is proportional to the mean, thereby giving greater weight to observations with higher counts. For citations to both patent references and nonpatent references, approximately 85% of firms are below the mean, and thus Poisson QML would overweight firms that make a greater number of citations to public research. We also estimated models using quasi-Poisson maximum likelihood. In these results, patent citations appeared to be weaker indicators of knowledge flows than in the results featured here.

report behaviors. Also, in most cases, the survey questions related to the featured variables are separated on the questionnaire by unrelated questions, which reduces priming effects and further mitigates possible spurious correlations between variables. Finally, our analyses below demonstrate that there are numerous variables drawn from the CMS that are not related to our survey measure of the use of public research and are not related to one another.

We now turn to our regression analyses that estimate the relationship between the variables comprising the different blocks of correlates and, respectively, the survey- and citation-based measures of knowledge flows. Table 3 reports descriptive statistics and correlations for the featured variables. Table 4 provides a list of variables and their corresponding measures. Columns (1)–(9) of Table 5 provide the regression results organized by the three blocks of correlates of knowledge flows. Columns (10)–(12) provide results for the fourth block of measures that are thought to correlate with citations but not be indicative of knowledge flows. Columns (13)–(15) present all correlates together on the RHS. For each block, there are three columns of results for an identical set of predictor variables. Each of the three columns corresponds to one of the three LHS measures: the survey measure (survey), citations to patent references (PR), and citations to nonpatent references (NPR).

4.1. Searching for Errors of Omission: Correlates of Knowledge Flows

4.1.1. Channels of Knowledge Flows. A number of channels of knowledge flows from public research to industry have been considered in previous studies, including publications, public meetings, consulting, and collaborative research with university scientists (e.g., Cockburn and Henderson 1998, Cohen et al. 2002). The channels of "open science" have attracted particular attention (Hicks 1995, Sorenson and Fleming 2004). These include the traditional means of dissemination of academic research, notably publications and conferences. Because the primary medium of open science is a codified (i.e., citable) document such as a journal article, we would expect that a firm's reliance upon open science should be reflected in both the survey's reports of use of public research in R&D projects and in patent citations to public research.

In contrast to open science, firms may also rely upon private, often contract-based interactions between public research scientists and industrial R&D personnel as channels of knowledge flows. These private interactions, such as cooperative research ventures, consulting, or contract R&D, are arguably more effective for transferring more complex, less codified knowledge and know-how (Cockburn and

Table 3   Descriptive Statistics and Correlation Matrix

	Mean	Std. dev.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) *Survey* (% of R&D projects)	0.20	0.23															
(2) *Patent references* (count)	2.44	9.60	0.14*														
(3) *Nonpatent references* (count)	6.79	27.64	0.21*	0.76*													
(4) *Open science* (factor score)	0.04	0.86	0.49*	0.12*	0.15*												
(5) *Private interactions* (factor score)	0.09	0.82	0.34*	0.05	0.04	0.22*											
(6) *Industrial scientists* (percentage)	0.16	0.17	0.20*	0.05	0.09*	0.17*	0.16*										
(7) *Suggesting new R&D projects* (0/1)	0.37	0.48	0.37*	0.14*	0.17*	0.30*	0.36*	0.20*									
(8) *Completion of existing R&D projects* (0/1)	0.42	0.49	0.30*	0.17*	0.15*	0.31*	0.42*	0.11*	0.43*								
(9) *Basic research* ($mil)	2.05	15.77	0.13*	0.35*	0.36*	0.10*	0.12*	0.06	0.14*	0.13*							
(10) *Applied research* ($mil)	7.70	31.84	0.12*	0.38*	0.38*	0.08*	0.18*	0.05	0.15*	0.16*	0.60*						
(11) *Development* ($mil)	12.79	55.37	0.03	0.10*	0.06	−0.02	0.08*	−0.01	0.07	0.08*	0.09*	0.51*					
(12) *Patent effectiveness* (percentage)	0.39	0.31	0.14*	0.05	0.12*	0.11*	0.14*	0.07	0.07	0.01	0.13*	0.12*	0.01				
(13) *Secrecy* (percentage)	0.50	0.31	0.09*	−0.04	−0.02	0.12*	0.11*	0.06	0.04	0.04	0.07	0.02	−0.01	0.18*			
(14) *Citing propensity*	10.84	9.62	0.09*	0.01	0.10*	0.08*	0.06	0.06	0.10*	0.05	0.06	0.02	−0.02	−0.02	0.03		
(15) *R&D* ($mil)	22.13	87.08	0.12*	0.46*	0.36*	0.04	0.14*	0.02	0.09*	0.13*	0.61*	0.88*	0.93*	0.10*	0.05	−0.02	
(16) *Firm patents* (count)	10.86	44.96	0.14*	0.85*	0.65*	0.10	0.06	0.07	0.13*	0.18*	0.48*	0.40*	0.46*	0.10*	−0.02	−0.05	0.52*

*$p < 1\%$.

Henderson 1998, Cohen et al. 1998, Zucker et al. 1998, Thursby and Thursby 2002). Although such interactions occasionally produce citable outputs, such as reports, these documents may not be publicly disclosed, and thus the duty to cite does not apply. Furthermore, the most important of these channels, consulting (Cohen et al. 2002, Thursby et al. 2009), typically involves face-to-face communication that is not reflected in citable sources. Thus, the knowledge conveyed via such private interactions may not be readily reflected in patent citations.

We used exploratory factor analysis of survey responses to construct factor scores as our measures of open science and private channels.[7] Both measures are based on a question that asks respondents to report on a four-point scale the importance to the firm's R&D of different channels of knowledge flows from public research institutions. We define open science to include publications, conferences, and informal communication, and "private interactions" to include consulting with faculty, contract research, and collaborative research with public research scientists.

[7] Composite measures were also constructed as the mean of the survey items for each channel with comparable results.

Loadings from the factor analysis support the two distinct constructs of open science and private relationships.[8]

In addition to the channels of open science and private interactions, the employment of academically trained science and engineering Ph.D.'s should also facilitate the flow of public research to the firm. Ph.D.'s are better equipped to understand frontier academic research and, given their training, are also more likely to look toward public research as a primary resource (Allen 1977). Thus, we expect labs with a higher fraction of academically trained personnel to cite more public research. We utilized the CMS to

[8] The factor loadings for open science are publications and reports (0.73), public conferences and meetings (0.80), and informal information exchange (0.70). The factor loadings for contract-based interactions are cooperative R&D with academic scientists (0.66), contract research with universities or research institutes (0.72), and consulting with university faculty (0.58). As an alternative assessment of the reliability of these measures, we also calculated the Cronbach's alpha coefficients where a value of 0.70 or higher is widely considered a reliable measure. The Cronbach's alpha coefficient for open science is 0.86, and the alpha for contract-based relationship is 0.80, suggesting a high degree of reliability that these measures reflect latent constructs of the two respective channels of knowledge flows.

Table 4   Variables and Measures

Name	Source	Measure
Knowledge flows from public research		
Use of public research in R&D projects (*Survey*)	Survey	Reported fraction of R&D projects that use public research findings, five-point scale (e.g., 0%–10%, 11%–40%, etc.) and recode to center values (0%, 25%, 50%, 75%, 95%)
Citations to patent references (*PR*)	NBER	Number of patent citations to patent reference where the assignee is a university, government lab, or nonprofit research institute
Citations to nonpatent references (*NPR*)	NBER, Delphion, SCI	Number of patent citations to nonpatent (e.g., scientific publications) references where at least one author is affiliated with a university, government lab, or nonprofit research institute
Channels of knowledge flows		
Open science	Survey	Factor score of the importance of publications, conferences, and informal communication as channels of knowledge from public research
Private interactions	Survey	Factor score of the importance of faculty consulting, contract research, and collaborative R&D as channels of knowledge from public research
Industrial scientists	Survey	Fraction of total R&D employees who are M.D. or Ph.D. scientists or engineers
Uses of public research		
Suggest new R&D projects	Survey	Dummy that equals 1 if public research was an important source of knowledge that suggested new projects, 0 otherwise
Completion of existing R&D projects	Survey	Dummy that equals 1 if public research was an important source of knowledge that contributed to the completion of a firm's existing projects, 0 otherwise
Composition of R&D activity		
Basic research	Survey	Log of the amount of R&D budget directed toward scientific research with no specific commercial objectives
Applied research	Survey	Log of the amount of R&D budget directed toward scientific or engineering research with specific commercial objectives
Development	Survey	Log of the amount of R&D budget directed toward technical activity translating research findings into products or processes
Patenting and citing behavior		
Patent effectiveness	Survey	Percentage of firm's product and process innovations for which patents were effective at providing a competitive advantage
Secrecy	Survey	Percentage of firm's product and process innovations for which secrecy was effective at providing a competitive advantage
Citing propensity	NBER	Average number of backward citations excluding citations to public research per patent; reflects firm's overall level of citing
Controls		
Firm patents	NBER	Log of the number of patents
Industry fixed effects	Survey	24 International Standard Industrial Classification dummy variables

construct our measure of industrial scientist employment, which is the reported fraction of R&D personnel who were Ph.D.'s or M.D.'s.

Columns (1)–(3) of Table 5 report estimates of the relationship between the importance of the different types of channels of knowledge flows and our three measures of knowledge flows. Column (1) shows that, as expected, the channels of both open science and private interactions are significantly associated with the survey measure of the reported fraction of R&D projects that use public research. In contrast, column (2) shows that citations to patent references are not significantly related with either channel. As shown in column (3), citations to nonpatent references are significantly associated with the

channels of open science, which is expected given that the principal media of open science—scientific publications—are readily citable. Nevertheless, we observe that the effect of open science is notably smaller for nonpatent references than for the survey measure; a one-standard-deviation increase in open science increases the percentage of R&D projects that use public research by 183% but increases the number of nonpatent references by only 17%. Neither citations to patent nor nonpatent references are significantly related with private interactions. These latter results suggest that patent citations—including nonpatent references—may underestimate the contribution of public research that flows through the typically less-codified channels of consulting, contract R&D, and

Table 5 "Block" Analyses of Measures of Knowledge Flows from Public Research

Dependent variable: Regression model:	(1) Survey Ordered logit	(2) PR Negative binomial	(3) NPR Negative binomial	(4) Survey Ordered logit	(5) PR Negative binomial	(6) NPR Negative binomial	(7) Survey Ordered logit	(8) PR Negative binomial	(9) NPR Negative binomial	(10) Survey Ordered logit	(11) PR Negative binomial	(12) NPR Negative binomial	(13) Survey Ordered logit	(14) PR Negative binomial	(15) NPR Negative binomial
Channels of knowledge flows															
Open science	1.22*** [0.11]	0.06 [0.05]	0.19** [0.08]										1.13*** [0.12]	0.01 [0.05]	0.17** [0.07]
Private interactions	0.67*** [0.10]	−0.06 [0.07]	0.00 [0.08]										0.53*** [0.11]	−0.11* [0.06]	−0.07 [0.07]
Industrial scientists	0.93*** [0.46]	0.82*** [0.27]	1.97*** [0.45]										0.65 [0.52]	0.73*** [0.22]	1.29*** [0.34]
Uses of public research															
Suggesting new R&D projects				1.06*** [0.19]	0.15 [0.10]	0.40** [0.14]							0.74*** [0.20]	0.14 [0.11]	0.13 [0.13]
Completion of existing R&D projects				0.74*** [0.17]	0.09 [0.10]	0.20 [0.13]							0.21 [0.20]	0.06 [0.10]	0.11 [0.13]
Firm composition of R&D activity															
ln(*Basic*)							0.03** [0.01]	−0.00 [0.01]	−0.01 [0.01]				0.01 [0.01]	−0.01 [0.01]	−0.01 [0.01]
ln(*Applied*)							0.07*** [0.03]	0.04*** [0.02]	0.07** [0.03]				−0.00 [0.02]	0.03* [0.01]	0.04** [0.02]
ln(*Development*)							0.05** [0.02]	0.03* [0.01]	0.01 [0.02]				0.03 [0.03]	0.04** [0.01]	0.01 [0.02]
Firm patenting behavior															
Patent effectiveness										0.50 [0.32]	−0.28 [0.18]	0.38** [0.18]	0.38 [0.32]	−0.29 [0.18]	0.34 [0.22]
Secrecy										0.26 [0.26]	−0.08 [0.16]	−0.42** [0.19]	0.03 [0.27]	−0.09 [0.15]	−0.41** [0.19]
Citing propensity										−0.00 [0.01]	0.03*** [0.00]	0.06*** [0.01]	−0.01 [0.01]	0.03*** [0.00]	0.05*** [0.01]
Control variables															
Industry fixed effects (23)	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
ln(*R&D*)	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
ln(*Patent*)	0.00 [0.08]	0.91*** [0.04]	1.07** [0.06]	−0.03 [0.06]	0.90*** [0.04]	1.04*** [0.06]	0.10 [0.08]	0.88*** [0.03]	1.08*** [0.05]	0.08 [0.07]	0.96*** [0.04]	1.15*** [0.05]	−0.06 [0.06]	0.91*** [0.03]	1.12*** [0.04]
Log-likelihood	−784.49	−821.46	−1,071.81	−845.23	−822.29	−1,078.52	−886.51	−821.34	−1,085.97	−887.12	−811.06	−1,043.18	−770.13	−797.69	−1,025.41
Observations	676	676	676	676	676	676	676	676	676	676	676	676	676	676	676

Note. The dependent variables are *Survey* (fraction of R&D projects that use public research), *PR* (number of citations to patent references), and *NPR* (number of citations to nonpatent references). Robust standard errors clustered on industry in brackets.

*** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

cooperative R&D. Of concern is that consulting is one of the most important channels through which public research flows to industrial R&D (Cohen et al. 2002).

Turning to the role of industrial scientists, we find a positive significant association with the survey measure of knowledge flows as well as with both citation-based measures. Given that scientists tend to rely upon publications and other scientific literature as a key source of knowledge, they would be more likely to cite scientific publications. The larger effect size for nonpatent references in column (3) is, however, notable. A one-standard-deviation increase in industrial Ph.D. scientists increases the survey measure by 16.9% and citations to patent references by 14.8%. In contrast, it is associated with a 39.4% increase in citations to nonpatent references. In addition to reflecting knowledge flows, the relationship with nonpatent references may reflect the academic socialization of Ph.D. scientists into the practice of generously citing the work of others (Merton 1957, Sorenson and Fleming 2004). If the larger number of citations reflect norms around citing in addition to actual knowledge flows, then the fraction of Ph.D.-level scientists in a lab could also account for an "error of commission," as considered in §4.4. Teasing out such a normative effect from Ph.D.'s greater reliance upon public research is, however, difficult.

4.1.2. Uses of Public Research in Firm R&D Projects. Public research may be used in two ways by firms: it can either suggest new R&D projects or contribute to the completion of existing R&D projects. One might interpret the former role as contributing to firms' technological opportunities and the latter as reflecting the role of public research institutions as repositories of scientific and engineering knowledge. One might expect that when public research suggests new projects, it is more likely to be cited in patents than if its contribution is confined more to helping firms execute existing projects. Consider, for example, a product patent. Although the patent will describe the novel features of the product, it is less likely to describe the methods employed in its invention. If these methods were informed by public research, then citations will not reflect that contribution. The CMS provides measures of these two uses of public research: a binary response variable that equals one if public research was an important source of ideas for new R&D projects in the prior three years, and a second binary response variable that equals one if public research contributed to the completion of existing R&D projects.

The significant, positive coefficients for these two variables shown in column (4) of Table 5 suggests that both of these contributions of public research significantly predict managers' reports of knowledge flows

to a firm's R&D projects. In contrast, as shown in column (5), citations to patent references are not significantly related with either use of public research. However, the results in column (6) show that citations to nonpatent references do appear to reflect the role of public research in stimulating new R&D projects; firms that use public research as a source of new ideas make 20.9% more citations to nonpatent references than firms that do not use public research in this way. At the same time, nonpatent references are not significantly associated with the use of public research in the completion of existing R&D projects. In sum, although these results suggest that patent citations to nonpatent references capture one important type of contribution of public research to industrial R&D, they also suggest that citations understate the important role of public research as a source of knowledge contributing to project completion.[9]

4.1.3. Composition of Firm R&D Activity. To the limited extent that a firm may conduct basic research, such research activity is likely to build more upon knowledge flows from public research than would the firm's downstream applied research and development activities (Rosenberg 1985, 1990). The results of basic research, however, are less likely to satisfy the patentability requirement of utility in most industries, and are thus less likely to be patented (Pavitt 1991, Jaffe et al. 1993, Rosenberg and Nelson 1994). As a result, public research used by firms in their basic research is less likely to be observed in patent citations. Applied R&D activities, on the other hand, are directed toward the creation of technological innovations that are more likely to be patentable. To measure the composition of firms' R&D activity, we included three measures of a firm's expenditures on, respectively, basic research, applied research, and development. Basic research is defined as scientific research with no specific commercial objectives; applied research is research activity directed toward specific commercial objectives; development is technical activity directed toward translating research findings into products or processes. These measures were constructed by multiplying the survey-reported total R&D budget (in dollars) for each lab by a survey response on the share of the lab's R&D activity directed toward basic research, applied research, and development, which accounted for 100% of the lab's R&D activities.

The results in column (7) of Table 5 show a positive and significant relationship between firms' basic and applied research expenditures and managers'

---

[9] The survey results reported in Cohen et al. (2002) suggest that the role of public research in the completion of existing R&D projects is at least as important as the role of public research in suggesting new R&D projects.

reported share of R&D projects that use public research. For citations to both patent references and nonpatent references, we observe a significant relationship with applied research, but no significant relationship with a firm's basic research, suggesting that, to the extent that firms conduct basic research, the contribution of public research to that activity may be missed by their patent citations.[10]

4.2. Searching for Errors of Commission: Correlates of Patenting and Citing

In this subsection, we consider possible sources of errors of commission tied to firms' appropriability and citing strategies. For example, firms patent a greater share of their innovations when they believe patents to be more effective in protecting their innovations (Arora et al. 2008), yielding more citations to all sources, including public research. In contrast, firms that rely more heavily on secrecy rather than patents to protect their innovations—perhaps to reduce disclosure that might enable competitors to imitate patented inventions—may patent less, resulting in fewer citations. Concern over secrecy may also influence citations in the patents that firms do file if they seek to conceal information that might enable rivals to invent around a patent (Horstmann et al. 1985, Friedman et al. 1991). In either event, firms may cite public research (or not) for reasons that have little to do with actual knowledge flows. Our measures of firms' appropriability strategies are drawn from the CMS and reflect R&D managers' views of the percentage of product and process innovations for which they consider patents and secrecy, respectively, to be effective means of protection.

Firms' strategies for citing prior art may also distort the accuracy of citations as a measure of knowledge flows. For example, firms concerned with the risk of litigation or wishing to strengthen their patents may cite more prior art to diminish the threat of invalidity countersuits (Allison and Lemley 1998, Harhoff et al. 1999). In contrast, firms wishing to amass patents for defensive or cross-licensing purposes may make fewer prior art references to maximize the likelihood of issuance (Jaffe et al. 1993, Lampe 2012).

Thus, a firm's strategic citing behavior may introduce variation in its citations that is unrelated to knowledge flows, potentially constituting yet another source of nonclassical measurement error. We measured a firm's propensity to cite prior art as a lab's average number of backward citations per patent for the sample period, 1991–1993, excluding citations to public research.[11]

The results in column (10) of Table 5 show that neither patent effectiveness, secrecy, nor citing propensity exhibit a relationship with the survey-reported use of public research, which is consistent with the survey measure not being influenced by factors that are not directly related to actual knowledge flows from public research. In contrast, citations to both patent and nonpatent references have significant positive relationships with citing propensity. Patent effectiveness and secrecy are both significantly associated with nonpatent references. To provide a sense of the magnitude of these possible errors of commission, a one-standard-deviation increase in patent effectiveness increases citations to nonpatent references by 12.4%; a one-standard-deviation increase in secrecy is associated with a decline in nonpatent references of 12.2%; and a one-standard-deviation increase in citing propensity is associated with an increase in citations to nonpatent references of 70.5%. Thus, it appears that firms' appropriability and citing strategies may influence citations to public research, suggesting a source of variation in citations that does not directly reflect knowledge flows, and thus may constitute a source of errors of commission.

4.3. Robustness Tests

We conducted a number of analyses to examine the robustness of our results for the block analyses. In these analyses, rather than estimating each block separately, we estimated all four blocks together, as reported in columns (13)–(15) of Table 5. Given collinearity across selected RHS variables (see Table 3), the strength of the relationships between these variables and the three measures of knowledge flows diminish, as expected. Still, some overarching results merit mention. First, citations to patent references continue to perform poorly relative to citations to nonpatent references. Consider in particular the significant relationship with the channels of open science, that suggests nonpatent references are a better measure of knowledge flows from public research than patent references.

We further examined the robustness of our results by employing alternative measures and estimation methods. To create survey and citation measures that

---

[10] In a corollary analysis designed to further consider the premise of our argument that the output of firms' basic research will tend not to be patented, we directly examined whether firms' basic research expenditures are reflected in patent counts, and, as an alternative, in firms' scientific publications. We found that a firm's patent counts exhibited no significant relationship with its basic research activity, and scientific publications exhibit a strong, significant positive relationship. This finding suggests that patent citations not only obscure the role of public research knowledge flows in informing firms' basic research, but that patents themselves are a poor indicator of the output of a firm's basic research activity. This is important in light of the occasional use of firm patent citations to public research as a proxy for firms' basic research in prior studies.

[11] We also performed this analysis including backward citations to all sources with qualitatively identical results.

are more directly comparable, we first recoded the categorical survey response of the percentage of R&D projects that use public research to the midpoint of each category to create a percentage-based measure. We then replaced the number of citations with the share of patents that cite at least one patent or nonpatent reference, respectively. We estimated all models using fractional logistic regression and found nearly identical qualitative results.

In another robustness test, we recognize that, unlike citation count measures, our survey measure does not provide information on the intensity of the use of public research in each project, but rather the breadth of use across projects. To consider this potential limitation, we reproduced our featured results using a survey measure of the frequency with which a firm's R&D personnel obtain useful information from public research, as reported on a five-point scale (rarely or never, semi-yearly, monthly, weekly, daily). To the degree that this measure better reflects intensity of use, it should provide a robustness test of our featured survey measure of the share of R&D projects that use public research. Results are qualitatively identical to our featured measure reflecting breadth of use.

We further tested the robustness of the patent citation results to different controls and levels of patenting activity. First, we replaced our control for a firm's overall level of patenting with a control for the total number of backward citations (i.e., to both public research as well as to firms and other sources), with qualitatively identical results. Second, we examined whether firms with more patents—and thus more observable citations—differ from firms with fewer patents. We did this by first restricting the sample to firms with a patent count equal to or greater than the mean (10.86) and then to firms with a patent count less than the mean. Results between the two samples were nearly identical.

In light of the Alcacer et al. (2009) finding that 40% of citations to patents were added by examiners, as a final robustness test we considered whether citations added by patent examiners may explain the weak results for patent references. The challenge for this exercise is that examiner-added citations were not publicly identifiable by the USPTO until 2001, eight years after our sample period. Proceeding on the assumption that the relationship between public research and industrial R&D had not changed appreciably between 1993 and 2001, we matched the CMS labs included in this study to patent data that identified examiner-added citations for patents filed in 2001.[12] Because many of the labs included

in our featured analysis did not patent in 2001, the matched sample dropped to 351 labs. We first replicated our block analyses using all patent references—both those made by the firm and those introduced by examiners—to benchmark against our featured analyses in Table 5, with nearly identical results. We then removed examiner-added citations to assess whether the results for patent references improved. Surprisingly, they did not. Finally, we constructed a measure of the share of examiner-added citations (per Alcacer et al. 2009) and included it as a control variable in our primary analysis, with no change in the results. Thus, at least for citations to public research as reflected in these data, we find no evidence of patent-examiner influence on the sources of systematic error that we investigate.

4.4. Isolating Sources of Measurement Error

To summarize the results from our block analyses, patent citations—especially nonpatent references—appear to reflect some dimensions of knowledge flows shared with the survey measure (e.g., open science), which we interpret as elements of X_1 in Equation (5). Our block analyses also suggest that other dimensions of knowledge flows, corresponding to X_2 in Equation (5), are associated with the survey measure but not observed in patent citations (e.g., a firm's basic research activity). Finally, our analyses suggest that firms' appropriability and citing strategies—corresponding to P in Equation (5)—affect citations but are not related to the survey measure, and may not be indicative of knowledge flows. Yet, three questions remain. First, even if patent citations mismeasure knowledge flows, are they still informative of knowledge flows from public research? Second, once we control for the common sources of variation between the two types of measures of knowledge flows, are there still systematic sources of systematic measurement error? Finally, for those dimensions of knowledge flows that are related to both the survey and citation measures, do the two types of measures reflect these dimensions comparably?

To consider these questions, we first regress one measure of knowledge flows onto the other to estimate the shared variation, which should reflect the common dimensions of true knowledge flows captured by both measures. These estimates should provide a sense of the strength of the common component of variation. We then also include in a second specification our measures for elements of X_1, X_2, and P from Equations (5) and (6). After accounting for the shared variation between the two measures of knowledge flows, any remaining significant coefficients should signal possible sources of error.[13]

---

[12] We thank Bhaven Sampat for suggesting this approach and for graciously providing access to patent examiner data.

[13] We also applied this approach to each block considered separately with similar results to the full specification discussed in the text.

Table 6   Isolating the Sources of Measurement Error

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Errors of omission				Errors of commission			
Dependent variable:	*Survey* %StdX	*Survey* %StdX	*Survey* %StdX	*Survey* %StdX	*PR* %StdX	*PR* %StdX	*NPR* %StdX	*NPR* %StdX
ln(*PR*)	34.8**	31.7**						
ln(*NPR*)			45.4**	36.7**				
Survey					9.1*	8.1*	35.2***	20.7***
Open science		163.2***		160.5***		−1.9		6.2
Private interactions		57.1***		57.4***		−9.9**		−8.6
Industrial scientists		9.4		9.5		12.2***		22.3***
Suggesting new R&D projects		42.2***		41.8***		4.9		3.3
Completion of Existing R&D projects		11.2		11.5		2.3		3.1
ln(*Basic*)		4.1		5.0		−2.7		−8.6*
ln(*Applied*)		−1.6		−1.7		14.2**		18.7**
ln(*Development*)		10.7		12.0		12.2**		3.6
Patent effectiveness		14.3		12.4		−9.6*		9.6
Secrecy		0.9		1.2		−2.5		−11.1**
Citing propensity		−11.6		−12.6		29.1***		61.4***

Notes. $N = 676$. Values reported are the percentage change in the dependent variable for a one-standard-deviation change in the independent variable; dependent variables are *Survey* (fraction of R&D projects that use public research), *PR* (number of citations to patent references), and *NPR* (number of citations to nonpatent references); ordered logistic regression estimates are reported for the survey measure and negative binomial regression estimates are reported for the count citation measures; robust standard errors clustered on industry.

*** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

Whether this reflects an error of omission or commission depends upon what we assume about the dependent variable in each instance. In addition to our maintained assumption that the survey measure reflects dimensions of true knowledge flows that are not fully shared with patent citations, we also assume that it is unrelated to any factors that are not indicative of knowledge flows. In this event, when regressing the survey measure against patent citations and the other independent variables considered, the significance of coefficient estimates for the variables other than patent citations should indicate an error of omission. Similarly, we will assume that patent citations are not indicative of any dimensions of knowledge flows beyond what is reflected in the shared variation with the survey measure. Thus, when regressing the citation measure against the survey measure and other independent variables, any significant coefficients of variables other than the survey measure should indicate an error of commission.

To facilitate comparisons across regression models, Table 6 reports the percentage change in the dependent variable for one-standard-deviation change in each coefficient estimate. The results in columns (1)–(8) demonstrate that, across specifications, both patent references and nonpatent references exhibit a significant relationship with the survey measure. For example, as shown in column (2), a one-standard-deviation increase in patent references is associated with a 31.7% increase in the survey measure, and in column (4) a one-standard-deviation increase in nonpatent references is associated with a 36.7% increase in the survey measure. The strong, shared variation between the survey- and citation-based measures of knowledge flows increases confidence that each measure reflects some common dimensions of true knowledge flows. Indeed, notwithstanding any bias that may characterize patent citations as a measure of knowledge flows from public research, the substantial shared variation suggests that citations—even to patent references—likely reflect some component of true knowledge flows.

Once we control for the shared variation, however, we still find evidence of sources of errors of omission in the citation measures. In columns (2) and (4) of Table 6, we see little change from the qualitative results shown for the full specification in column (13) of Table 5 for the other variables corresponding to $\mathbf{X}_2$, suggesting that, even after controlling for the shared variation, patent citations may understate the same dimensions of knowledge flows as highlighted earlier. The results for variables reflecting patenting and citing behaviors are also insignificant, as they were in the full specification in Table 5.[14] The results in column (4) of Table 6 also suggest that nonpatent

---

[14] We also performed regressions that combine patent and nonpatent references into a single measure with results that are similar to, but weaker than, those with only nonpatent references. Thus, combining citations to patent and nonpatent references appears to be an inferior measure relative to citations to nonpatent references alone.

references may understate the influence of open science as indicated by the positive significant coefficient even after controlling for the shared variation.

We now consider errors of commission by regressing both citation measures onto the survey measure. If the survey measure controls for the shared variation with patent citations and for those dimensions of knowledge flows corresponding to the survey but not the citation measures, we can expect that the dimensions associated with $\mathbf{X}_1$, such as open science, should no longer be significant. Nor should we expect any of the measures corresponding to $\mathbf{X}_2$ to be significant. In addition, as previously noted, we should expect any remaining significant variables to reflect possible sources of errors of commission corresponding to $\mathbf{P}$. Columns (5)–(8) in Table 6 show that, as expected, after controlling for the survey measure, the correlates that we believe to correspond to knowledge flows are no longer significant, with the exception of industrial scientists. We also note that the relationship with nonpatent references is notably greater than that with patent citations; a one-standard-deviation increase in the survey measure is associated with a 8.1% increase in patent references but a 20.7% increase in nonpatent references.[15]

The large and highly significant coefficient for industrial scientists, even after controlling for the survey measure of knowledge flows, is striking. This suggests that as a firm's R&D employees become more populated by science and engineering Ph.D.'s, we will observe more citations to public research beyond what we would expect based on a lab's reported use of public research alone. The sociology of science (Merton 1957, Sorenson and Fleming 2004) suggests an explanation: academically trained Ph.D.'s are socialized into the practice of generously citing the work of others and the larger share of citations to public research made by Ph.D.'s simply reflects those scientists' greater propensity to cite beyond the actual use of public research.[16] Alternatively, if we depart from our maintained assumptions and allow that patent citations may reflect dimensions of knowledge flows that the survey variable does not, then this finding may suggest that Ph.D.'s more accurately attribute the contribution of public research to a firm's R&D relative to other R&D employees.

As an additional analysis, to examine further the strength of the relationship between our two types of measures, we consider cross-industry differences in the shared variation between the citation and survey measures. To ensure adequate sample size, we aggregated observations to five industry groups: biomedical (pharmaceuticals, biotechnology, medical devices); chemicals; ICT (computers, semiconductors, telecommunications); machinery; and electrical equipment. Table 7 reports marginal effect estimates for simple specifications that regress both citation-based measures onto the survey measure by industry group. We observe that across industry groups, patent references demonstrate no significant relationship with the survey measure. We also see, however, that nonpatent references are significantly associated with the survey measure in biomedical, chemicals, and ICT. These industries are distinguished from other industries in that they rely more upon public research (per the survey measure), cite a greater number of nonpatent references per patent, and employ a greater share of science and engineering Ph.D.'s as R&D personnel. One way to interpret these findings is that both the survey measure and nonpatent references are more accurate indicators of the contribution of public research in industries where that contribution is greater.

4.5. Magnitude and Direction of Bias

Our comparison of survey and patent citation-based measures of knowledge flows suggests sources of systematic measurement error in patent citations that may lead to bias when citations are employed as measures of knowledge flows in regression analyses. We now examine the magnitude and direction of this possible bias when backward patent citations are used as an independent variable by estimating a simple, illustrative model of the impact of knowledge flows on R&D labs' innovative performance:

$$q_i = \alpha + \beta k_i + \delta r_i + v_i, \tag{7}$$

where q is firm i's innovative performance measured as forward citation-weighted patent counts, k is knowledge flows from public research, r is firm R&D measured in log form, and v is an error term.

Although prior studies have used backward citations as their measure of k in similar specifications, our analyses suggest that patent citations as a measure of k will not fully capture the impact of knowledge flows because of errors of omission. Consequently, this measure may bias the estimated effect of the knowledge flow variable, as well as introduce bias into estimates of other independent variables that are correlated with the measurement error. To assess whether such errors of omission might indeed result in biased coefficient estimates,

---

[15] Auxiliary analyses of seemingly unrelated regressions comparing the coefficients confirm that these differences are significant ($\chi^2 = 10.78^{***}$ comparing columns (5) and (7), and $\chi^2 = 5.05^{**}$ comparing columns (6) and (8) in Table 6), providing additional evidence that nonpatent references correspond more closely to the survey measure of knowledge flows than do patent references.

[16] The notion that Ph.D.'s provide substantially more academic articles as prior art than are relevant to the invention is supported by interviews of both inventors and patent attorneys conducted by one of the authors.

Table 7   Shared Variation by Industry Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Industry (avg. %R&D projects):	Biomedical (30%)		Chemicals (20%)		ICT (25%)		ElecEquip (13%)		Machinery (15%)	
% of cites to public research:	(18%)	(82%)	(29%)	(71%)	(40%)	(60%)	(65%)	(35%)	(47%)	(53%)
Dependent variable:	*PR*	*NPR*	*PR*	*NPR*	*PR*	*NPR*	*PR*	*NPR*	*PR*	*NPR*
%R&D projects (Survey)	0.45	2.20***	0.50	1.47**	0.45	1.30***	0.59	0.12	1.01	1.26
	(0.44)	(0.49)	(0.78)	(0.61)	(0.56)	(0.50)	(0.56)	(1.12)	(0.74)	(1.15)
ln(*Patent*)	0.91***	1.20***	1.01***	1.20***	0.90***	0.98***	0.74***	1.29***	1.03***	1.15***
	(0.07)	(0.08)	(0.07)	(0.09)	(0.07)	(0.07)	(0.12)	(0.22)	(0.08)	(0.12)
Industry fixed effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Constant	−0.71***	−1.38***	−2.17***	−1.85***	−1.36***	−1.66***	−0.93***	−1.20***	−1.50**	−3.92***
	(0.21)	(0.24)	(0.34)	(0.33)	(0.26)	(0.31)	(0.23)	(0.36)	(0.69)	(0.81)
Log-likelihood	−164.43	−229.46	−141.48	−219.24	−97.20	−126.58	−57.09	−56.30	−84.83	−100.09
Observations	99	99	104	104	62	62	58	58	107	107

Notes. Marginal effect estimates from negative binomial regressions reported; dependent variables are *PR* (number of citations to patent references) and *NPR* (number of citations to nonpatent references). The groups are biomedical (pharmaceuticals, biotechnology, and medical devices), chemicals, information and computer technology (computers, semiconductors, and telecommunications), machinery, and electrical equipment; robust standard errors clustered on industry in parentheses.

    ***$p < 1\%$; **$p < 5\%$; *$p < 10\%$.

we introduce a measure that reflects those elements of knowledge flows not captured in patent citations corresponding to $\mathbf{X}_2$, by assuming that our survey measure reflects at least some of these elements. To construct this measure, we regressed the survey measure onto patent citations and used the residual to approximate that component of knowledge flows not explained by patent citations, denoted as $\hat{\mathbf{X}}_2$.[17] We then included $\hat{\mathbf{X}}_2$ in Equation (7) along with the citation measure, which should reflect elements of $\mathbf{X}_1$. The coefficient estimate for $\hat{\mathbf{X}}_2$ should convey a sense of the magnitude of bias associated with those sources of errors of omission in the patent citation measure.[18]

As previously noted, another source of potential bias is errors of commission. If factors accounting for errors of commission are correlated with both our LHS citation measure of innovative performance and our RHS citation measure of knowledge flows, but not with true knowledge flows, they may also bias the estimate of β. A correlation between the sources of errors of commission and our measure of innovative performance may occur because, for example, firms that believe that patents are more effective at appropriating returns may both innovate and patent more because of stronger appropriability, thereby increasing both the number of backward citations and forward citations. To determine the impact of errors of commission on the estimation of the model, we examine the sensitivity of the estimated coefficients on citation measures of knowledge flows to the inclusion of the variables representing the sources of errors of commission considered above.

In this analysis, we limited our attention to nonpatent references, which, per our analyses above, appear to better reflect knowledge flows from public research than do patent references. To compute $\hat{\mathbf{X}}_2$, we converted our citation measure to the percentage of patents that cite nonpatent references to be comparable to our survey-based measure that is also expressed as a percentage. Column (1) in Table 8 presents the fractional logistic regression used to construct $\hat{\mathbf{X}}_2$. We then predicted the percentage of R&D projects that used public research and subtracted this from the observed measure to obtain the residual measure $\hat{\mathbf{X}}_2$, which reflects that component of knowledge flows reflected in the survey measure but not in citations.

We estimated the innovative performance models using Poisson quasi-maximum likelihood and report marginal effect estimates. The results in columns (2)–(7) provide specifications with the survey and citation measures entered separately and

---

[17] An alternative approach to correcting for measurement error bias is the use of an instrumental variable, which can be a second and independent measure of the latent variable of interest (Bound et al. 2001, Carroll et al. 2006). To be a valid instrument, a variable must be correlated with the mismeasured variable, but not correlated with either the measurement error or the dependent variable after partialling out the mismeasured variable. Although our survey measure satisfies the first criterion, sources of nonclassical measurement error in patent citations themselves—especially elements of X_2—make satisfying the second two criteria impossible. As a consequence, using the survey measure as an instrument for patent citations is invalid and may overestimate or underestimate measurement error bias (Bound et al. 2001).

[18] This exercise arguably provides a conservative test of the bias associated with the use of patent citations since some component of the relationship between forward citations on the LHS and backward citations on the RHS is likely because not only are these both patent-based measures, but these measures reflect citations to and by the same patents.

Table 8   **Estimating the Magnitude and Direction of Bias**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	*%Survey*	*CWPC*	*CWPC*	*CWPC*	*CWPC*	*CWPC*	*CWPC*	*CWPC*
Regression model:	FracLogit	PQML	PQML	PQML	PQML	PQML	PQML	PQML
%NPR (X_1)	0.10***	118.70***	122.11***			107.47***	119.73***	122.49***
	(0.02)	(42.71)	(42.82)			(38.04)	(40.18)	(39.26)
%Survey ($X_1 + X_2$)				156.85***	157.66***	139.52***		
				(47.95)	(49.88)	(46.01)		
%Survey_residual ($\hat{X}_2$)							140.53***	140.94***
							(61.45)	(45.79)
Patent effectiveness (P)			40.98		18.49	9.63		9.17
			(46.57)		(46.90)	(45.55)		(45.59)
Secrecy (P)			−93.76*		−87.27*	−85.03*		−84.85*
			(52.66)		(49.77)	(51.01)		(50.96)
Citing propensity (P)			−0.68		−0.88	−1.16		−1.16
			(1.39)		(1.20)	(1.19)		(1.19)
ln(*R&D*)		85.63***	84.82***	79.57***	79.73***	77.96***	77.06***	77.84***
		(18.46)	(17.93)	(14.35)	(14.41)	(15.33)	(19.25)	(15.30)
Industry fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
Observations	676	676	676	676	676	676	676	676

Notes. Marginal effect estimates reported. Column (1) regresses *%Survey* onto *%NPR* using fractional logistic regression to estimate $\hat{X}_2$; columns (2)–(6) dependent variable is *citation-weighted patent counts* (*CWPC*) using Poisson quasi-maximum likelihood (PQML); columns (1)–(6) robust standard errors in parentheses; columns (7) and (8) bootstrapped standard errors in parentheses.

   ***p < 1%; **p < 5%; *p < 10%.

together for comparison. The results in column (7) enable a comparison of the estimated effect of knowledge flows on innovative performance as reflected by citations to nonpatent references (%NPR) with that of the unobserved dimensions reflected by $\hat{X}_2$. We see that both %NPR and the computed residual component, $\hat{X}_2$, are positive and significant. Furthermore, the estimated marginal effects for both measures are roughly comparable, suggesting that they each reflect unique and important dimensions of the impact of public research. Although this is only an illustrative model, these results suggest that typically "unobserved" dimensions of knowledge flows have a positive, significant relationship with firms' innovative performance, and, in turn, that the use of patent citations alone as a measure may substantially underestimate the influence of public research on industrial R&D—perhaps by half.[19]

Regarding the impact of errors of commission, a comparison of the estimated coefficient for %NPR between either columns (2) and (3) or between columns (7) and (8) suggests that they result in little bias. Indeed, consistent with this result, we observe little independent effect of the sources of errors of commission on innovative performance; there is only a weak negative relationship with secrecy. Finally, we also observe little evidence that nonclassical measurement error in patent citations—whether because of errors of omission or commission—biases

estimates of other independent variables. After correcting for both errors of omission and errors of commission, the coefficient estimate for R&D—an additional predictor of innovative performance—changes little. Nevertheless, the potential for bias in this or other independent variables remains; the significance and magnitude of this bias likely depends upon the sample employed and the empirical specification.

Our illustrative analysis suggests, first, that errors of omission, not commission, appear to be of greater concern. Using citations as a measure of knowledge flows from public research may consequently account for a significant underestimation of the impact of public research on innovative performance. Second, patent citations appear to capture an important component of that impact, although the estimate of this effect may be inflated in this analysis because of the employment of citation-based measures for both the independent and dependent variable.

5. Discussion and Implications for Research

The strategy and innovation literatures regularly use patent citations to measure knowledge flows, yet we know relatively little about the validity of these measures. The research of Jaffe et al. (1998, 2002) and Duguet and MacGarvie (2005) suggests that, although such citations appear to reflect knowledge flows, they are "noisy." This paper considered whether patent citations may actually be subject to sources of systematic measurement error that compromise their role as measures of knowledge flows in analyses where

---

[19] We performed regressions with patent counts as the measure of innovative performance with comparable results.

such measures serve as either independent or dependent variables (see, e.g., Jaffe et al. 1993, Henderson et al. 1998, Almeida and Kogut 1999, Mowery et al. 2002, Rosenkopf and Almeida 2003). By matching managers' reports on the use and character of knowledge flows from public research with contemporaneous patent data for those managers' R&D labs, we explored the virtues and limitations associated with patent citations as indicators of knowledge flows from public research.

We searched for two sources of measurement error. First, we considered dimensions of knowledge flows from public research that citations fail to reflect, which we termed errors of omission. Second, we looked for sources of errors of commission—factors related to citations that are not indicative of knowledge flows. Consistent with our expectations, we found that citations, particularly to the nonpatent literature, appear to reflect the research output of public research that shows up in open, documented ways, but do not reflect the knowledge that is transmitted via other, typically more private channels, such as consulting or cooperative ventures. What citations also appear to miss is the knowledge that helps firms address the technical challenges they encounter in their R&D. Furthermore, citations do not appear to reflect the extent to which firms use university research in their own basic research.

We also found evidence for errors of commission. We observed a strong relationship between firms' patent citations to public research and firms' overall citing propensity, suggesting that some component of the variation across firms in their patent citations to public research may be driven not only by knowledge flows, but by firms' concern with the strength and validity of their patents. We also observed a strong negative relationship between both patent and nonpatent references to public research and the degree to which firms feature secrecy in their appropriability strategies. Similarly, though far from conclusive, our results suggest that employment of industrial Ph.D. scientists and engineers, although surely a correlate of true knowledge flows, may also account for an overstatement of such flows, perhaps reflecting conformity to academic norms that encourage attribution.

Sources of error notwithstanding, our exercise to isolate the sources of error showed significant shared variation between the survey measure and citations to both patent and nonpatent references. These findings suggest that, despite the sources of measurement error identified, patent citations likely reflect meaningful aspects of knowledge flows from public research. The result for industrial scientists in this exercise also raised the possibility that patent citations

may capture features of those flows that our survey measure fails to reflect.

Finally, our estimation of an illustrative model of firm innovative performance provides a sense of the magnitude and direction of bias when patent citations are used as a RHS measure of knowledge flows from public research. Our results suggested that nonpatent references—the better of the two citation-based measures—lead to a substantial underestimate of the influence of public research. It is also important to note, however, that in this exercise patent citations were significantly associated with firms' innovative performance, suggesting that, notwithstanding any bias, citations likely capture some of the influence of knowledge flows from public research.

A consistent, robust finding across our analyses is that, compared to citations to other patents, citations to nonpatent references correspond much more closely to managers' reports of the use of public research. The implication is that, relative to citations to other patents, patent citations to the nonpatent literature are the better measure of knowledge flows from public research. However, we caution against combining both patent and nonpatent references into a single measure; this combined measure appears to be inferior to citations to nonpatent references alone.

Our results have implications for how we might view prior research that use patent citations to measure knowledge flows and, more narrowly, nonpecuniary spillovers. For example, in their seminal paper, Jaffe et al. (1993) found that citations made by firms to university patents were more likely to be geographically localized than citations to a set of control patents. They interpreted their findings as evidence that nonpecuniary R&D spillovers from universities to firms are localized.[20] Their paper also paved the way for the use of citations as a measure of knowledge flows more generally. With particular reference to universities, our findings suggest, however, that patent citations do not adequately reflect flows that move through more private channels, notably faculty consulting, even though such channels represent a significant pathway through which university research informs industrial R&D (Cohen et al. 2002). If one believes that private interactions tend to be especially localized, then one might conclude that the impact of knowledge flows may generally be even more localized than that of nonpecuniary spillovers.

---

[20] Note that Jaffe et al. (1993) employ patent citations as a measure of nonpecuniary spillovers rather than knowledge flows generally. Thus, their study is consistent with our results in that citations do appear to reflect the knowledge flowing through the channels of open science, which one may view as the key channel for such spillovers from public research. What is puzzling, however, is that in our analysis it is nonpatent references, not patent references as employed by Jaffe et al., that reflect this link.

Our results offer a number of implications for research. First, to the extent that our survey measure reflects knowledge flows from public research, we conclude that patent citations—albeit a more "objective" and widely available measure—likely overlook key dimensions of knowledge flows. This in turn implies that one may be legitimately suspect of their reliability as simple descriptors of the influence of public research. Second, it appears that systematic measurement error—and particularly errors of omission—may be of concern. How large a concern will depend on whether patent citations are used as a dependent or independent variable, the degree to which the sources of measurement error might be correlated with other independent variables, and how researchers interpret their findings.

So how should researchers use patent citations to measure knowledge flows from public research? First, our analysis suggests that when measuring knowledge flows from public research, citations to nonpatent references are a better measure than citations to patent references. Second, despite the advantages of nonpatent references, scholars should recognize that even such citations do not fully reflect the flow of knowledge to firms, but rather reflect the flow of more codified knowledge. This implies that restricting the use of citation measures to research questions regarding specific activities or types of knowledge flows, such as using nonpatent references to examine the contribution of published science to firm innovation (Gittelman and Kogut 2003, Sorenson and Fleming 2004), may be advisable.

Second, we would recommend, when possible, using controls for dimensions of knowledge flows not reflected in patent citations. For example, one might include measures that control for the extent of private or contract-based relationships and/or firms' basic research activities. In addition, our own preliminary analysis suggests that publications coauthored between academics and industrial R&D personnel appear to be correlated with what we called private interactions.[21] Perhaps even more promising, the National Science Foundation's newly developed survey of industrial R&D collects data on consulting between firms and universities.[22] These data could be used to supplement citation-based measures in assessing the impact of public research on industrial R&D and productivity growth. In any event,

future researchers using citation data should explicitly acknowledge those dimensions of knowledge flows they are attempting to measure and, where possible, include controls to account for dimensions of knowledge flows that are not well captured.

Our ability to generalize our findings regarding citation-based measures to studies that use patent citations as a measure of knowledge flows across firms is limited. First, unlike public research, the outputs of firm R&D are more likely to be patented, and much less likely to be published. Thus, the recommendation to use nonpatent references as a more accurate measure for cross-firm flows does not apply. Nevertheless, citations to other firms' patents likely suffer from some of the same errors of omission previously identified. For example, they are unlikely to reflect flows of knowledge that depend heavily on more private interactions, such as the tit-for-tat exchanges described by von Hippel (1988) and others. Furthermore, errors of commission may also apply to cross-firm flows. For example, firms' propensities to cite prior art more extensively to strengthen the validity of their patents (Allison and Lemley 1998, Alcacer et al. 2009) suggest that, to mitigate future charges of invalidity, firms may over-cite the patents of other firms. Going beyond our study, one might also conclude that the more incremental output of firm R&D that benefits other firms may not be reflected in citations because such output is less likely to be patentable. Nevertheless, our recommendations above still hold: scholars using patent citations to measure knowledge flows across firms should explicitly acknowledge the dimensions of knowledge flows they are attempting to measure, as well as include controls to account for dimensions of knowledge flows that may not be fully reflected in patent citations.

---

[21] This finding is consistent with the work of Cockburn and Henderson (1998) and Zucker et al. (1998), who use coauthored publications as a measure of less codified knowledge flows between universities and pharmaceutical firms.

[22] See question 4-19 of the "2010 Business R&D and Innovation Survey," available at http://www.nsf.gov/statistics/srvyindustry/about/brdis/surveys/srvybrdis_2010.pdf (accessed October 16, 2012).

Appendix

Table A.1 Descriptive Statistics by Use of Public Research and Patenting Activity

	(1)	(2)	(3)	(4)	(5)	(6)
	Do not patent		Patent			
			Cite		Do not cite	
	Use	Do not use	Use	Do not use	Use	Do not use
Observations	238	280	265	168	102	142
%Sample (%)	19.9	23.4	22.2	14.1	8.5	11.9
R&D ($mil)	6.0	1.1	62.3	12.4	5.4	4.1
Number of patents	0.0	0.0	19.2	11.4	1.4	1.5
%R&D projects that use public research (Survey) (%)	36.2	0.0	37.1	0.0	35.0	0.0
%Patents that cite (%)	0.0	0.0	50.7	43.0	0.0	0.0
%Cites to patent references (PR) (%)	0.0	0.0	3.9	3.0	0.0	0.0
%Cites to nonpatent references (NPR) (%)	0.0	0.0	6.9	3.8	0.0	0.0
Avg. no. of backward cites per patent	0.00	0.00	15.15	14.22	10.41	18.16
Avg. no. of cites to PubRes patent	0.00	0.00	1.71	0.83	0.00	0.00
Avg. no. of cites to PatRef per patent	0.00	0.00	0.42	0.33	0.00	0.00
Avg. no. of cites to NPR per patent	0.00	0.00	1.21	0.46	0.00	0.00

Table A.2 Use of Public Research as Predictor of Patenting Activity

Dependent variable:	(1) *Patent* ($Y = 1$)	(2) *Cite public research*	(3) *No. of total citations*	(4) *No. of patent references*	(5) *No. of nonpatent*
Method	Probit	Probit	PQML	PQML	PQML
%R&D projects (Survey)	0.05 [0.18]	0.51** [0.25]	1.21*** [0.32]	0.88*** [0.29]	1.35*** [0.34]
ln(*R&D*)	0.28*** [0.04]	0.24*** [0.05]	0.63*** [0.05]	0.58*** [0.07]	0.69*** [0.06]
Industry dummies (23)	Included	Included	Included	Included	Included
Constant	−3.60*** [0.54]	−3.47*** [0.69]	−8.98*** [0.88]	−9.37*** [1.06]	−10.95*** [0.95]
Observations	1,227	676	676	676	676
Log-likelihood	−679.80	−368.16	−1,502.19	−1,010.13	−1,231.33

$^{***}p < 1\%$; $^{**}p < 5\%$.

Table A.3 Supplemental "Block" Analyses of Measures of Knowledge Flows from Public Research

Dependent variable:	(1) *Survey* %StdX	(2) *PR* %StdX	(3) *NPR* %StdX	(4) *Survey* %StdX	(5) *PR* %StdX	(6) *NPR* %StdX	(7) *Survey* %StdX	(8) *PR* %StdX	(9) *NPR* %StdX	(10) *Survey* %StdX	(11) *PR* %StdX	(12) *NPR* %StdX	(13) *Survey* %StdX	(14) *PR* %StdX	(15) *NPR* %StdX
Channels of knowledge flows															
Open science	183.3***	4.9	17.2**										164.2***	0.5	15.8**
Private interactions	73.8***	−5.1	0.3										55.2***	−8.7*	−0.6
Industrial scientists	16.9**	14.8***	39.4***										11.5	13.1***	24.3***
Uses of public research															
Suggesting new R&D projects				66.3***	7.3	20.9***							42.9***	7.0	6.3
Completion of existing R&D projects				42.9***	4.6	10.2							10.7	3.1	5.8
Firm composition of R&D activity															
ln(*Basic*)							21.0**	−1.3	−4.0				3.9	−3.1	−7.7
ln(*Applied*)							28.2**	18.1**	29.3**				−0.5	13.8**	17.2**
ln(*Development*)							19.5**	9.6*	3.4				10.9	13.4**	5
Firm patenting behavior															
Patent effectiveness										16.6	−8.2	12.4**	12.4	−8.5	10.9
Secrecy										8.5	−2.4	12.2**	1.0	−2.8	−12.1**
Citing propensity										−2.8	30.9***	70.5***	−9.3	28.9***	63.4***

Note. Values reported are the percentage change in the dependent variable for a one-standard-deviation change in the independent variable.
$^{***}p < 1\%$; $^{**}p < 5\%$; $^{*}p < 10\%$.

References

Agrawal A, Henderson R (2002) Putting patents in context: Exploring knowledge transfer from MIT. *Management Sci.* 48(1):44–60.

Alcacer J, Gittelman M (2006) Patent citations as a measure of knowledge flows: The influence of examiner citations. *Rev. Econom. Statist.* 88(4):774–779.

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

Allen TJ (1977) *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D Organization* (MIT Press, Cambridge, MA).

Allison JR, Lemley MA (1998) Empirical evidence on the validity of litigated patents. *AIPLA Quart. J.* 26(3):185–275.

Almeida P, Kogut B (1999) Localization of knowledge and the mobility of engineers in regional networks. *Management Sci.* 45(7):905–917.

Arora A, Ceccagnoli M, Cohen WM (2008) R&D and the patent premium. *Internat. J. Indust. Organ.* 26(5):1153–1179.

Bound J, Brown C, Mathiowetz N (2001) Measurement error in survey data. Heckman JJ, Leamer EE, eds. *Handbook of Econometrics* (Elsevier, Amsterdam), 3705–3843.

Branstetter L, Ogura Y (2005) Is academic science driving a surge in industrial innovation? Evidence from patent citations. NBER Working Paper 11561, National Bureau of Economic Research, Cambridge, MA.

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) *Measurement Error in Nonlinear Models*, Second ed. (Chapman & Hall, London).

Cockburn IM, Henderson RM (1998) Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *J. Indust. Econom.* 46(2):157–182.

Cohen WM, Nelson RR, Walsh JP (2000) Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper 7552, National Bureau of Economic Research, Cambridge, MA.

Cohen WM, Nelson RR, Walsh JP (2002) Links and impacts: The influence of public research on industrial R&D. *Management Sci.* 48(1):1–23.

Cohen WM, Florida R, Randazzese L, Walsh J (1998) Industry and the academy: Uneasy partners in the cause of technological advance. Noll RG, ed. *Challenges to Research Universities* (Brookings Institution Press, Washington, DC), 171–199.

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

Fleming L, Sorenson O (2004) Science as a map in technological search. *Strategic Management J.* 25(8–9):909–928.

Friedman DD, Landes WM, Posner RA (1991) Some economics of trade secret law. *J. Econom. Perspect.* 5(1):61–72.

Gittelman M, Kogut B (2003) Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logit of citation patterns. *Management Sci.* 49(4):366–382.

Griliches Z (1990) Patent statistics as economic indicators: A survey. *J. Econom. Literature* 28(4):1661–1707.

Hall BH, Jaffe AB, Trajtenberg M (2001) The NBER patent citations data file: Lessons, insights, and methodological tools. NBER Working Paper 8498, National Bureau of Economic Research, Cambridge, MA.

Harhoff D, Narin F, Scherer FM, Vopel K (1999) Citation frequency and the value of patented inventions. *Rev. Econom. Statist.* 81(3):511–515.

Henderson R, Jaffe AB, Trajtenberg M (1998) Universities as a source of commercial technology: A detailed analysis of university patenting, 1965–1988. *Rev. Econom. Statist.* 80(1):119–127.

Hicks D (1995) Published papers, tacit competencies and corporate management of the public/private character of knowledge. *Indust. Corporate Change* 4(2):401–424.

Horstmann I, Macdonald GM, Slivinski A (1985) Patents as information-transfer mechanisms—to patent or (maybe) not to patent. *J. Political Econom.* 93(5):837–858.

Jaffe AB, Fogarty MS, Banks BA (1998) Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation. *J. Indust. Econom.* 46(2):183–205.

Jaffe AB, Trajtenberg M, Fogarty MS (2002) The meaning of patent citations: Report on the NBER/Case-Western Reserve study of patentees. Jaffe AB, Trajtenberg M, eds. *Patents, Citations, and Innovations: A Window on the Knowledge Economy* (MIT Press, Cambridge, MA), 379–401.

Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *Quart. J. Econom.* 108(3):577–598.

Lampe R (2012) Strategic citation. *Rev. Econom. Statist.* 94(1):320–333.

Lemley MA, Sampat B (2012) Examiner characteristics and patent office outcomes. *Rev. Econom. Statist.* 94(3):817–827.

MacGarvie M (2006) Do firms learn from international trade? *Rev. Econom. Statist.* 88(1):46-60.

Merton RK (1957) Priorities in scientific discovery: A chapter in the sociology of science. *Amer. Sociol. Rev.* 22(6):635–659.

Mowery DC, Ziedonis AA (2002) Academic patent quality and quantity before and after the Bayh-Dole Act in the United States. *Res. Policy* 31(3):399–418.

Mowery DC, Sampat BN, Ziedonis AA (2002) Learning to patent: Institutional experience, learning, and the characteristics of U.S. university patents after the Bayh-Dole Act, 1981–1992. *Management Sci.* 48(1):73–89.

Narin F, Hamilton KS, Olivastro D (1997) The increasing linkage between U.S. technology and public science. *Res. Policy* 26(3):317–330.

National Science Board (2012) *Science and Engineering Indicators 2012* (National Science Foundation, Arlington, VA).

Pavitt K (1991) What makes basic research economically useful. *Res. Policy* 20(2):109–119.

Peri G (2005) Determinants of knowledge flows and their effect on innovation. *Rev. Econom. Statist.* 87(2):308–322.

Rosenberg N (1985) The commercial exploitation of science by American industry. Clark KB, Hayes RH, Lorenz C, eds. *The Uneasy Alliance: Managing the Productivity-Technology Dilemma* (Harvard Business School Press, Boston), 19–51.

Rosenberg N (1990) Why do firms do basic research (with their own money)? *Res. Policy* 19(2):165–174.

Rosenberg N, Nelson RR (1994) American universities and technological advance in industry. *Res. Policy* 23(3):323–348.

Rosenkopf L, Almeida P (2003) Overcoming local search through alliances and mobility. *Management Sci.* 49(6):751–766.

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

Scherer FM (1983) The propensity to patent. *Internat. J. Indust. Organ.* 1(1):107–128.

Singh J (2005) Collaborative networks as determinants of knowledge diffusion patterns. *Management Sci.* 51(5):756–770.

Singh J, Agrawal A (2011) Recruiting for ideas: How firms exploit the prior inventions of new hires. *Management Sci.* 57(1): 129–150.

Sorenson O, Fleming L (2004) Science and the diffusion of knowledge. *Res. Policy* 33(10):1615–1634.

Thursby J, Fuller AW, Thursby M (2009) U.S. faculty patenting: Inside and outside the unviersity. *Res. Policy* 38(1): 14–25.

Thursby JG, Thursby MC (2002) Who is selling the ivory tower? Sources of growth in university licensing. *Management Sci.* 48(1):90–104.

von Hippel E (1988) *The Sources of Innovation* (Oxford University Press, New York).

Zucker LG, Darby MR, Brewer MB (1998) Intellectual human capital and the birth of U.S. biotechnology enterprises. *Amer. Econom. Rev.* 88(1):290–306.

CORRECTION

In this article, "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research" by Michael Roach and Wesley M. Cohen (first published in *Articles in Advance*, November 9, 2012, *Management Science*, DOI: 10.1287/mnsc.1120.1644), the National Institutes of Health grant information has been added to the acknowledgments section.