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

Michael Roach
Kenan-Flagler Business School
The University of North Carolina
michael.roach@unc.edu

Wesley M. Cohen
The Fuqua School of Business
Duke University
National Bureau of Economic Research
wcohen@duke.edu

This Version: July 2011
Comments and suggestions are welcome.
Please do not cite without permission from the authors.

We would like to thank Jim Adams, Juan Alcacer, Ajay Agrawal, Ashish Arora, Sharon Belenzon, Lee Branstetter, Ernie Berndt, Jim Bessen, Lee Branstetter, Ronnie Chatterji, Iain Cockburn, Maryann Feldman, Steven Klepper, Mike Lenox, Megan MacGarvie, Sangsoo Park, David Ridley, Scott Rockart, Henry Sauermann, Mike Scherer, Jerry Thursby, John Walsh, and Charlie Williams for their valuable discussions and insights, as well as seminar participants at the National Bureau of Economic Research, Harvard Business School, Duke University, Georgia Institute of Technology, and University of North Carolina. We especially thank Bhaven Sampat for graciously providing us with access to patent examiner data. All errors are our own.

ABSTRACT
In this paper, we assess the validity and accuracy of firm backward patent citations as a measure of knowledge flows from public research (i.e., university, government and other nonprofit research) by comparing them to a survey-based measure of public research. Employing a newly constructed dataset that matches patents to survey data at the level of the R&D lab, we identify sources of measurement error associated with backward citations to both patent and nonpatent references, respectively. We also consider the magnitude and direction of potential bias in backward citations as a measure of knowledge flows in a simple empirical model of industrial innovative performance. In brief, we find that patent citations miss a good deal of the content of knowledge flows from public research, especially those flows that tend to be more private, contract-based and not codified. We also show that, as a consequence, measuring knowledge flows with patent citations can lead to a substantial underestimate of the effect of public research on firms' innovative performance. However, our analysis also suggests that, of the two types of patent citations, nonpatent references (e.g., journals, conferences, etc.) are a better measure of knowledge originating from public research.

Keywords: patents, citations, knowledge flows, universities, innovation, measurement

1. Introduction

Over the past 20 years, firms' patent citations have been widely used as a measure of knowledge flows into firms, which have long been thought to be a critical determinant of innovation. Citations are extensively used, for example, as measures of knowledge flows from universities and other public research institutions (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 national borders (Duguet and MacGarvie, 2005; Singh, 2005; MacGarvie, 2006). Indeed, patent citations are likely the most widely employed measure of knowledge flows in the economics, management and policy literatures. This widespread use reflects some notable advantages of patent citations as a data source. In addition to providing indicators of phenomena of interest, their coverage across industries and over time is comprehensive. Moreover, they are readily available, not only from the government, but in more user-friendly form from private sources as well as the NBER patent database (Hall et al., 2001).

To date, however, we know surprisingly little about how well citations measure knowledge flows, though a number of studies, some of them quite recent, provide a basis 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, 2011). Some of these studies, such as those noting the large number of cites contributed by patent examiners, call into question how well patent citations might be used descriptively as indicators of knowledge flows. The absence of alternative, comparable measures of knowledge flows have, however, limited our ability to assess whether citations may be subject to sources of systematic measurement error that compromise their role in analyses of the determinants of innovation and productivity growth, or of knowledge flows themselves.

In this paper, we use a newly developed dataset to compare patent citations to contemporaneous R&D lab managers' reports to assess citations as a measure of knowledge flows from one particular type of knowledge flow—that originating from public research institutions. 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 better understand what dimensions of knowledge flows citations reflect, as well as factors reflected in citations that have little to do with knowledge flows.

Our empirical analysis has three goals. First, we attempt to identify dimensions of knowledge flows that citations should capture and do. We then probe for two possible sources of measurement error: dimensions of knowledge flows that citations should capture but do not, which we term "errors of omission," and, second, factors significantly associated with citations that are not reflective of knowledge

flows, termed "errors of commission." We look for these possible sources of measurement error through both descriptive and regression analyses, employing additional survey responses that should correlate with knowledge flows to explore similarities and differences between patent citations and our survey measure of knowledge flows.

In brief, we find that citations to nonpatent references correspond more closely to managers' reports of the use of public research than do citations to patent references. Moreover, nonpatent references appear to be significantly related with the channels of open science, the employment of Ph.D. industrial scientists and engineers, and the role of public research in suggesting new R&D projects. At the same time, we find evidence of different sources of systematic measurement error in citations to both patent and nonpatent references, most of which suggest a systematic understatement of knowledge flows. For example, patent citations appear to obscure knowledge flows through contract-based relationships with academic scientists (e.g., consulting or cooperative R&D ventures), possibly due to the private, less codified way in which knowledge moves through this channel. Citations also do not reflect the contribution of public research to a firm's basic research, perhaps because the outputs of this activity are less likely to be patented relative to applied research and development. At the same time, citations are significantly associated with factors not related with knowledge flows, such as firms' use of secrecy and strategic citing practices. Finally, in an exploration of the bias associated with these sources of measurement error, we find that the use of citation-based measures as an independent variable leads to a substantial underestimation of the impact of knowledge flows from public research on firm innovative performance.

The next section of the paper provides the background and describes our approach to identifying possible sources of measurement error in patent citations as indicators of knowledge flows from public research. Section 3 describes the data and presents summary statistics comparing and relating the citation-based and the survey-based measures of knowledge flows. In Section 4, we begin our search for the sources of possible measurement error by comparing regression analyses of the two types of data. In addition, we illustrate the magnitude and direction of possible bias associated with the use of patent citations by estimating a simple model of innovative performance that employs the survey measure to control for knowledge flows not captured by patent citations as a predictor. A discussion of the results and future directions concludes the paper.

2. Background & 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 both our interpretation of results that rely upon them, as well as guide efforts to improve their utility. There is cause for believing that

patent 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, there is a question about how well patent citations reflect knowledge flows since 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 claims of the patented invention (Jaffe, et al., 1993). Moreover, what gets cited is influenced by not only the inventor, but also by firms' citing strategies (Lampe, 2011) as well as patent attorneys and 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 and, in turn, as a justification for federal support of public research (e.g., Narin et al., 1997). Second, and more commonly, they have been used as dependent variables in regression analyses examining either the output and commercialization of university research (Henderson, et al., 1998; Mowery et al., 2002) or the use of public research in industrial R&D (Jaffe, et al., 1993). Third, they have also been used as regressors in models explaining the influence of 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 to the degree that associated measurement error is correlated with included right hand side variables. And even where such error is orthogonal to any of the independent variables, the predictive power of the model may be compromised. Where citation-based measures of knowledge flows are employed as right hand side 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, then the direction and magnitude of bias may extend beyond simply the mismeasured variable to affect other variables of interest as well (Bound et al., 2001). In any event, we can expect bias and inconsistency in coefficient estimates.

As noted above, patent citations may be associated with two types of measurement error. First, there may be what we call "errors of omission," which 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. To the degree that firms use knowledge flows from public research in their own basic research, the absence of a patent should cause patent citations to understate the contribution of public research to basic research.

A second possible class of measurement error in patent citations—what we call "errors of commission"—is where firms' citations to public research reflect something other than actual knowledge flows from public research.[2] For example, firms may cite public research less for its contribution to their R&D and more for reasons having to do with their patenting strategies and citing practices. To illustrate, firms with a particular concern over the validity of their patents may include more citations of all sorts in their attempts to make valuable patents less vulnerable to validity challenges in the courts (Allison and Lemley, 1998; Lampe, 2011).

To look for either source of measurement error, we begin with the simple measurement model:

$$k = \beta k^* + v, \tag{1}$$

where k is patent citations, k^* is true, unobservable knowledge flows, and v is the error term.[3] If true knowledge flows, k^*, were observable, then the coefficient β would estimate how well k accurately measures k^*. It is often implicitly assumed that any error in k is "classical," that is, uncorrelated with the true variable k^* such that $\mathrm{Cov}(v, k^*) = 0$ and $\mathrm{E}(v) = 0$. Such classical measurement error in a dependent variable will neither bias nor reduce the consistency of coefficient estimates, but will simply reduce their efficiency. When classical measurement error characterizes an independent variable, however, it may result in attenuated, or downward-biased, and inconsistent coefficient estimates. To the degree, however, that either of these assumptions are violated—that is, either v is correlated with k^*or $\mathrm{E}(v) \neq 0$—then "nonclassical" measurement error may lead to biased and inconsistent estimates (Bound, et al., 2001; Carroll et al., 2006).

To search for possible sources of nonclassical measurement error, we partition k^* into two sets of correlates of true knowledge flows, such that $k^* = k_1 + k_2$, where k_1 is a set of variables reflecting those dimensions of true knowledge flows k^* that a given measure, such as patent citations, should capture and does, and k_2 is a set of variables that the measure should capture but does not. Let us further assume that there is a variable p that is *not* correlated with true knowledge flows k^* but is correlated with the measure,

---

[1] U.S. Federal Government, Office of Management and Budget Circular A-11, available online at

[2] By error of commission, we mean citations may reflect something other than knowledge flows. Note, however, that sources of error of commission can include factors that may cause citations to either under or overstate the contribution of such flows.

[3] This specification is quite similar to the Pakes and Griliches (1984) knowledge production function $p = \beta k^* + v$, where p represents patents as an indicator of knowledge increments, k^* is true knowledge increments, and $\mathrm{cov}(k^*, v) = 0$ by construction.

k. Thus, p contains information on k above and beyond that shared with the true measure k^*. Substituting $\boldsymbol{\beta}_1 k_1 + \boldsymbol{\beta}_2 k_2$ for k^* and adding p into equation (1) yield the following estimable expression for k:

$$k = \alpha + \boldsymbol{\beta}_1 k_1 + \boldsymbol{\beta}_2 k_2 + \gamma p + v \qquad (2)$$

When estimating equation (2) using only patent citations as the measure of knowledge flows, variables reflecting k_2 and p would be omitted and thus subsumed into a composite error term u such that:

$$u = \boldsymbol{\beta}_2 k_2 + \gamma p + v.$$

Since k_2 is a component of k^*, k^* will be correlated with u, violating the requirement for classical measurement error that Cov(u, k^*) = 0. Furthermore, to the degree that the measure k is correlated with p, we would expect $\gamma p \neq 0$, violating the second classical measurement error assumption that E(u) = 0. As a consequence, patent citations contain two possible sources of nonclassical measurement error that may impact not only the accuracy of estimates of knowledge flows, but also bias estimates of other variables of interest that are themselves correlated with components k_2 or p. While many studies that use patent citations are careful to recognize the limitations associated with these measures(Jaffe, et al., 1993), the pervasive assumption is that sources of measurement error are not systematically related with other variables of interest and will, at worse, bias estimates downward.

For us, the first challenge is to identify the sources of systematic error in patent citations as a measure of knowledge flows—that is, the analogues to k_2 and p. As noted above, our approach is to employ an alternative measure of knowledge flows from public research drawn from survey data, k_s, where the subscript s denotes the survey measure. We compare our survey measure to patent citation measures of knowledge flows, denoted as k_c. We will regress both k_s and k_c on a common set of variables, **K**, that we expect to be elements of k^* (i.e., correlates of true knowledge flows from public research), as well as a set of variables that we expect to be elements of ***P*** (i.e., correlates of patent citations but not knowledge flows). Accordingly, we begin our search for sources of measurement error by estimating the following two equations:

$$\mathrm{k_c} = \alpha + \boldsymbol{\beta}_c \mathbf{K} + \boldsymbol{\gamma}_c \mathbf{P} + \mathrm{v_c} \qquad (3)$$

$$\mathrm{k_s} = \alpha + \boldsymbol{\beta}_s \mathbf{K} + \boldsymbol{\gamma}_s \mathbf{P} + \mathrm{v_s} \qquad (4)$$

Our initial goal is to identify: 1.) elements of ***K*** that are associated with both measures of knowledge flows (assumed in our analysis to correspond to k_1 in equation (2)); and 2.) elements of ***K*** that

are significantly associated with one measure of knowledge flows but not the other (assumed in our analysis to correspond to k_2). Our expectation, as argued below in Section 3.1, is that the survey measure k_s will be associated with more dimensions of knowledge flows than will the citation-based measure, k_c. We will further argue that variables not associated with patent citations but associated with the survey measure of knowledge flows are plausible candidates as sources of errors of omission. In addition to considering elements of $\boldsymbol{K}$ that should ideally be associated with any measure of knowledge flows from public research, we are also looking for elements of $\boldsymbol{P}$, that is, variables correlated with k_c that have little direct relationship with knowledge flows from public research, nor, we conjecture, k_s. We believe such variables are candidates as sources of errors of commission. In this analysis, we will assume that, while 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, our central premise is that a direct comparison of the correlates of these two independent measures will shed light on possible sources of measurement error in patent citations.

In our consideration of patent citations below, we distinguish between citations to other patents versus citations to nonpatent references, including, for example, scientific and engineering journals. Although most studies employing patent citations as a measure of knowledge flows from public research rely only upon citations to other patents, we also consider citations to the nonpatent literature in the belief that such citations should more accurately measure the use of public research since nonpatent references are the primary form of public research output (see below; Narin, et al., 1997; Agrawal and Henderson, 2002), and only a small fraction of public research is actually patented (Rosenberg, 1990; Pavitt, 1991; Agrawal and Henderson, 2002).

3. Data

The data for this study are drawn from a novel dataset that combines survey and patent data at the firm R&D lab level. The dataset begins with the Carnegie Mellon Survey (CMS), which contains responses from R&D managers regarding their business units' R&D and patenting activities. Given that CMS was administered in 1994 and asks respondents about their respective labs' R&D activities over the prior three years, we begin with patents applied for between 1991 and 1993. Patents from the NBER patent database (Hall, et al., 2001) were matched to each CMS R&D lab by matching the company names and addresses drawn from the CMS to assignee names and inventor addresses listed on each patent. To assign the patents to a CMS lab, we calculated the proximity of each inventor's residential address to the CMS lab belonging to the same company as the inventor's employer (i.e., the assignee). We then assigned the patent to a CMS lab when the lab had at least one inventor living within 35 miles. For the matched patents, 88% of all listed inventors reside within 35 miles of the matched lab.

While obtaining the patent citations is straightforward, obtaining the nonpatent references, including references to scientific publications, conference proceedings, and other published document, is not. To access nonpatent references, we first retrieved the full record for each identified patent from Delphion, a commercial patent database operated by Thompson Scientific. Since nonpatent references are not provided in a standardized format, a software program was developed to assign each of the more than 55,000 nonpatent references into one of 35 categories, which include journal articles, conference proceedings, government reports, and books. For all journal articles covered in the *Science Citation Index* (SCI) database, we manually retrieved the full record for each publication, which includes 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 labs that patent to allow us to make the most direct comparison possible between the two measures of knowledge flows.[4] Our final dataset of 676 matched labs provides a comprehensive list of patent and nonpatent references as well as inventors' institutional affiliations for patents matched to R&D labs responding to the CMS.

3.1. Measures of knowledge flows

As noted above, our dataset includes two different sources of measures of knowledge flows from public research from: 1.) backward citations to patent and nonpatent references provided in patents, and 2.) reports from R&D lab managers on the use of public research in their R&D projects.

Citation-based measures of knowledge flows – In our patent-based data, we distinguish between two types of backward citations to public research: citations to patent references (PR) and citations to nonpatent references (NPR). Citations to patent references tend to be the most widely employed measure of knowledge flows, due largely to 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 since publications, conference proceedings, and the like are the primary form of public research output (Narin, et al., 1997; Agrawal and Henderson, 2002). This reflects the fact that both incentives and professional norms in academia favor the publication of research findings over more commercial forms of research output such as patents (Merton, 1969; Etzkowitz, 1998; Owen-Smith and Powell, 2001; Agrawal and Henderson, 2002). Moreover, only a small fraction of public research is actually patented (Agrawal and

[4] For the full matched sample of 1,246 observations, 51.7% of R&D labs report using public research to some degree in their R&D projects. Of these labs, approximately 23.6% do not patent, and thus do not cite public research. Yet there is no difference in the survey reported level of knowledge flows between those firms that do patent (36.3% of R&D projects) and those that do not (36.9%). Furthermore, in a corollary probit analysis not reported here, we find that there is no significant relationship between the use of public research and the likelihood of a firm patenting, suggesting that there is no sample selection bias associated with patenting versus nonpatenting labs. Nonetheless, confining our sample to patenting labs would tend to paint patent citations in a more favorable light.

Henderson, 2002), and thus we expect that citations to patents likely reflect different types of knowledge flows relative to citations to publications and other nonpatent references. Finally, recent research on patent examiner effects finds that less than 10% of all nonpatent references are made by examiners (Lemley and Sampat, 2010),[5] in contrast to the more than 40% of patent references inserted by examiners (Alcacer and Gittelman, 2006; Alcacer, et al., 2009). These studies suggest that relative to patent references, nonpatent references may more accurately reflect inventors' views of prior art.

We designate patent references (PR) to public research as citations where the patent assignee is listed as a U.S. university, government lab, research institute, or hospital.[6] We designate nonpatent references (NPR) to public research as citations where at least one author was affiliated with a U.S. university, government lab, research institute, or hospital. While this clearly includes references to papers coauthored with industrial scientists, the observed frequency of such citations for these data is less than 5%. For our matched data, the vast majority of references to public research are nonpatent references such as scientific journal articles (49.8%), conference proceedings, working papers, and reports (8.0%), and edited academic volumes or textbooks (21.5%). Almost 80% of all citations to public research are made to nonpatent references, underscoring the importance of incorporating these measures of knowledge flows. For patent references and nonpatent references taken together, the majority of citations to public research are to universities (83%), and, of those, citations to nonpatent references (mostly scientific publications affiliated with universities) account for 75% of all citations.

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 use public research, reported on a five point categorical scale (less than 10%, 10-40%, 41-60%, 61-90%, and greater than 90%). A strength of this measure, and a contrast with patent citations, is that, since it reflects the views of R&D lab managers, it ideally reflects an informed understanding of the lab's sources of information and knowledge, and thus potentially reflects knowledge flows from public research across the full array of information channels and uses of that knowledge. Thus, our survey measure likely captures dimensions of knowledge flows missed by patent-based measures given that not all inventions are patented and the possibility that citations do not comprehensively characterize the knowledge underlying a firm's inventions. In addition, the survey measure is not influenced by external parties such as patent attorneys or patent examiners, and is thus unlikely to overstate knowledge flows in the same way that citations might. Furthermore, note that, unlike surveys conducted at the firm level, the respondents to the CMS are

---

[5] In an analysis discussed below, we matched examiner data for patents issued after 2001 to the CMS labs and find that, for our sample labs, examiners added 8.9% of all nonpatent references and 37.1% of patent references to public research.

[6] A limitation of this approach is that we are unable to determine which patents may have been assigned to a firm yet were developed in collaboration with a university or other public research institution. As a result, patent citations to patent prior art may understate knowledge flows from public research.

R&D managers who are more likely than other managers or executives to understand the sources of knowledge flows behind the firm's innovations.

Although offering advantages, our survey measure is subject to its own 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 knowledge flows in each project, but rather the breadth of knowledge flows across projects.[7] Third, as with all survey data, respondents may misreport the use of public research due to inaccurate recall, lack of familiarity with actual knowledge flows, socially desirable response to over or understate actual knowledge flows, or for other reasons that may be either randomly or systematically related with the true measure. As such, we do not claim that our survey provides a true measure of knowledge flows or is more accurate than patent citations, but rather that it provides an alternative indicator of knowledge flows that will allow us to cast new light on the possible limits and virtues of patent citations.

Descriptive Comparison of Measures – Table 1 compares the reported fraction of R&D projects that use public research (%Survey) to the fraction of patents that cite at least one public research reference (%Patent) by industry, and Figure 1 graphs this relationship. For purposes of this comparison, we will employ the midpoint from each of the survey response categories. In aggregate, the share of patents citing public research (30.4%) is higher in most industries than the share of R&D projects reported as using public research (20.2%). Consistent with prior research (Narin, et al., 1997; Cohen, et al., 2002), biotechnology exhibits the highest use of public research, at 51% for the survey measure and 87% for the patent measure, followed by pharmaceuticals (30% and 59%), medical devices (23% and 37%), semiconductors (24% and 31%), and computers (25% and 30%). Given that the survey-based measure reflects midpoints of broad categorical ranges, we cannot claim that these absolute differences are meaningful. Of greater interest are the correlations between these two very 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.[8] These four citations measures are: (1) the percentage of patents that make at least one citation to a patent or nonpatent reference from public research, (2) the total number of citations to both patent and nonpatent references, (3) the number of citations to patent references only, and (4) the number of references to nonpatent references only. The different industry-level citation-based measures are highly correlated with the survey-reported use of public research,

---

[7] As reported below, we test the robustness of our analysis to this potential limitation using a CMS measure of the frequency of interaction with public research with qualitatively identical results.

[8] These four citations measures are: (1) the percentage of patents that make at least one citation to a patent or nonpatent reference from public research, (2) the total number of citations to both patent and nonpatent references, (3) the number of citations to patent references only, and (4) the number of references to nonpatent references only.

ranging from 0.87 for the industry average percentage of patents that cite public research to 0.51 for the industry average number of patent references. At the firm level, however, the correlations drop dramatically (ranging from 0.14 to 0.23), and further still in partial correlations that control for industry effects, ranging from a high 0.16 for the number of nonpatent references to a low of 0.12 for the number of patent references. While these correlations suggest that the survey and citation measures likely reflect a common underlying latent variable of knowledge flows, the relatively weak relationships, especially within industries, suggest that they differ considerably across firms.

3.2. Independent Variables

As noted above, our effort to identify possible sources of systematic measurement error begins with: 1.) variables we believe to be correlated with knowledge flows, **K**; and 2.) variables correlated with patenting and citing, **P**, that are not believed to be related to knowledge flows. The Carnegie Mellon Survey (CMS) provides the data from which we construct both of these types of measures. For the former type of correlate—that will help assess errors of omission—the CMS data provide three categories of measures that characterize, respectively: 1.) the nature of the channels through which knowledge is conveyed from public research institutions to firms' R&D labs; 2.) the ways in which public research is used by firms' R&D units; and 3.) the type of R&D conducted by the firms.

For the second type of correlate—that will help assess errors of commission—the CMS provides measures bearing on each R&D manager's perception of the strength of patents as well as that of secrecy, an alternative to patenting. From the patent data, we are also able to construct each lab's propensity to cite prior art.[9] The construction of each variable is detailed below.

4. Analysis of Measures of Knowledge Flows

In the following section, we conduct a series of "block" analyses that compare patent citations with the survey response as measures of knowledge flows by estimating equations (3) and (4). As introduced in the previous section, the four blocks are: 1.) the channels of knowledge flows, 2.) the uses of public research, 3.) the composition of firm R&D activities, and 4.) patenting and citing strategy and behavior. In these analyses we use ordered logit regression to estimate equations where the survey measure is the dependent variable.[10] For the patent citation equations we measure knowledge flows as the numbers of citations to patent references (PR) and nonpatent references (NPR), respectively, and estimated these equations using negative binomial regression. We use the number of citations rather than

[9] As discussed below, we do not include citations to public research in this measure.
[10] In a robustness test discussed below, 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.

the share of citations for comparability to the extant literature which predominately uses citation counts as dependent or independent variables (see, for example, Henderson, et al., 1998; Branstetter and Ogura, 2005; Duguet and MacGarvie, 2005).[11] Given that the mean number of patent citations to public research is small relative to the maximum and the presence of overdispersion, we employ negative binomial regression because it assigns less weight to larger values when adjusting for overdispersion relative to Poisson quasi-maximum likelihood.[12] In addition to coefficient estimates, we also report the percentage change in the independent variable for a one standard deviation change in each dependent variable to standardize results and facilitate comparisons. Given differences across the three measures and regression methods, however, direct comparisons across models should be interpreted with caution.

While 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 descriptive and diagnostic in nature. As such we are careful to interpret our results as suggesting simple relationships between variables and do not attempt to infer causality. An additional concern with relying upon survey measures is common methods bias, which could conceivably magnify the correlations across our survey variables. The question is whether this is an issue when both the RHS and LHS variables are drawn or constructed from the CMS. First, as will be seen below, there are numerous variables drawn from the survey that are not correlated; they are often intended to measure different phenomena, often employ different response scales, and while some are based on Likert scales, others report behavior. Perhaps most importantly, our analyses below demonstrate that there are numerous variables drawn from the CMS that are not related at all to our survey measure of the use of public research. Also, in most cases, the survey questions related to featured variables are separated on the questionnaire by unrelated questions, which reduces priming effects and further mitigates possible spurious correlations between variables.

We now turn to our "block" regression analyses comparing the relationship between the citation- and survey-based measures of knowledge flows, respectively, and the different blocks of correlates. Table 3 reports descriptive statistics and correlations for the featured variables, while Table 4 provides a list of variables and their measures. Columns 1-12 of Table 5 provide results for the survey and patent citation measures of knowledge flows organized by the three sets—or blocks—of correlates of knowledge flows and a fourth block of measures thought to correlate with patenting and citing behavior but not with knowledge flows. For each block, there are three columns of results for an identical set of predictor

[11] To check the robustness of our results, we also performed fractional logistic regressions with both the fraction of total citations that cite public research and the fraction of patents that cite public research with qualitatively similar results.
[12] Poisson quasi-maximum likelihood assumes that variance is proportional to the mean. 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 ran models using quasi-Poisson maximum likelihood and found comparable, although notably weaker results for both the number of citations to patent and nonpatent references, respectively.

variables. Each of the three columns corresponds to one of the three LHS measures: 1.) the survey measure (Survey), 2.) citations to patent references (PR), and 3.) citations to nonpatent references (NPR).

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

Channels of Knowledge Flows – A number of channels of knowledge flows from public research have been considered in previous studies, including publications, public meetings, consulting, and collaborative research with university scientists (e.g., Dasgupta and David, 1987; 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 vehicles through which academics disclose their research, notably publications and public meetings and conferences. Since the primary medium of open science is a codified (i.e., citable) document such as a journal article, conference proceeding, etc., we would expect that a firm's reliance upon a channel of open science should be reflected in both the reported use of public research in R&D projects as well as in citations to public research in patents.

Private, contract-based interactions between public research scientists and industrial R&D personnel such as cooperative research ventures, consulting, or contract R&D also constitute important channels conveying public research to firms, and are particularly effective in the transfer of more complex, less codified knowledge and know-how (Cockburn and Henderson, 1998; Cohen et al., 1998; Zucker et al., 1998; Thursby and Thursby, 2002). While such interactions occasionally produce citable outputs, including publications and reports, the most important of these different channels, namely consulting (Cohen, et al., 2002; Thursby et al., 2009), typically does not. Moreover, since these channels often involve face-to-face communication, a good deal of the knowledge that is transmitted is not typically provided via citable sources. Thus, although these more private channels are important for conveying public research to industrial R&D labs, we suspect that the knowledge conveyed will not be readily reflected in patent citations.

We use exploratory factor analysis of survey responses to the Carnegie Mellon Survey to construct our measures of open science and private channels. Both measures are based on a question that asks respondents to report the importance, on a four-point scale, to the firm's R&D of different channels that convey information about the R&D activities or research findings of public research. We define "open science" channels to include publications, conferences, and informal communication, and "private interactions" channels to include consulting with faculty, contract research, and collaborative research with public research scientists. Loadings from the factor analysis support the two distinct constructs of

open science and private relationships.[13] Factor scores were computed and are used as measures in the regression models.[14]

In addition to the channels of open science and private interactions, the employment of academically-trained M.D. or Ph.D.-level scientists and engineers should also facilitate the flow of public research to the firm. M.D.'s and especially Ph.D.'s are both better equipped to understand frontier academic research, and, given their training, are also more likely to look toward public research as a primary resource. For both of these reasons, we would expect labs with a higher fraction of such academically trained personnel to be more likely to cite public research.[15] We utilize the CMS to construct our measures of industrial scientist employment, which is the reported fraction of R&D personnel who are M.D. or Ph.D. scientists or engineers.

Estimates of the relationship between the importance of the different types of channels of knowledge flow and our three measures of knowledge flows are reported in Columns 1-3 of Table 5. Table 6 presents the Table 5 results in the form of the percentage change of the dependent variable for a one standard deviation increase in each independent variable. Column 1 of Table 5 shows that, as expected, the channels of both open science and private, contract-based relationship 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, on the other hand, are significantly associated with the channels of open science, which is expected given that the principal media of open science—scientific publications and other written documents—are readily citable. Our interpretation of these results is that patent citations—including nonpatent references—may underestimate that component of the contribution of public research that flows through the typically less codified channels of consulting, contract R&D and cooperative R&D. Of concern here is that consulting in particular is one of the most important channels through which public research flows to industrial R&D labs (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 both citation-based measures, although the relationship is strongest with nonpatent references. Given that scientists tend to rely upon publications and other

---

[13] As a general heuristic, a factor loading of 0.70 is considered sufficient to include in a construct measures. 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). An alternative approach is to construct a composite measure of the average of the three respective measures. The Cronbach's alpha coefficient for "open science" is 0.86, while the alpha for "contract-based relationship" is 0.80, suggesting high reliability that they reflect latent constructs of the channels of knowledge flows (a Cronbach's alpha of 0.70 or higher is widely considered a reliable measure).

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

[15] We would also expect this relationship to differ between scientists and engineers. Relative to natural scientists, Ph.D. engineers rely more upon technical reports, reverse engineering, and informal communication with peers (Allen, 1977) .

scientific literature as a key source of knowledge, they are more likely to cite scientific publications (Allen, 1977) and the significant relationship between the fraction of R&D personnel that are scientists and patent citations is expected. However, the considerably larger effect size for the nonpatent reference in Column 3 is striking. A one standard deviation increase in industrial scientists increases the survey measure by 16.9% and citations to patent references by 14.8%. In contrast, it is associated with an increase in citations to nonpatent references of 39.4%. One possible explanation for this relationship with nonpatent references in particular is that it reflects norms more than actual knowledge flows. Ph.D. scientists' greater propensity to cite public research may reflect academic socialization into the practice of generously citing the work of others (Merton, 1957; Sorenson and Fleming, 2004). If the larger number of citations related to the employment of more Ph.D.-level employees is more a matter of norms around citing, then the fraction of Ph.D. and M.D. level scientists in a lab could also account for an "error of commission." On the other hand, teasing out such a normative component from the likelihood that Ph.D.'s greater reliance on public research is difficult. We explore the normative interpretation below.

Uses of Public Research in Firm R&D Projects – Public research may be used in two ways by firms. It can either suggest new projects or contribute to the completion of existing projects. One might interpret the former role as contributing to firms' technological opportunities, and the latter as reflecting the degree to which public research institutions serve as repositories of scientific and engineering knowledge and know-how. Our expectation is that, to the degree that public research suggests new projects, it is more likely to lead to patentable inventions—and thus be linked to citations—than if its contribution is confined more to helping firms execute existing projects. To show this, consider a product patent. That patent will describe the features of the product, not the methods employed to come up with it. To the degree that those methods are suggested by, say, commonly understood theories, know-how, or even textbooks, they are unlikely to be reflected in citations. This point illustrates the difference noted above between citations in scholarly papers versus citations in patents. Although a method for achieving some technical end may well be cited in an academic paper, it is less likely to be cited in a patent on a new product.[16] The CMS provides measures of these two uses of public research, specifically a binary response variable denoted as "suggesting new projects," that equals one if public research was an important source of ideas for new projects in the prior three years, and a second binary response variable denoted as "contributing to the completion of existing projects," that equals one if public research contributed to the completion of existing projects.

[16] Moreover, one might conjecture that, to the degree that public research is a source of new projects, it may be more salient to inventors (Levin et al., 1987; Pavitt, 1991; Mowery and Rosenberg, 1998), and thus more likely to be cited in firm patents.

The highly significant, positive coefficients for these two variables shown in Column 4 in Table 5 suggests that both of these ways in which public research may inform industrial R&D—that is, suggesting new projects and contributing to the completion of existing projects—significantly predict managers' reports of the use of public research in a firm's R&D projects. In contrast, as shown in column 5, the relationship between citations to patent references and the uses of public research is weak. Citations to nonpatent references, on the other hand, appear to reflect the degree to which public research stimulates new R&D projects but not its contribution to existing projects. In sum, although these results suggest that patent citations—especially to the nonpatent literature—capture one important type of contribution of public research to industrial R&D, they also suggest that citations may understate the less readily observed though important role of public research as a source of knowledge contributing to project completion.[17]

Composition of Firm R&D Activity – To the limited extent that a firm may conduct basic research, that basic research activity tends to build more intensely upon knowledge flows from public research than does its downstream applied research and development activities (Rosenberg, 1985, 1990). The results, however, of basic research—if truly basic—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). Applied research and development activities, on the other hand, are directed toward the creation of technological innovations that are more likely to be patentable. Thus, while a firm's basic research activity may draw more heavily upon public research than a firm's applied research or development efforts, its basic research output is less likely to be patented. As a result, public research knowledge flows used by firms in their basic research—the very type of research most reliant upon public research—are less likely to be observed in patent citations. Consequently, we would expect citation-based measures to understate the contribution of public research to their basic R&D activity.

To measure the composition of firms' research activity, we include three measures of 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, and development is technical activity directed toward translating research findings into products or processes. These measures are constructed by multiplying the survey-reported R&D budget (in dollars) for each lab by another survey response on the share of the lab's total R&D activity directed toward each type of R&D, which together account for 100% of a lab's R&D activities.

---

[17] Indeed, the survey results reported in Cohen et al. (2002) suggest that this latter role of public research is at least as important as the former role of suggesting new R&D projects.

The results in Column 7 in Table 5 demonstrate a positive and significant relationship between firms' basic and applied research expenditures and managers' reported share of R&D projects that use public research. For patent 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 activity, 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—even to nonpatent references.[18]

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

Thus far, we have considered correlates of knowledge flows with a view toward identifying possible "errors of omission" in patent citations. In the following exercise, we search for possible sources of "errors of commission," which are variables related to patent citations but not to knowledge flows that may inflate our estimates of knowledge flows. One variable considered above that may have this property is a firm's employment of Ph.D. and M.D. scientists to the degree that such employees cite academic research for normative reasons. In this subsection, we consider possible sources of errors of commission tied to firms' patenting and citing behaviors that grow out of firms' appropriability strategies.

Firms patent a greater share of their innovations when they believe patents to be more effective in protecting their innovations (Arora et al., 2008), and thus tend to have more citations from all sources. Moreover, firms that believe patents to be less effective, say in preventing invent-arounds, may also rely more heavily on secrecy rather than patents to protect their innovations (Horstmann et al., 1985; Friedman et al., 1991). The use of secrecy may also influence citations directly as firms attempt to conceal some aspects of the underlying knowledge upon which the invention was built. Thus, to the extent that firms rely on secrecy, we would expect them to patent less, and, where they do patent, to cite less to the degree that they try not to disclose all the features of their innovations. In either event, they will have fewer citations. Our measures of firms' reliance upon patents and secrecy are drawn from the Carnegie Mellon Survey, and reflect R&D managers' views of the percentage of product and process innovations for which they consider patents and secrecy to be effective as means of protection.

Aside from patenting behavior, firms may also cite public research in ways unrelated to knowledge flows from public research. For example, firms concerned with the risk of litigation or

---

[18] In a corollary analysis designed to consider further the premise of our argument that the output of firm basic research will tend not to be patented, we examine directly whether firms' basic research expenditures are reflected in patent counts, and, as an alternative, whether those expenditures are alternatively reflected in firms' scientific publications. Although not reported here, we find that, as expected, firm scientific publications indeed exhibit a strong positive relationship with a firm's basic research activity while patent counts exhibit no significant relationship. This finding offers an additional implication. It 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 common use of firm patent citations to public research as a proxy for firms' basic research activity.

wishing to strengthen what they view as especially valuable patents may cite more prior art to strengthen their patents against the possibility of invalidity countersuits (Allison and Lemley, 1998; Harhoff et al., 1999). Firms may also make fewer prior art references to maximize the likelihood of issuance with a view toward building their portfolios for defensive purposes (Jaffe, et al., 1993; Lampe, 2011), despite the fact that such patents may be less likely to hold up in court if challenged. Consistent with the importance of strategic citing, we would expect a firm's overall propensity to cite may influence citations to public research in ways unrelated to knowledge flows, and that some component of a firm's citing propensity may be a source of non-classical measurement error that predicts citations to—but not the use of—public research in a firm's R&D projects. We measure a firm's propensity to cite as a lab's average number of backward citations per patent, and, to avoid an obvious source of endogeneity, we exclude citations to public research from this computation.[19]

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 unrelated 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 decreases nonpatent references by 12.2%, and a one standard deviation increase in citing propensity increases citations to nonpatent references by 70.5%. Thus, it would appear that firms' appropriability and citing strategies may be influencing citations to public research, introducing the prospect of what we are calling error of commission.

4.3. Robustness Tests

In this section, we examine the robustness of our results from our block analyses. First, rather than estimating each block separately, as above, we estimate all four blocks simultaneously with the full specification as reported in Columns 13-15 of Table 5. Given collinearity across the right hand side variables, we would of course expect the strength of the relationships between these variables and the three measures of knowledge flows to diminish, and they do. But some overarching results merit mention. First, citations to patent references continue to perform poorly overall. Citations to nonpatent references perform notably better; consider in particular the significant relationship with the channels of open science—suggesting they are arguably a better measure of knowledge flows from public research.

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

Nonpatent references, however, continue to exhibit a negative relationship to secrecy, suggesting sensitivity to that component of firms' appropriability strategies.

To further examine the robustness of the results, we conducted several analyses using alternative measures and estimation methods. To create survey and citation measures that are more directly comparable, we first recoded the Likert survey response 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 (Papke and Wooldridge, 1996). The survey measure produced virtually identical qualitative results compared to the featured results above. The share of patents that cite patent and nonpatent references produced similar, but generally weaker, results in comparison to the number of citations. The only difference worth noting is that the coefficient for Ph.D. scientists and engineers is not significantly related with the share of patents that cite public research.

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 examine this potential limitation, we reproduced our featured results using a survey measure of the frequency with which a firm's R&D personnel use public research, as reported on a five-point scale (rarely or never, semiyearly, monthly, weekly, daily). To the degree that this measure better reflects intensity of use, then it should provide a robustness test of our featured survey measure of the share of R&D projects that use public research. In results not reported here, we find qualitatively identical results between this frequency-based measure and the 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 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 do this by first restricting the sample to firms with equal to or greater than the mean number of patents (8.0) and those with less than the mean with nearly identical results.

In a final robustness test, we examined whether patent examiner added citations might explain the overall weak results for patent references.[20] The challenge for this exercise, however, is that examiner-added citations were not made available by the USPTO until 2001. Accordingly, using patent applications filed in 2001 matched to 351 of the CMS labs used in this study, we first replicated our block

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

analyses using all patent references—both those made by the firm and those made by examiners—with nearly identical results to our featured results in Table 5. 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 and included it as control variable with contemporaneous patent references with no change in the results. Thus, at least for citations to public research, patent examiners seem to have no influence on citations as measures of knowledge flows.

4.4. Isolating Sources of Measurement Error

To summarize the results thus far, patent citations—especially nonpatent references—appear to reflect some dimensions of knowledge flows in common with the survey measure (e.g., open science), which we assume to correspond to k_1 in equation (2) above. Our block analyses above also suggest that there is a second component of knowledge flows, assumed to correspond to k_2, that is associated with the survey measure but not observed in patent citations (e.g., private, contract-based relationships). Finally, our analyses suggest a third component, corresponding to p in equation (2), that reflects elements of patenting and citing behavior unrelated with the survey measure but significantly associated—both positively and negatively—with patent citations (e.g., citing propensity, secrecy).

However, our block analyses leave open several questions. First, even if patent citations mismeasure true knowledge flows as suggested above, are they still informative indicators of knowledge flows from public research? Second, once we control for the common source of variation between the two types of measures of knowledge flows, does there still appear to be systematic sources of measurement error? Also, where some dimensions of knowledge flows are related to both the survey measure and the citation measures, do the two types of measures reflect these dimensions comparably?

To consider these questions, we can regress one measure of knowledge flows onto the other to estimate the strength of the shared component. Furthermore, once the shared component is partialled out of a specification that includes all of our measures for **K** and **P** from equations (3) and (4) above (per columns 13-15 of Table 5), any remaining significant coefficient should signal that there is some component of the associated correlate that still represents a source of error. For example, including nonpatent references as a predictor of the survey measure will account for the shared variation between citations and the survey measure. To the degree, however, that another variable, such as open science, still exhibits a significant relationship with the survey measure after controlling for patent citations, then some component of that measure of open science may still contribute to k_2. Similarly, we can conduct the same exercise including the survey measure as a predictor of patent citations to control for k_1, thereby isolating elements of p and conceivably other elements of k_2.

Applying the outlined approach, we first regress the survey measure onto citations to patent references and nonpatent references, respectively, while also including variables for **K** and **P**. We interpret the statistical significance of the citation coefficients as indicators of the strength of the common component of variation with the survey measure. The coefficient estimates presented in Columns 1 through 4 of Table 7 demonstrate that, across specifications, both patent references and nonpatent references exhibit a significant relationship with the survey measure. As shown in Table 8, a one standard deviation increase in patent references is associated with a 31.7% increase in the survey measures while a one standard deviation increase in nonpatent references is associated with a 36.7% increase in the survey measure. The fact of a strong common component between these two very different measures of knowledge flows—survey-based and citation-based—increases confidence that each measure reflects some element of true knowledge flows. At the same time, we find little change in the qualitative results for the other variables that comprise **K**, suggesting that even after controlling for the shared component, patent citations may understate these dimensions of knowledge flows, at least to the extent that the survey measure reflects them. The results for variables reflecting patenting and citing behaviors are also insignificant, as they were in the full specification in Table 5.[21] In addition to missing some dimensions of knowledge flows, the results from column (4) suggest that nonpatent references, though picking up some of the influence of open science channels per the results from Table 5, may understate that influence.

We now consider errors of commission by regressing both citation measures onto the survey measure. If the survey measure partials out the shared variation with patent citations and also controls for those dimensions of knowledge flows corresponding to the survey but not the citation measures, then we expect that the dimensions associated with k_1, such as open science, should no longer be significant. Furthermore, if we assume that the survey measure reflects the other key dimensions of knowledge flows for which we have measures, then any remaining significant relationship between the other variables and patent citations should reflect potential sources of errors of commission—that is, factors related to patenting, citing and related behaviors that have little to do with knowledge flows. Columns 5-8 in Table 7 illustrate 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 measures is associated with a 8.1% increase in patent

[21] We also performed regressions with patent and nonpatent references entered together, however due to high collinearity of 0.79 neither coefficient was significant. However, regressions that combine both patent and nonpatent references into a single measure produces results that are similar to, but much less precise, 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 but a 20.7% increase in nonpatent references. These results again illustrate that nonpatent references more closely correspond to the survey measure of knowledge flows than do patent references.

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 becomes increasingly more populated by PhD or M.D. level scientists and engineers, we observe more citations to public research beyond what we would expect based on a lab's reported use of public research alone. One interpretation of this finding is that Ph.D. scientists and engineers more accurately attribute knowledge flows from public research to a firm's R&D projects relative to other R&D employees who may undercite the academic literature in their patent applications. The sociology of science (Merton, 1957; Sorenson and Fleming, 2004) suggests an alternative explanation—that Ph.D. scientists are socialized into the practice of generously citing the work of others, perhaps contributing to greater rates of citation relative to other R&D employees who do not have Ph.D.s. Thus, for firms that employ academically trained scientists, a greater share of citations to public research may simply reflect scientists' greater propensity to cite, in which case the number of citations to university and government scientific publications are overstating the use of public research.[22]

4.5. Magnitude and Direction of Bias

Our comparison of survey and patent citation-based measures of knowledge suggests sources of systematic measurement error in patent citations that may lead to bias. However, our prior analysis does not show the magnitude or even direction of potential bias when backward patent citations are used as an independent variable. To provide a sense of the magnitude and direction of the bias that may be introduced by using patent citations as a RHS measure of knowledge flows, we estimate the following simple, illustrative model of the impact of knowledge flows on R&D labs' innovative performance:

$$q_i = \alpha + \beta k_i + \delta r_i + v_i, \qquad (5)$$

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

As discussed above, when using patent citations as a measure of knowledge flows, elements of k_2 and p are unobserved. Consequently, using patent citations as a measure may bias the estimated effect of

[22] This notion is supported anecdotally through interviews conducted by one of the authors who was told by both inventors and patent attorneys that PhD scientists and engineers provide substantially more academic articles as prior art than are relevant to the invention.

the knowledge flows variable, and may also introduce bias into estimates of other independent variables to the degree that those independent variables are correlated with the measurement error. Using our survey measure of knowledge flows from public research, we construct a measure, $\hat{k}_2$, to approximate k_2 in equation (2) above. This is also orthogonal to patent citations (see below). Thus, in our model of innovative performance, $\hat{k}_2$ should reflect those measured dimensions of knowledge flows captured by the survey measure, but not by the patent citation measure.

Although we expect that, consistent with prior research (Gittelman and Kogut, 2003), patent citations to nonpatent references will be positively and significantly related with firm innovative performance, we also expect that the estimated effect of the patent citation measure to not fully reflect the effect of knowledge flows. The coefficient estimate for $\hat{k}_2$ should, however, provide a sense of the magnitude and direction of bias associated with those errors of omission that are picked up by our survey measure. To the extent that variables spawning errors of commission may also be correlated with innovative performance, perhaps indirectly, they may also bias the estimate of our citation-based measure of knowledge flows. The sources of errors of commission, however, may also affect forward citations reflected in other firms' patents. For example, one might expect greater use of secrecy to reduce own patenting, and, in turn, forward citations to the firm's patents. For both of these reasons, we control for the sources of errors of commission directly by including them as predictors in the empirical specification.

We restrict the following analysis to nonpatent references, which appear to better reflect knowledge flows from public research than patent references. In order to compute $\hat{k}_2$, we need to convert our citation measure to percentage units. We employ the percentage of patents that cite nonpatent references to be comparable with our survey-based measure. Column 1 in Table 9 presents the fractional logistic regression used to construct $\hat{k}_2$, where the percentage-based survey measure (%Survey) is regressed onto the percentage of patents that cite nonpatent references (%NPR). From this regression we then predict the percentage of R&D projects that used public research and subtract this from the observed measure to obtain $\hat{k}_2$.

For the innovative performance models we use Poisson quasi-maximum likelihood and report the marginal effect estimates. Column 2 provides the baseline results that estimate the relationship between patent citations to public research (%NPR) and firm innovative performance (CWPC). Column 3 adds measures to control for errors of commission. Although the coefficient for nonpatent references diminishes slightly when the controls are added, the change is not significant. Columns 4 and 5 report comparable results for the survey measure of knowledge flows. Interestingly, the estimated marginal effect for the survey for the survey measure is similar to that for patent citations to nonpatent references.

Nevertheless, when included together in Column 6 both %NPR and %Survey are highlight significant, suggesting that they each reflect unique dimensions of knowledge flows. Column 7 includes %NPR and the computed residual component, $\hat{k}_2$, both of which are positive and significant. Furthermore, the estimated marginal effects for both measures are roughly comparable, suggesting that they each reflect unique but important dimensions of the impact of public research.[23] Although only for a simple, illustrative model, these results suggest that the typically "unobserved" component of knowledge flows has a positive significant relationship with firm innovative performance. Most importantly, the results suggest that the impact of public research on firm innovation may be substantially underestimated—perhaps by half—when relying upon patent citations alone.

The results suggest less of an impact of our errors of commission on the influence of the citation measure of knowledge flows, %NPR, on innovative performance, though this may be due to offsetting effects.[24] We do observe, however, a negative impact that an emphasis of secrecy might have on others' citations to the firm's patents—but even that effect is slight. Finally, we find only modest evidence that nonclassical measurement error in patent citations biases estimates of other independent variables. In this simple model, we only have one other independent variable of note, namely R&D. After correcting for both errors of omission and errors of commission we find that the coefficient estimate of R&D diminishes little. However, the potential for bias in other independent variables exists; the significance and magnitude of this bias likely depends upon the specific sample and empirical specification.

5. Discussion and Implications for Research

Patent citations are increasingly used as measures of knowledge flows in the strategy and innovation literatures. Yet, we know relatively little regarding the validity of these measures. The research of Jaffe et al. (1998; 2002) and Duguet and MacGarvie (2005) suggest that, while such citations appear to reflect knowledge flows, they are noisy. Until now, however, with the exception of Duguet and MacGarvie's (2005) use of the Community Innovation Survey, we have had little against which we could directly compare citations as indicators of knowledge flows. Thus, we have had little basis for assessing whether patent citations may be subject to sources of systematic measurement error that compromise their role as measures of knowledge flows in analyses where such measures serve as either independent or dependent variables (see, for example, 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

[23] We conjecture that the effect on innovation performance captured by the coefficient estimate for patent citations may still overstate the "true" contribution to the degree that forward citations as a measure of innovative performance also likely suffer from sources of measurement error, and some portion is shared with the sources that affect backward citations to public research.
[24] For example, when either patent effectiveness and citing propensity is entered separately, the coefficient estimate for %NPR decreases. Similarly, we find that when our secrecy variable is entered separately, the coefficient for %NPR rises.

and character of knowledge flows from public research with contemporaneous patent data for those managers' R&D labs, we are able to advance our understanding of 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 sources associated with dimensions of knowledge flows from public research that citations fail to reflect, what we call sources of "errors of omission." Possible sources of errors of omission considered in our analyses include ways in which such flows are used, channels through which such flows move, and the R&D activities of the firm that might affect the intensity of use of such flows. Second, we also looked for sources of "errors of commission"—factors related to citations but not directly related to knowledge flows, which may also distort citations as an indicator of flows of knowledge from public research. One possible source of such errors of commission includes firms' appropriability strategies that affect their citation behavior. These might include the degree to which firms rely upon secrecy to protect their inventions, or their overall propensity to cite prior art to defend against subsequent charges of invalidity.

Our initial correlational analyses show a reasonably close correspondence between citation-based measures of knowledge flows and our survey-based measure at the industry level. Where things become more worrisome, however, is within industries (i.e., controlling for industry fixed effects). Although the correlation is still significant, it is modest. Indeed, the correlation coefficients—depending upon exactly which citation-based measure we employ—drop by approximately three quarters. A consistent and robust finding, however, is that, of the two types of citations found in patents—those to other patents versus those to the nonpatent literature—citations to nonpatent references appear to correspond much more closely to managers' reports of the use of public research. Our finding that citations to patent references convey little systematic information regarding knowledge flows from public research is consistent with the results of Agrawal and Henderson (2002) based on interviews of MIT faculty. The most obvious 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.

In our search for sources of systematic error in citations as a measure of knowledge flows, we find that citations to the nonpatent literature partially reflect public research that is disclosed through the channels of open science, as well as public research in its role as a source of new project ideas (i.e., "technological opportunities"). Such citations appear to miss—at least to the extent that our survey measure does not—public research that is conveyed through more contract-based, private channels such as consulting and contract R&D. They also appear to miss the contribution of public research to the completion of existing R&D projects. Thus, our analysis of the different classes of correlates of knowledge flows from public research point to an understanding of what patent citations may or may not reflect that accords well with priors. Citations, particularly to the nonpatent literature, appear to reflect

the research output of public research institutions that shows up in open, often documented ways. What tends to get cited, however, is knowledge embodied in the patented inventions themselves, not the kind of academic know-how and expertise that may help firms overcome the technical challenges that they face in getting to those inventions. The one caveat to this story is that citations appear to not reflect the extent to which firms use university research in their own basic research. But basic research is only a small share of what firms do in their R&D.

We also find evidence for errors of commission. We observe 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 less by knowledge flows and more by firms' concern with the strength and validity of their patents. We also observe a strong negative relationship between patent citations—whether to other patents or to the nonpatent literature—with the degree to which firms feature secrecy in their appropriability strategies. Similarly, though far from conclusive, there is some suggestion in our data that employment of industrial Ph.D. scientists and engineers, while surely a correlate of true knowledge flows, may overstate such flows, perhaps reflecting conformity to academic norms that encourage attribution.

Finally, we estimate a simple, illustrative model of innovative performance to address two important questions: Where patent citations serve as a right hand side measure of knowledge flows from public research, are they associated with an over- or underestimation of the effect of knowledge flows on firm innovative performance? Second, if there is bias, just how important might it be? The answers to both questions undoubtedly depend upon the particular model and specification at hand. Providing some insight into both issues, however, our simple empirical model of firm innovative performance suggests that patent citations to the nonpatent literature—the better of the two citation-based measures of public research knowledge flow—lead to a substantial understatement of the influence of public research.

This study is not without limitations. First, our survey-based measures of knowledge flows and their correlates may suffer from their own sources of measurement error and respondent biases. Second, the cross-sectional nature of our survey data do not allow us to account for potential endogeneity associated with firms' prior activities that may amplify the relationship between the correlates of knowledge flows and the use of public research. Notwithstanding these limitations, we believe that the survey data provide both a valuable point of comparison and a wealth of other measures that allow us to better understand what patent citations reflect and not.

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 thereof. 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—particularly errors of omission—may be of concern. How much of a concern, however, depends upon whether patent citations are used as a dependent or independent variable, and the degree to which the sources of measurement error might be correlated with other independent variables. For example, if, as suggested by many, geographic proximity favors the use of university research in firm innovation due to the fact that an important component of this knowledge requires face-to-face communication (Jaffe, et al., 1993; Cockburn and Henderson, 1998), then relying upon citations may underestimate the mediating influence of proximity if one accepts our analyses suggesting that citations do a poor job of reflecting the knowledge and know-how that are communicated in this fashion.

Our analyses of the sources of errors of omission and commission also suggest there may be differences across industries in the accuracy with which citations may reflect the influence of public research. For example, to the extent that firms within an industry conduct relatively more basic research, engage more frequently in contract-based interactions such as consulting, or rely more heavily on secrecy in protecting their innovations, our analyses suggest greater underestimation of the influence of public research. One way to compensate for these limitations is to include measures that control, for example, for the extent of private or contract-based relationships and/or firms' basic research activities.

So how should researchers use patent citations as a measure of knowledge flows from public research? First, as suggested by our analysis, citations to nonpatent references appear to offer greater utility than citations to patent references as a measure of knowledge flows from public research. Second, despite the advantages of nonpatent references, scholars should recognize that even such citations do not fully reflect the flow of knowledge to firm innovation as has been implicitly assumed (Narin, et al., 1997), but rather reflect the flow of more codified knowledge that is cited in firms' patented inventions. This implies, first, that more special purpose use of citations, such as that of 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, the use of controls for dimensions of knowledge flows not reflected in patent citations. For example, 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 above.[25] Perhaps even more promising, in its newly developed survey of industrial R&D, the National Science Foundation now collects data on consulting between firms and universities.[26] These data could be used conceivably to supplement

[25] This finding is consistent with the work of Cockburn & Henderson (1998), who use coauthored publications as a measure of tacit, less codified knowledge flows between universities and pharmaceutical firms.
[26] See question 4-19 in: http://www.nsf.gov/statistics/srvyindustry/about/brdis/surveys/srvybrdis_2010.pdf.

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 which dimensions of knowledge flows they are attempting to measure and, where possible, include additional controls to account for dimensions of knowledge flows that are not well captured.

Having considered patent citations as a measure of public research, our ability to generalize our findings to studies that use patent citations as a measure of knowledge flows within and between 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 does not apply. Second, the measures used in this study were aggregated to the lab level of analysis, and so it is unclear how well the results translate to patent-level studies. That being said, some of the findings can be generalized to patent citations more broadly. For example, it is conceivable that citations to firms likely suffer from some of the same "errors of omission" identified above. 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 Schrader (1991). Furthermore, firms' propensities to cite prior art more extensively to strengthen the validity of their patents (Allison and Lemley, 1998; Alcacer, et al., 2009) might suggest that firms will cite partner firms engaged in strategic alliances or former employers of recently hired scientists in an abundance of caution to mitigate future charges of invalidity. As a consequence, citations to such firms may more closely correspond to a firm's citing strategy than knowledge flows per se. Going beyond our study, one might also believe that more incremental output of firm R&D that benefits other firms may also not be reflected in citations because such output is less likely to be patentable.

REFERENCES

Agrawal A, Henderson R. 2002. "Putting Patents in Context: Exploring Knowledge Transfer from MIT." *Management Science*, 48(1), pp. 44-60.

Alcacer J, Gittelman M. 2006. "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." *The Review of Economics and Statistics*, 88(4), pp. 774-79.

Alcacer J, Gittelman M, Sampat BN. 2009. "Applicant and Examiner Citations in U.S. Patents: An Overview and Analysis." *Research Policy*, 38, pp. 415-27.

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

Allison JR, Lemley MA. 1998. "Empirical Evidence on the Validity of Litigated Patents." *AIPLA Quarterly Journal*, 26(3), pp. 185-275.

Almeida P, Kogut B. 1999. "Localization of Knowledge and the Mobility of Engineers in Regional Networks." *Management Science*, 45(7), pp. 905.

Bound J, Brown C, Mathiowetz N. 2001. "Measurement Error in Survey Data," In *Handbook of Econometrics,* ed. JJ Heckman, EE Leamer, 3705-843. Elsevier.

Branstetter L, Ogura Y. 2005. "Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations." *NBER Working Paper No. 11561*.

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

Cockburn IM, Henderson RM. 1998. "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery." *Journal of Industrial Economics*, 46(2), pp. 157-82.

Cohen WM, Florida R, Randazzese L, Walsh J. 1998. "Industry and the Academy: Uneasy Partners in the Cause of Technological Advance," In *Challenges to Research Universities,* ed. RG Noll. Washington, D.C.: Brookings Institution Press.

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 No. 7552*.

Cohen WM, Nelson RR, Walsh JP. 2002. "Links and Impacts: The Influence of Public Research on Industrial R&D." *Management Science*, 48(1), pp. 1-23.

Dasgupta P, David PA. 1987. "Information Disclosure and the Economics of Science and Technology," In *Arrow and the Ascent of Modern Economic Theory,* ed. GR Feiwel. New York: New York University Press.

Duguet E, MacGarvie M. 2005. "How Well Do Patent Citations Measure Flows of Technology? Evidence from French Innovation Surveys." *Economics of Innovation and New Technology*, 14(5), pp. 375-93.

Etzkowitz H. 1998. "The Norms of Entrepreneurial Science: Cognitive Effects of the New University-Industry Linkages." *Research Policy*, 27, pp. 823-33.

Fleming L, Sorenson O. 2004. "Science as a Map in Technological Search." *Strategic Management Journal*, 25(8-9), pp. 909-28.

Friedman DD, Landes WM, Posner RA. 1991. "Some Economics of Trade Secret Law." *Journal of Economic Perspectives*, 5(1), pp. 61-72.

Gittelman M, Kogut B. 2003. "Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logit of Citation Patterns." *Management Science*, 49(4), pp. 366-82.

Griliches Z. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature*, 28(4), pp. 1661-707.

Hall BH, Jaffe AB, Trajtenberg M. 2001. "The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools." *NBER Working Paper No. 8498*.

Harhoff D, Narin F, Scherer FM, Vopel K. 1999. "Citation Frequency and the Value of Patented Inventions." *Review of Economics and Statistics*, 81(3), pp. 511-15.

Henderson R, Jaffe AB, Trajtenberg M. 1998. "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988." *Review of Economics and Statistics*, 80(1), pp. 119-27.

Hicks D. 1995. "Published Papers, Tacit Competencies and Corporate Management of the Public/Private Character of Knowledge." *Industrial & Corporate Change*, 4(2), pp. 401-24.

Horstmann I, Macdonald GM, Slivinski A. 1985. "Patents as Information-Transfer Mechanisms - to Patent or (Maybe) Not to Patent." *Journal of Political Economy*, 93(5), pp. 837-58.

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." *Journal of Industrial Economics*, 46(2), pp. 183-205.

Jaffe AB, Trajtenberg M, Fogarty MS. 2002. "The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Study of Patentees," In *Patents, Citations, and Innovations: A Window on the Knowledge Economy,* ed. AB Jaffe, M Trajtenberg. Cambridge, MA: MIT Press.

Jaffe AB, Trajtenberg M, Henderson R. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics*, 108(3), pp. 577-98.

Lampe R. 2011. "Strategic Citation." *Review of Economics and Statistics*, Forthcoming.

Lemley MA, Sampat B. 2010. "Examiner Characteristics and Patent Office Outcomes." *Stanford Law and Economics Olin Working Paper No. 369*.

Levin RC, Klevorick AK, Nelson RR, Winter SG. 1987. "Appropriating the Returns from Industrial R&D." *Brookings Papers on Economic Activity*, 1987(3, Special Issue On Microeconomics), pp. 783-831.

MacGarvie M. 2006. "Do Firms Learn from International Trade?" *Review of Economics and Statistics*, 88(1), pp. 46-60.

Merton RK. 1957. "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review*, 22(6), pp. 635-59.

Merton RK. 1969. "Behavior Patterns of Scientists." *American Scientist*, 58, pp. 1-23.

Mowery D, Rosenberg N. 1998. *Technology and the Pursuit of Economic Growth*. New York, NY: Cambridge University Press.

Mowery DC, Sampat BN, Ziedonis AA. 2002. "Learning to Patent: Institutional Experience, Learning, and the Characteristics of Us University Patents after the Bayh-Dole Act, 1981-1992." *Management Science*, 48(1), pp. 73-89.

Narin F, Hamilton KS, Olivastro D. 1997. "The Increasing Linkage between Us Technology and Public Science." *Research Policy*, 26(3), pp. 317-30.

Owen-Smith J, Powell WW. 2001. "To Patent or Not: Faculty Decisions and Institutional Success at Technology Transfer." *Journal of Technology Transfer*, 26, pp. 99-114.

Pakes A, Griliches Z. 1984. "Patents and R&D at the Firm Level: A First Look," In *R&D, Patents, and Productivity,* ed. Z Griliches. Chicago: University of Chicago Press.

Papke LE, Wooldridge JM. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates." *Journal of Applied Econometrics*, 11, pp. 619-32.

Pavitt K. 1991. "What Makes Basic Research Economically Useful." *Research Policy*, 20(2), pp. 109-19.

Rosenberg N. 1985. "The Commercial Exploitation of Science by American Industry," In *The Uneasy Alliance: Managing the Productivity-Technology Dilemma,* ed. KB Clark, RH Hayes, C Lorenz. Boston, MA: Harvard Business School Press.

Rosenberg N. 1990. "Why Do Firms Do Basic Research (with Their Own Money)?" *Research Policy*, 19, pp. 165-74.

Rosenberg N, Nelson RR. 1994. "American Universities and Technological Advance in Industry." *Research Policy*, 23, pp. 323-48.

Rosenkopf L, Almeida P. 2003. "Overcoming Local Search through Alliances and Mobility." *Management Science*, 49(6), pp. 751.

Rosenkopf L, Nerkar A. 2001. "Beyond Local Search: Boundary-Spanning, Exploration, and Impact in the Optical Disk Industry." *Strategic Management Journal*, 22(4), pp. 287-306.

Scherer FM. 1983. "The Propensity to Patent." *International Journal of Industrial Organization*, pp. 107-28.

Schrader S. 1991. "Informal Technology Transfer between Firms: Cooperation through Information Trading." *Research Policy*, 20, pp. 153-69.

Singh J. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science*, 51(5), pp. 756-70.

Singh J, Agrawal A. 2011. "Recruiting for Ideas: How Firms Exploit the Prior Inventions of New Hires." *Management Science*, 57(1), pp. 129-50.

Sorenson O, Fleming L. 2004. "Science and the Diffusion of Knowledge." *Research Policy*, 33(10), pp. 1615-34.

Thursby J, Fuller AW, Thursby M. 2009. "US Faculty Patenting: Inside and Outside the Unviersity." *Research Policy*, 38(1), pp. 14-25.

Thursby JG, Thursby MC. 2002. "Who Is Selling the Ivory Tower? Sources of Growth in University Licensing." *Management Science*, 48(1), pp. 90-104.

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

Zucker LG, Darby MR, Brewer MB. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review*, 88(1), pp. 290-306.

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

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

Table 2 Correlations with Survey Measure of Knowledge Flows

	Obs.	(1) % Patents that cite	(2) # Total references	(3) # Patent references	(4) # Nonpatent references
Correlations					
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*
Chronbach's alpha [c]					
Industry-level alpha [a]	24	0.93	0.83	0.46	0.87
Firm-level alpha	677	0.37	0.33	0.24	0.36

Notes: Correlations and alphas 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 ISIC industries represented; [b] partial correlations controlling for 24 ISIC industries; [c] scale reliability coefficient (standardized items).

Table 3 Descriptive Statistics and Correlation Matrix

		Mean	StDev	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1)	Survey (%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*

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, 5-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 university, government lab, or non-profit 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 non-profit research institute
Channels of Knowledge Flows		
Open science	Survey	Factor score of the importance of publications, conferences, and informal communication as a sources of knowledge from public research
Private interactions	Survey	Factor score of the importance of faculty consulting, contract research, and collaborative R&D as a sources of knowledge from public research
Industrial scientists	Survey	Fraction of total R&D employees who are MD or PhD 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 a providing a competitive advantage
Secrecy	Survey	Percentage of firm's product and process innovations for which secrecy was effective a 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; used when patent citations are measure of knowledge flows to control for the level of patenting activity
Industry dummies	Survey	24 ISIC dummy variables

Table 5 Comparison of Measures of Knowledge Flows from Public Research

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Dependent Variable	Survey	PR	NPR	Survey	PR	NPR	Survey	PR	NPR	Survey	PR	NPR	Survey	PR	NPR
Regression Model	Ordered Logit	Negative Binomial	Negative Binomial	Ordered Logit	Negative Binomial	Negative Binomial	Ordered Logit	Negative Binomial	Negative Binomial	Ordered Logit	Negative Binomial	Negative Binomial	Ordered Logit	Negative Binomial	Negative Binomial
Channels of Knowledge Flows															
Open Science	1.22***	0.06	0.19**										1.13***	0.01	0.17**
	[0.11]	[0.05]	[0.08]										[0.12]	[0.05]	[0.07]
Private Interactions	0.67***	-0.06	0.00										0.53***	-0.11+	-0.07
	[0.10]	[0.07]	[0.08]										[0.11]	[0.06]	[0.07]
Industrial Scientists	0.93**	0.82***	1.97***										0.65	0.73***	1.29***
	[0.46]	[0.27]	[0.45]										[0.52]	[0.22]	[0.34]
Uses of Public Research															
Suggesting New R&D Projects				1.06***	0.15	0.39***							0.74***	0.14	0.13
				[0.19]	[0.10]	[0.14]							[0.20]	[0.11]	[0.13]
Completion of Existing R&D Projects				0.72***	0.09	0.20							0.21	0.06	0.11
				[0.17]	[0.10]	[0.13]							[0.20]	[0.10]	[0.13]
Firm Composition of R&D Activity															
ln(Basic)							0.03**	-0.00	-0.01				0.01	-0.01	-0.01
							[0.01]	[0.01]	[0.01]				[0.01]	[0.01]	[0.01]
ln(Applied)							0.07**	0.04***	0.07**				-0.00	0.03**	0.04**
							[0.03]	[0.02]	[0.03]				[0.02]	[0.01]	[0.02]
ln(Development)							0.05**	0.03+	0.01				0.03	0.04**	0.01
							[0.02]	[0.01]	[0.02]				[0.03]	[0.01]	[0.02]
Firm Patenting Behavior															
Patent Effectiveness										0.50	-0.28	0.38**	0.38	-0.29	0.34
										[0.32]	[0.18]	[0.18]	[0.32]	[0.18]	[0.22]
Secrecy										0.26	-0.08	-0.42**	0.03	-0.09	-0.41**
										[0.26]	[0.16]	[0.19]	[0.27]	[0.15]	[0.19]
Citing Propensity										-0.00	0.03***	0.06***	-0.01	0.03***	0.05***
										[0.01]	[0.00]	[0.01]	[0.01]	[0.00]	[0.01]
Control Variables															
Industry dummies (23)	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.
ln(patent)		0.91***	1.07**		0.90***	1.04***		0.88***	1.08***		0.96***	1.15***		0.91***	1.12***
		[0.04]	[0.06]		[0.04]	[0.06]		[0.03]	[0.05]		[0.04]	[0.05]		[0.03]	[0.04]
ln(rd)	0.06	0.04	0.08	0.11**	0.03	0.07				0.17***	0.02	0.04			
	[0.04]	[0.03]	[0.05]	[0.04]	[0.03]	[0.05]				[0.04]	[0.03]	[0.04]			
log-likelihood	-784.49	-821.46	-1071.81	-845.23	-822.29	-1078.52	-886.51	-821.34	-1085.97	-887.12	-811.06	-1043.18	-770.13	-797.69	-1025.41
Observations	676	676	676	676	676	676	676	676	676	676	676	676	676	676	676

Notes: The dependent variables are Survey (fraction of R&D projects that use public research), PR (number of citations to patent references), NPR (number of citations to nonpatent references); ordered logistic regression estimates are reported for the survey measure, negative binomial regression estimates are reported for the citation count measures; robust standard errors in brackets; *** $p < 1\%$, ** $p < 5\%$, + $p < 10\%$

Table 6 Percentage Change in Measures of Knowledge Flows

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Dependent Variable	Survey	PR	NPR	Survey	PR	NPR	Survey	PR	NPR	Survey	PR	NPR	Survey	PR	NPR
	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%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; *** $p < 1\%$, ** $p < 5\%$, + $p < 10\%$.

Table 7 Isolating the Sources of Measurement Error

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Errors of Omission				Errors of Commission			
Dependent Variable	Survey	Survey	Survey	Survey	PR	PR	NPR	NPR
Regression Method	Ordered Logit				Negative Binomial			
Measures of Knowledge Flows								
ln(PR)	0.36**	0.33**						
	[0.15]	[0.16]						
ln(NPR)			0.32***	0.26**				
			[0.11]	[0.11]				
Survey					0.38+	0.34+	1.31***	0.81***
					[0.21]	[0.20]	[0.28]	[0.28]
Channels								
Open Science		1.13***		1.12***		-0.03		0.07
		[0.12]		[0.12]		[0.05]		[0.07]
Private Interactions		0.55***		0.55***		-0.12**		-0.11
		[0.11]		[0.11]		[0.06]		[0.07]
Industrial Scientists		0.52		0.53		0.68***		1.19***
		[0.53]		[0.53]		[0.23]		[0.35]
Uses of Public Research								
Suggesting New R&D Projects		0.73***		0.72***		0.11		0.07
		[0.20]		[0.20]		[0.11]		[0.12]
Completion of Existing R&D Projects		0.21		0.22		0.05		0.06
		[0.20]		[0.20]		[0.10]		[0.12]
Composition of R&D Activity								
ln(Basic)		0.01		0.01		-0.01		-0.02+
		[0.01]		[0.01]		[0.01]		[0.01]
ln(Applied)		-0.00		-0.00		0.04**		0.05**
		[0.03]		[0.02]		[0.01]		[0.02]
ln(Development)		0.03		0.03		0.03**		0.01
		[0.03]		[0.03]		[0.01]		[0.02]
Patenting Behavior								
Patent Effectiveness		0.43		0.38		-0.33+		0.30
		[0.32]		[0.32]		[0.19]		[0.21]
Secrecy		0.04		0.04		-0.09		-0.38**
		[0.27]		[0.27]		[0.15]		[0.19]
Citing Propensity		-0.01		-0.01		0.03***		0.05***
		[0.01]		[0.01]		[0.00]		[0.01]
Control Variables								
Industry dummies (23)	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.
ln(patent)	0.05	-0.23**	0.02	-0.24**	0.93***	0.91***	1.14***	1.13***
	[0.10]	[0.12]	[0.10]	[0.11]	[0.03]	[0.03]	[0.04]	[0.04]
log-likelihood	-890.61	-768.01	-889.16	-767.55	-825.12	-796.49	-1077.88	-1021.57
Observations	676	676	676	676	676	676	676	676

Notes: N = 676. The dependent variables are survey (fraction of R&D projects that use public research), PR (number of citations to patent references), NPR (number of citations to nonpatent references); ordered logistic regression estimates are reported for the survey measure,; negative binomial regression estimates are reported for the count citation measures; Robust standard errors in brackets; *** p < 1%, ** p < 5%, + p < 10%.

Table 8 Percentage Change in Measures of Knowledge Flows

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Errors of Omission				Errors of Commission			
Dependent Variable	Survey	Survey	Survey	Survey	PR	PR	NPR	NPR
	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%StdX	%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***

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%.

Table 9 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	QPML	QPML	QPML	QPML	QPML	QPML	QPML
%NPR (k_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 ($k_1 + k_2$)				156.85***	157.66***	139.52***		
				(47.95)	(49.88)	(46.01)		
%Survey_residual ($\hat{k}_2$)							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(RD)		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 dummies (23)	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.
Observations	676	676	676	676	676	676	676	676

Notes: N = 676. Marginal effect estimates reported for all models; Column 1 regresses %Survey onto %NPR (percentage of patents that cite public research) using fractional logistic regression to estimate $\hat{k}_2$; Columns 2-6 are results for citation-weighted patent counts (CWPC) using quasi-Poisson maximum likelihood (QPML); Columns 1-6 Robust standard errors in brackets; Columns 7-8 bootstrapped standard errors in brackets; *** p < 1%, ** p < 5%, + p < 10%.

Figure 1 Mean comparison of measures by industry