Available online at www.sciencedirect.com
ScienceDirect
Research Policy 36 (2007) 850–863
research policy
www.elsevier.com/locate/respol

Minerva unbound: Knowledge stocks, knowledge flows and new knowledge production☆

Lynne G. Zucker[a,e,f,*], Michael R. Darby[b,e,f], Jonathan Furner[c,e], Robert C. Liu[d,e], Hongyan Ma[c,e]

[a] *Departments of Sociology and Public Policy, University of California, UCLA Box 951551, Los Angeles, CA 90095-1551, United States*
[b] *Departments of Management, Economics and Public Policy, UCLA Anderson School of Management, UCLA Box 951481, Los Angeles, CA 90095-1481, United States*
[c] *Department of Information Studies, University of California, UCLA Box 951520, Los Angeles, CA 90095-1520, United States*
[d] *Deloitte Consulting LLP, 350 S. Grand Avenue, Suite 200, Los Angeles, CA 90071-3462, United States*
[e] *Center for International Science, Technology, and Cultural Policy, UCLA School of Public Affairs, University of California, UCLA Box 951656, Los Angeles, CA 90095-1656, United States*
[f] *National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138, United States*

Available online 1 May 2007

Abstract

Regional growth of new knowledge in nanotechnology, as measured by counts of articles and patents in the open-access digital library NanoBank, is shown to be positively affected both by the size of existing regional stocks of recorded knowledge in all scientific fields, and the extent to which tacit knowledge in all fields flows between institutions of different organizational types. The level of federal funding has a large, robust impact on both publication and patenting. The data provide support for the cumulative advantage model of knowledge production, and for ongoing efforts to institutionalize channels through which cross-organizational collaboration may be achieved.

Keywords: Bibliometrics; Boundary Spanning; Collaboration; Nanotechnology; Organization type

☆ This research has been supported by the National Science Foundation (grant SES-0304727) and the University of California's Industry-University Cooperative Research Program (grants PP9902, P00-04, P01-02, and P03-01). We are indebted to our research team members Jason Fong, Yaqiang Qi, and Emre Uyar. Nanotechnology data used in this paper is part of NanoBank.org, a community resource building project funded by NSF. In this paper, we both introduce and demonstrate some of the potential uses for this new database, as well as test the underlying data in preparation for public launch of NanoBank. Certain data included herein are derived from the *Science Citation Index Expanded, Social Sciences Citation Index, Arts and Humanities Citation Index* of the Institute for Scientific Information®, Inc. (ISI®), Philadelphia, PA, USA: ©Copyright Institute for Scientific Information®, Inc. 2005. All rights reserved. Certain data included herein are derived from the Zucker–Darby Knowledge, Innovation, and Growth Project's Science & Technology Agents of Revolution (STAR) database ©Lynne G. Zucker and Michael R. Darby. All rights reserved. This paper is a part of the NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of their employers or the National Bureau of Economic Research.
* Corresponding author at: Departments of Sociology and Public Policy, University of California, UCLA Box 951551, Los Angeles, CA 90095-1551, United States. Tel.: +1 310 825 9155; fax: +1 310 454 2748.
E-mail addresses: zucker@ucla.edu (L.G. Zucker), darby@ucla.edu (M.R. Darby), jfurner@ucla.edu (J. Furner), robcliu@ucla.edu (R.C. Liu), hym@ucla.edu (H. Ma).

0048-7333/$ – see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.respol.2007.02.007

1. Introduction

What factors influence the rates at which new knowledge is produced in technological fields? The particular study reported in this paper is part of a more general research program driven by this question, whereby we seek to understand the processes that determine the productivity of authors and inventors in new technology, as measured by counts of articles and patents (see, e.g., Zucker et al., 1998a,b; Zucker et al., 2002). In the study reported here, our counts of documents are obtained by statistical analysis of the contents of NanoBank, an open-access digital library of articles and patents in the field of nanotechnology (Zucker and Darby, 2006). Our approach is guided by a theoretical conception of the production of scientific knowledge as an activity that is deeply embedded in a complex network of social structures and practices, and that the forms taken by these structures and practices are crucial determinants of the forms taken by knowledge production in later periods in the same place. While it is conventional to refer to science as cumulative, we argue and demonstrate that there is a significant cumulative effect even when the knowledge produced is discontinuous and revolutionary in some respects.

We present the results of tests of two related hypotheses. The first of these is that the frequency of publication, during a given period and in a given geographical region, of articles and patents *relating to nanotechnology* is correlated with the size of the existing "knowledge stock" of all other (non-nanotechnology) articles and patents *in all fields of science* previously published in that region. The second hypothesis is that the frequency of publication, during a given period and in a given geographical region, of articles and patents relating to nanotechnology is correlated with the extent to which articles and patents in the existing knowledge stock of the region are *co-authored* by affiliates of institutions of different organizational types.

The results of our tests allow us to draw two sets of conclusions. In the first place, we are able to differentiate the respective merits of two competing kinds of claims about the ways in which existing knowledge stocks affect the evolution of new fields of knowledge such as nanotechnology. In the second place, we are able both to evaluate, on the basis of their impact on productivity, ongoing efforts to institutionalize channels through which cross-institutional collaboration (or "knowledge flow") may be achieved, and to demonstrate the utility of a method by which the impact of stocks of tacit knowledge (as opposed to that of stocks of recorded knowledge) may be estimated.

In the course of our investigations of the links between knowledge stocks, knowledge flows, and knowledge production, we are also able to assess the impact (on productivity of knowledge in nanotechnology) of the cumulative stock of funding dollars awarded by the National Science Foundation (NSF) to nanotechnology projects initiated by institutions in a given region.

The paper is structured as follows. Firstly, we contextualize our hypotheses by considering the impact, on the production of new knowledge, of general knowledge stocks (Section 2), and of barriers to the flow of knowledge across institutional boundaries (Section 3). We then provide a justification of our focus on geographically localized knowledge flow (Section 4), before describing our methods of measuring knowledge (Section 5), of identifying "nano-relevant" documents (Section 6), and of categorizing those documents by organizational type and geographical region (Section 7). In Section 8, we describe our methods of data analysis; in Section 9, we present the results of the tests of our first hypothesis, about the impact of knowledge stocks; and in Section 10, we present the results of the tests of our second hypothesis, about the impact of knowledge flows. Finally, we draw our conclusions (Section 11).

2. General knowledge stocks: their impact on the production of new knowledge

Researchers in the economics of scientific knowledge have long been concerned to assess the impact of knowledge production on economic growth (see, e.g., Stephan, 1996; Foray, 2004). How closely do measurements of the rates at which new knowledge is produced correlate with measurements of the rates at which the economy grows as a whole? A number of production functions have been proposed that model the relationship between output quantities of goods and services and input quantities of knowledge. Considerable attention has also been paid to the task of identifying the conditions under which rates of knowledge production (and thus economic productivity in general) can increase most rapidly. Correspondingly, production functions have been developed that may be used to predict the rate at which new knowledge will be produced in the future (see, e.g., Adams, 1990).

Comprehensive functions of this latter kind typically quantify inputs of three principal types: time, physical resources, and human (i.e., intellectual) resources. In practice, the intellectual capital accessible to an institution includes both (i) knowledge that is recorded or codified in documents, and (ii) the tacit knowledge or know-how that is stored only in the minds of the institution's scientists and researchers. Research in economics (see, e.g., Griliches, 1990) suggests that the impact of cumulated general knowledge stocks (i.e., knowledge capital) on the production of new knowledge is *positive*. In other words, the larger the existing cumulated stock of general knowledge, the faster new knowledge is produced: rates of production of new knowledge increase in direct proportion to sizes of stocks of existing knowledge. Such a finding supports the general claim, widely accepted by economists, that scientific knowledge is strongly cumulative (see, e.g., Stigler, 1983; Machlup, 1984). According to the cumulative advantage model of the knowledge production process, having one idea increases the likelihood of having another: knowledge begets knowledge.

The validity of the cumulative advantage model has been questioned from various perspectives. On one hand, it is criticized for its reliance on the assumption that the existing knowledge base consists of propositions about the world that are themselves true. Work in the history and philosophy of science (HPS) and science and technology studies (STS) tends to emphasize the sharp continuity breaks in the evolution of fields of knowledge that are observed to follow breakthrough discoveries (see, e.g., Kuhn, 1962), and thus to suggest that the size of existing knowledge stocks is either *irrelevant* for predictions of future knowledge productivity, or even *negatively correlated* in the sense that the existence of large stocks of prior general knowledge actively retards the making and acceptance of discoveries in new fields.

From another perspective, the cumulative advantage model has been attacked for failing to take into account the gradual obsolescence (and depreciation in economic value) of knowledge, and the variation in obsolescence rates among fields. Nevertheless, the correlation between size of existing knowledge stock and rate of production of new knowledge has been observed even when the value of previously-produced knowledge is discounted (conventionally by 20% per year) to reflect the way in which older knowledge becomes obsolete over time (see, e.g., Griliches, 1990).

Thirdly, it has been argued that the positive externalities of knowledge in a given field are most frequently restricted or localized to third parties working in that field—i.e., that the stock of knowledge produced in one field has little impact on the rate of production of new knowledge in other fields (see, e.g., Antonelli, 2001).

The study described in this paper supplies data that can be interpreted as further evidence of the validity of the cumulative advantage model, addressing in particular the third criticism mentioned above. We report measurements of (i) the size of prior stocks of knowledge in all fields, and (ii) the rate of production of new knowledge in the field of nanotechnology, and find that the correlation between sets of measurements of the two kinds is typically positive. We believe that this finding reflects the capability of disruptive, breakthrough discoveries in science still to draw upon, or at least not contradict, some of the concepts, formulae, and machines developed in pre-existing science. While some elements in pre-existing science may be radically changed or eliminated, other old elements are imported largely unchanged, while still others are transformed by their use for new purposes, yielding hybrid elements that mix old and new.

3. Barriers to the flow of knowledge across institutional boundaries: their impact on the production of new knowledge

A much remarked property of knowledge as an economic good is that it is capable of "spillover": it may be acquired and used freely by people working in institutions other than those in which the knowledge originates, while the originating institution retains access to that knowledge but receives no further compensation for its diffusion (see, e.g., Arrow, 1962). Knowledge production is thus said to generate "positive externalities" or benefits for third parties. Codified knowledge flows through text books, scientific equipment, and lecture hall presentations. Tacit knowledge is embodied in the person and is communicated with great difficulty, often by doing research together with the person learning. The freedom for tacit knowledge to flow across institutional boundaries is frequently identified as one of the essential conditions under which rates of knowledge production can increase most rapidly.

Ideally, the intellectual capital input to a knowledge production function should account not only for stocks of recorded knowledge but also for stocks of tacit knowledge. Recorded knowledge may be quantified using counts of publications (and perhaps qualified, more controversially, by weighting publication counts by counts of the citations made to those publications by other authors). It has proven more difficult, on the other hand, to measure stocks of tacit knowledge. Nevertheless, since there is evidence to suggest that researchers' engagement in collaborative activity is the principal way in which tacit knowledge is generated and shared, collaboration contains evidence of likely significant tacit knowledge stocks, once value and skill differentiation are controlled.

The simplest and most commonly occurring process by which knowledge is transferred across institutional boundaries involves people from different institutions interacting with one another, and (more specifically) collaborating on projects whose results include the production of some new knowledge. Micro-institutional approaches look at the social construction process, including transmission of knowledge and information (Zucker, 1977) and "institutionalization projects" (DiMaggio, 1982a,b).

Co-authoring across institutional boundaries serves to connect the two organizations, while simultaneously converting each co-author into quasi-members of the other co-authors' organizations, such that a subsequent promotion of one of the authors may include a letter of evaluation from an external collaborator (Zucker, 1991; Gartrell, 1987). This process of external evaluation of job performance also occurs in firms, e.g. investment banks where externally collected survey data from clients is used to determine bonuses and promotions (Eccles and Crane, 1988). Action becomes embedded in the virtual space created by the social network (Powell, 1990). Thus, the deepening of channels or conduits for the flow of knowledge can serve as a catalyst for changes in the degree and type of social embeddedness of knowledge-production activity Gulati (1999).

As the frequency and number of different scientists working with each other across institutional boundaries increases, the "gap" between institutions may begin to be converted into a "glissando" (Tyree et al., 1979). When this kind of boundary change occurs repeatedly across institutions of two organizational types (e.g., university and firm), the social distance between the two types is reduced and institutional differentiation may decrease, changing the context in which action is embedded. For new knowledge to have an impact, and perhaps become taken-for-granted, the knowledge not only needs to be produced, but also transmitted to others. Pre-existing practices of co-authorship and collaboration provide existing channels along which the new information is likely to flow. These collaborations can be viewed as micro-joint ventures, where success encourages more joint work, further embedding persons in cross-organizational contexts. As the number of these "micro-JVs" adds up, the cumulative effects may operate like an organization-level joint venture, and lead to a move to the other organization by one of the scientists.

The factors considered by researchers when choosing whether or not to engage in collaborative knowledge-production activity are various. Whatever the factors that are perceived by individuals as incentives in particular contexts, the force of such incentives is typically diminished by the existence of barriers to collaboration, such as the absence (or unreliability) of formalized structures of the kinds that might support cross-institutional interaction. It is clear, for instance, that the prior lack of channels for cross-institutional communication is a strong disincentive to collaborate on cross-institutional research, with the result that institutional boundaries function as knowledge envelopes, preventing knowledge from leaving the institution in which it was produced (Zucker et al., 1996).

In our prior work, we have examined the effects of direct ties, measured as numbers of articles co-authored by scientists in universities and scientists in firms, on commercialization of basic science discoveries made in universities (Zucker et al., 1998a,b, 2002; Zucker and Darby, 2001). It is not yet equally clear, however, what impact the prior existence of channels for cross-institutional communication has on the rate at which knowledge is produced in new fields. It might be hypothesized that the depth of such channels is positively correlated with knowledge productivity—in other words, that the more frequently cross-institutional collaboration has occurred (and knowledge has flowed) in the past, the greater quantity of knowledge will be produced in the immediate future. In this paper, we investigate this hypothesis in relation to the field of nanotechnology, reporting our findings that the absence of barriers to the flow of general (i.e., "non-nano") knowledge in the past has a positive impact on the rate at which new knowledge is produced in nanotechnology.

4. Geographically localized knowledge flow: a theory of localization

A stream of recent research on innovation in the U.S. has found evidence of "geographically localized knowledge spillovers" occurring in areas around major universities (Jaffe, 1986,1989a,b; Jaffe et al., 1993; Audretsch and Feldman, 1996; Henderson et al., 1998). The underlying assumption is that proximity to a major university itself provides technological opportunity; the localization is assumed to be due to the social ties between university and firm employees or to firm employees' access to seminars at the university. The importance of distance is strengthened by Adams and Jaffe (1996) finding that geographic distance is an important impediment to flow of technology even within the firm.

Zucker et al. (1998a,b) and Darby and Zucker (2001) find that firms are more likely to begin using biotechnology in U.S. and Japanese regions where and when "star" bioscientists are actively publishing, respectively. Although these findings have been cited as evidence of geographically localized knowledge spillovers, we read our results – and those of the other authors cited above – as only demonstrating geographical localization of knowledge. Zucker et al. (1998a,b, 2002) and Zucker and Darby (2001) show for California, the U.S., and Japan, respectively, that university effects on nearby firm R&D productivity are highly concentrated in the particular firms with bench-science working relationships with top academic scientists and practically absent otherwise. We identify these academic–firm *links* by the academic scientist publishing a journal article that also has one or more firm-affiliated authors. Fieldwork – supported by analysis of the timing of the academic scientists' first articles with a firm and its founding – indicates that these academic – firm co-publishing relationships most often connote that the academic scientist was a firm founder or at least presently has a significant financial interest in the firm. In our view, knowledge localization occurs because high levels of tacit knowledge characteristic of high-science discoveries require inventor involvement for successful transfer to firms. The implied temporary natural excludability produces high potential returns which help motivate the involvement of top scientists in the commercial application of their discoveries.

This paper builds on this literature in two ways: First, we hypothesize that both recently published and patented knowledge creation will have localized effects on the current production of knowledge. Second, we hypothesize that it will be more difficult to engage in collaborations across organizational type in a new area like nanotechnology in those local regions where there is little prior precedent for such collaborations than in those regions with a more extensive history of such collaborations. That is, the inherited institutional setting of a local region may limit or enhance the ability of scientists and engineers to engage in productive collaborations across organizational type.

5. Overcoming difficulties in the measurement of knowledge

The characteristics of knowledge that make it impossible for researchers to measure its quantity (let alone its quality) *directly* are well documented (see, e.g., Foray, 2004). In this paper, we adopt a strategy for measurement that will be familiar to sciento- and bibliometricians (see, e.g., Hullmann and Meyer, 2003; Moed et al., 2004): we use counts of articles (i.e., articles published in the scientific literature) and counts of patents as separate *indicators* of the quantities of knowledge produced in given fields, in given sectors, in given regions, in given periods.

The limitations of such an approach are also well documented; suffice to say that we are only too aware (i) that such counts obscure large variations in the quality of the knowledge recorded in articles and patents, and (ii) that a large but unknown proportion of the knowledge produced and used by researchers and by developers remains unrecorded in articles and patents. One common response to the former observation is to weight publication counts in accordance with counts of the citations received by publications. We have not made use of citation statistics in this study since the nature of the relationship between citedness and quality is itself unclear.

We have been developing an open-access digital library of articles and patents in the field of nanotechnology, covering all aspects of the science and technology of nanometer-scale structures and systems (Zucker and Darby, 2006). Nanotechnology is an emergent, highly interdisciplinary field. The criteria that we use when deciding whether or not to add a new article or patent to NanoBank are described in Section 6.

Each of the records in NanoBank includes metadata that specify, inter alia, the date of the article's publication or the patent's grant, the institutional affiliation of each of the article's authors or the patent's assignees (and the addresses of named institutions), and any discipline(s) or subject area(s) to which the article or patent has been assigned in previous acts of classification. These metadata can then be processed by computer (using procedures described in Section 7) in order to generate counts of articles and patents tabulated by year, by geographical area, by organizational type, by subject area, and by type of cross-institutional collaboration.

These counts – indicating the rate of production of new knowledge in nanotechnology – are then compared with counts of "non-nano" articles and patents that indicate the size of existing general knowledge stocks. These latter counts are generated in a similar manner from existing databases of general coverage produced by Thomson Scientific (formerly the Institute for Scientific Information, Inc., ISI), and the United States Patent and Trademark Office (USPTO).

6. Identifying nano-relevant documents

In order to populate NanoBank with articles and patents that treat topics related to nanotechnology (i.e., that are "nano-relevant" or, simply, "nano"), we need to filter such documents from initial universal sets of documents covering all subject areas. Nanotechnology is an interdisciplinary endeavor, and nano-relevant documents may be found inhabiting many disciplinary spaces, both expected and unexpected. The two universal document sets with which we began are (for articles) the union of *Science Citation Index Expanded*™, *Social Sciences*

Citation Index®, and *Arts & Humanities Citation Index*® produced by Thomson Scientific, and (for patents) the database of US patents produced by the Zucker–Darby Knowledge, Innovation, and Growth Project. Thomson Scientific's data cover more than 24 million articles from more than 8700 peer-reviewed journals; our patent data cover the 3,891,720 patents granted by the USPTO from 1976 to 2005.

The fraction of the content of these document sets that is nano-relevant is substantial, so the task of filtering nano-relevant documents is one that can be carried out only with automated assistance. We use two separate methods of distinguishing documents that are "nano": (a) Search for the occurrence of terms in a predefined list of 379 terms identified by subject specialists as nano-specific to indicate "nano." We found this method to be less reliable in identifying the latest and the earliest nano-relevant documents: the latest are more likely to include terms that are too new to have made it on to our list, while the earliest were written before the terms on our list were in common usage. (b) Discriminate between nano-relevant documents and others using a probabilistic procedure for the automatic identification of those terms (Sebastiani, 2002). This procedure is adaptive in the sense that it requires the computer to continuously learn from the training data provided by prior sets of judgments of the relevance or non-relevance of documents, and integrates techniques developed in the fields of information retrieval (IR), machine learning (ML), and natural language processing (NLP).

We begin by assuming the nano-relevance of the articles that make up the *Virtual Journal of Nanoscale Science & Technology* (a.k.a. *VJNano*; http://www.vjnano.org/), a weekly compilation of the latest research on nanoscale systems whose contents is selected manually, from a variety of source publications, by the members of an international editorial board. In the technical vocabulary of IR, *VJNano*'s selection policy is analogous to the output of a search with very high precision but potentially low recall: few, if any, non-nano-articles are selected, but it is not known how many nano-relevant articles are not selected.[1] Our task is then to identify those documents that have not been selected by *VJNano* but that are nevertheless nano-relevant; in other words, we want to classify the non-*VJNano* literature into nano (relevant) and non-nano (non-relevant). The terms that occur most frequently in *VJNano* (and that do not appear on a stop-word list of terms that occur frequently in English text in general) are weighted according to their frequencies and used in a query against a universal document set (Van Rijsbergen, 1979). The system's response to such a query is a list of documents ranked in order of their probability of relevance. In a "blind" feedback process, we then assume the nano-relevance of the top-ranked documents, use data on the frequencies of occurrence of terms in those newly-identified documents to modify the previous query, and submit the modified query against the universal document set (Efthimiadis, 1996). We repeat the process until we converge on a relatively consistent set of terms that changes little between iterations.

A refinement of this procedure involves the generation of separate ranked lists of highly discriminating terms for different subject areas. Every journal indexed by Thomson Scientific, for instance, is assigned to one or more subject categories, and every article appearing in a given journal inherits the category or categories to which that journal has been assigned. We were able to match with their counterparts in the Thomson Scientific databases 17,693 of the 22,732 articles appearing by December 2005 in *VJNano*, and then to generate a separate sub-classifier (i.e., a relatively stable query that may be used to retrieve nano-relevant documents) for each of the 235 subject categories to which Thomson Scientific's articles are assigned. Sub-classifiers can similarly be generated for each of the five categories of patent in the broad science and technology classification scheme based on the World Intellectual Property Organization (WIPO) International Patent Classification (IPC) that is used in the USPTO database. Even more specific sub-classifiers can be generated for every publication year.

7. Categorizing documents by organizational type and geographical region

Each Thomson Scientific record of an article includes fields that supply the name and address of each of the institutions with which any of that article's authors is affiliated; each USPTO record of a patent includes fields that similarly supply the name and usually the address of the institutional assignee-at-issue, and the home address of each inventor. We analyze the data contained in these fields in order to categorize the knowledge-producing institutions by organizational type or sector [i.e., as a firm, university/hospital, government/national laboratory, research institute/national professional organization, "unclassified" (type not yet determined)] and by geographical region. Institutions are classified by organizational type using a sequen-

[1] *VJNano* also covers a limited range of journals, those published by the American Institute of Physics or by the American Physical Society, plus a core group of interdisciplinary journals.

tial application of two methods: (i) the application of look-up tables, developed over several years of automatic and manual processing, that match variant names to preferred names and thence to organizational type; and (ii) keyword analysis for remaining unidentified organizations that infers organizational type from certain of the words, phrases, and abbreviations contained in an institution's name (so that, for example, a name that includes "Inc." is classed as a firm). Addresses are classified by county and then by geographical region using the Federal Information Processing Standard database (FIPS55; http://geonames.usgs.gov/fips55.html) maintained by the U.S. Geological Survey, and definitions of the 179 functional economic areas in the U.S. supplied by the U.S. Bureau of Economic Analysis (Johnson and Kort, 2004).

Statistical analysis of the contents of NanoBank yields data on the numbers of nano-relevant articles published in each of 235 subject categories and patents granted in each of 5 subject categories, in each of 5 sectors, in each of 179 U.S. regions, in each of the 24 years from 1981 through 2004. If an article or patent is associated with more than one subject category, organizational type, or region, each count is credited a fraction equivalent to the reciprocal of the number of associated types or regions. If an article is co-authored by authors affiliated with institutions of different organizational types within the same region, it is counted as an instance of cross-institutional collaboration or "knowledge flow."[2] Corresponding data on numbers of non-nano-articles and patents are derived by analyzing the residual contents (i.e., the contents left after subtracting nano-articles and patents) of the Thomson Scientific and USPTO databases.

Table 1 provides a summary of these data. Table 2 provides further detail on the relative frequencies of articles whose co-authors span institutional boundaries.

8. Data analysis methodology

Our data analysis is performed on panel data – a time series of cross sections – comprised of observations for each of the years 1981-2004 for each of the 179 U.S. functional economic areas (i.e., central urban areas plus their suburbs and exurbs or "regions") as defined by the U.S. Bureau of Economic Analysis. Most of these regions have a relatively low rate of publication and patenting while a smaller tail have much higher rates.

We analyze this data using Poisson regressions with random effects and robust standard errors estimated using the xtPoisson procedure in the Stata 9.0 statistical package. This procedure is both flexible and robust. For example, since the negative binomial is a special case of the random effects model (Kennedy, 1998, pp. 247–248), it permits the data to choose that form rather than have it imposed which potentially biases the coefficient estimates if the true model is not negative binomial.[3] Since we use the local region as the group variable in these estimations we are fitting separate constants for each region and crediting the model with explanatory power only to the extent that it explains variation over time relative to the mean rates of publication or patenting for each region. This provides something of an acid test for the significance of the coefficients individually and as a group or groups.

To be very specific, the general form of the random intercept Poisson model estimated by xtPoission can be expressed by the following equations:

$$\log(\lambda_{ij}) = \beta_{0j} + \sum_{k=1}^{K} \beta_k X_{kij} \tag{1}$$

$$\beta_{0j} = \eta_{00} + \alpha_{0j} \tag{2}$$

Taking (1) and (2) together, we have

$$\log(\lambda_{ij}) = \eta_{00} + \sum_{k=1}^{K} \beta_k X_{kij} + \alpha_{0j} \tag{3}$$

where λ_{ij} is the expected number of events occurred for the jth region at the ith year, β_{0j} indicates an intercept that is random for each region, and α_{0j} is an error term for the jth region that follows a log-gamma distribution with mean zero.

Compared to a simple Poisson model estimated based on the pooled data, our estimator allows for heterogeneity among regions by estimating the variance of α_{0j}. When the estimated variance is not significantly different from zero (i.e., when there is no heterogeneity among regions), the estimator will give identical results as a simple Poisson regression. In our case, all the estimated variances of α_{0j} are significantly different from zero, which suggests that we cannot treat observations from different regions as if observations from the same region, ignoring the hierarchical nature of the data. Although a random intercept model does not explicitly explore the source of heterogeneity (which is not the focus of the

[2] Our data on the extent of cross-institutional collaboration are currently limited to co-authorship of papers: we have not analyzed patents for instances of knowledge flow.

[3] The negative binomial is nested within the random effects model which itself is a very simple form of the hierarchical linear model (random coefficients) estimation.

Table 1
Variable names and summary statistics

Variables	U.S. regions				
	N	Mean	S.D.	Min	Max
Variables measured without differentiating organization types					
Nano-articles published in year	4296	25.690	82.330	0	1025.873
Nano-patents granted in year	4296	11.946	57.868	0	1329.448
Non-nano-articles knowledge stock	4296	5.598	13.707	0	124.281
Non-nano-patents knowledge stock	4296	0.992	2.565	0.001	35.842
NSF nano-funding stock for region (in US$ millions)	4296	0.092	0.301	0	5.842
Year	4296	1992.500	6.923	1981	2004
Region (BEA number)	4296	90.000	51.678	1	179
University and hospital variables					
University nano-articles	4296	18.950	57.632	0	651.666
University nano-patents	4296	1.518	7.044	0	104.239
University non-nano-articles knowledge stock	4296	4.183	9.630	0	86.983
University non-nano-patent knowledge stock	4296	0.028	0.086	0	0.971
Firm variables					
Firm nano-articles	4296	3.159	20.878	0	607.591
Firm nano-patents	4296	9.494	48.771	0	1157.022
Firm non-nano-articles knowledge stock	4296	0.550	1.890	0	21.713
Firm non-nano-patent knowledge stock	4296	0.901	2.357	0.001	33.294
Government and federal or national laboratories variables					
Government nano-articles	4296	1.763	9.459	0	158.970
Government nano-patents	4296	0.349	2.783	0	70.332
Government non-nano-articles knowledge stock	4296	0.233	1.344	0	20.469
Government non-nano-patent knowledge stock	4296	0.025	0.093	0	1.231
Research institutes and national professional organization variables					
Research inst. nano-articles	4296	0.663	3.644	0	78.172
Research inst. nano-patents	4296	0.132	0.857	0	15.367
Research inst. non-nano-articles knowledge stock	4296	0.127	0.484	0	6.723
Research inst. non-nano-patent knowledge stock	4296	0.004	0.014	0	0.143
Unclassified (not yet classifiable as above) variables					
Unclassified nano-articles	4296	1.156	7.716	0	173.494
Unclassified nano-patents	4296	0.453	2.166	0	63.295
Unclassified non-nano-articles knowledge stock	4296	0.504	1.676	0	22.003
Unclassified non-nano-patent knowledge stock	4296	0.035	0.102	0	1.389
Cross-organization-type knowledge flow variables					
Non-nano-articles knowledge flow	4296	62.929	232.531	0	3289.931
Firm-university non-nano-articles knowledge flow	4296	13.522	50.725	0	541.824
Firm-research institute non-nano-articles knowledge flow	4296	0.421	2.387	0	57.440
Firm-government non-nano-articles knowledge flow	4296	1.453	10.685	0	218.875
Firm-unclassified non-nano-articles knowledge flow	4296	2.096	10.357	0	206.753
University-research institute non-nano-articles knowledge flow	4296	9.094	42.095	0	628.370
University-government non-nano-articles knowledge flow	4296	6.542	33.200	0	530.031
University-unclassified non-nano-articles knowledge flow	4296	26.450	85.920	0	1180.020
Government-research institute non-nano-articles knowledge flow	4296	0.428	4.918	0	124.131
Government-unclassified non-nano-articles knowledge flow	4296	2.147	25.912	0	497.000
Research institute-unclassified non-nano-articles knowledge flow	4296	0.691	4.394	0	99.631

Notes: (1) Knowledge stocks are computed as a perpetual inventory of the indicated series with 20% per year depreciation and measured in thousands of cumulated articles or patents. (2) Each variable is measured each year 1981–2004 for each of the U.S. regions which are the 179 functional economic areas defined by the U.S. Bureau of Economic Analysis (Johnson and Kort, 2004). $N = 24$ years × 179 regions = 4296. (3) We include in our patent counts only those with an assignee at issue to an organization separate from the inventor(s). Patents are located by the address of the inventors and classified as to organization type by assignee. (4) The NSF nano-funding stock is computed like the knowledge stocks, but measured in millions of US dollars. (5) Knowledge flows are computed as the number of articles co-authored across organization types within one BEA.

Table 2
Organization type boundaries: spanning the fences

Organization type	Total number of ISI journal articles	All authors within the org. type		Authors across 2+ org. types	
		Number articles	Percent	Number articles	Percent
(A) Non-nano-articles					
Universities and hospitals	4,380,043	4,141,152	95	238,891	5
Firms	563,504	488,356	87	75,148	13
Government, national and federal laboratories	238,056	192,654	81	45,402	19
Research institutes and natl. professional orgs.	137,807	91,755	67	46,052	33
Unclassified (not yet classifiable as above)	539,765	404,577	75	135,188	25
(B) Nano-articles					
Universities and hospitals	81,409	75,729	93	5,680	7
Firms	13,570	10,691	79	2,879	21
Government, national and federal laboratories	7,572	5,422	72	2,150	28
Research institutes and natl. professional orgs.	2,846	2,198	77	648	23
Unclassified (not yet classifiable as above)	4,965	3,460	70	1,505	30
(C) All articles					
Universities and hospitals	4,461,452	4,216,881	95	244,571	5
Firms	577,074	499,048	86	78,026	14
Government, national and federal laboratories	245,628	198,076	81	47,552	19
Research institutes and natl. professional orgs.	140,653	93,953	67	46,700	33
Unclassified (not yet classifiable as above)	544,730	408,037	75	136,693	25

Note: These counts are based, respectively, on all non-nano-articles, all nano-articles, and all articles listed in the science citation index expanded, 1981–2004, and the organization matching program of the Zucker–Darby knowledge, innovation, and growth project.

current paper), it nevertheless takes it into account and gives adjusted estimates for regressors that we report.

We elected to use robust standard errors in order to provide *t*-statistics which are not biased by deviations from the assumed parametric model. Nonetheless, in response to a comment by an anonymous referee, we produced alternative estimates based on the bootstrapping procedure. Since these results are very similar to those reported here, we believe that the basic methodology is well suited to the data being analyzed. The full bootstrapping results are reported in the Appendix to Zucker et al. (2006), available at www.nber.org or from the authors.

9. The impact of general knowledge stocks on the production of knowledge in nanotechnology: empirical results

Table 3 summarizes the results of a Poisson regression with random effects analysis that indicates the strength of the relationship between three explanatory variables—the cumulative stock[4] of non-nano-articles written by authors affiliated with institutions in a given region, the cumulative stock of non-nano-patents granted to inventors affiliated with institutions in a given region, and (for purposes of comparison) the cumulative stock of funding dollars awarded by the National Science Foundation (NSF) to nanotechnology projects initiated by institutions in a given region[5] – and two dependent variables – the number of nano-articles written in the corresponding region in a given year, and the number of nano-patents granted in the corresponding region in a given year. There is no standard measure of goodness of fit like the R^2 for this procedure, perhaps because it would be controversial to take credit for the explanatory power of the differing regional means. We note that all the coefficients but one in Table 3 are highly significant and do make comparative evaluations of goodness of fit using Wald (1943) log-likelihood tests.

[4] The sizes of all stocks reported in our data are computed by cumulating counts for all previous years, and discounting by 20% annually to reflect depreciation.

[5] Nano-relevant awards were identified by searching the NSF "Award Search" website (http://nsf.gov/awardsearch/tab.do?dispatch=3) for the following 17 terms identified by subject specialists as indicators of nano-relevance: atomic force microscop*; buckminsterfullerene; c60; fullerene; giant magnetoresistance; langmuir blodgett; mesoscopic; nanocrystal*; nanoparticle; nanoscale; nanostructur*; quantum confin*; quantum dot; quantum well; quantum wire; scanning tunneling microscop*; self-assembled monolayer. The asterisk is a wild-card symbol: documents that contain any term beginning with the string of characters before the asterisk will be marked as "nano."

Table 3
Nano-publishing and patenting: knowledge stock effects Poisson regressions with random effects[a] for U.S. regions, 1981–2004

Explanatory variables for region and year	Nano-scale article and patent			
	Nano-article	Nano-article	Nano-patent	Nano-patent
Non-nano-articles Knowledge stock	0.076*** (0.001)	0.074*** (0.001)	0.073*** (0.001)	0.074*** (0.001)
Non-nano-patents knowledge stock	0.001 (0.001)	−0.013*** (0.001)	0.069*** (0.001)	0.058*** (0.001)
NSF nano-funding stock		0.439*** (0.008)		0.262*** (0.014)
Constant	1.974*** (0.118)	1.922*** (0.117)	0.974*** (0.091)	0.957*** (0.110)
Log likelihood	−30567.4***	−29168.4***	−15648.6***	−15482.7***
Wald log-likelihood Test for NSF nano-funding stock[b]	–	2798.0***	–	331.8***

Notes—Dependent variables: nano-articles published in year and nano-patents granted in year. Robust standard errors in parentheses below coefficient estimates. $N = 4296$. Significance levels: ˆ0.10, *0.05, **0.01 and ***0.001. Knowledge stocks (KS) are computed as a perpetual inventory of the indicated series with 20% per year depreciation and are measured in thousands of cumulated articles of patents.

[a] Group variable for random effects: region (BEA number), 179 groups.

[b] This Wald (1943) log-likelihood test is for the null hypothesis that adding the NSF nano-funding stock does not significantly improve the fit of the regression.

From these results in Table 3, we can determine that the rates of production both of nano-articles and of nano-patents are higher in regions with a larger cumulative stock of non-nano-articles. However, a larger cumulative stock of non-nano-patents has a positive impact only on the rate of production of nano-patents, and not on the rate of production of nano-articles. This finding is consistent with a linear model in which academic discoveries (which dominate the article counts) add to the research productivity of nearby firms (which dominate the patent counts), but not vice versa. Adding the NSF nano-funding stock variable very significantly increases the explanatory power of the regressions relative to the corresponding regressions without that variable. The same is true for all other regression specifications reported below. NSF nano-funding significantly increases regional research productivity whether measured by articles or patents—implying that both universities and firms are more productive in years and regions with above average support by the NSF. We cannot distinguish from these non-experimental data the extent to which NSF is independently creating discoveries by its funding or successfully identifying and facilitating the research efforts of the best and brightest scientists. This result comes through loud and clear regardless of which model specification we tried. In Table 3 it also makes the article impact of the cumulative non-nano-patent stock significantly negative. We are tempted to interpret this result as due to more competition for top academic scientists' time in areas with very successful firm research efforts, but note that the negative coefficient is not statistically significant using the alternative bootstrap standard errors discussed at the end of Section 8.

In Table 4, we break down the cumulative stocks by sector. For the dominant players in article production (universities) and in patent production (firms),[6] the strengths of the relationships between cumulative stocks and annual productivity are similar to those we see in Table 3. However, the results of the regression for other organizational types display considerable instability that is difficult to interpret: in research institutes, for example, cumulative non-nano-patent stocks have a strong positive impact on patent production but a strong negative impact on article production, whereas, in the same sector, cumulative non-nano-article stocks have a negative impact all round. Since the models in Table 3 can be considered to be nested in the models of Table 4 when the coefficients for the articles knowledge stocks are all constrained to equal the same number and the coefficients for the patents knowledge stocks are all constrained to equal the same number (possibly different from the one for articles), we can use another Wald test (reported at the bottom of Table 4) to see if relaxing the constraints significantly improves the explanatory power of the regressions. The reported χ^2 statistics are all extremely large, demonstrating that the more complicated models of Table 4 have much more explanatory power than the simple models of Table 3.

We believe that the significant anomalous signs for smaller participants in publishing and patenting results from multicollinearity among the regressors. While multicollinearity is always a matter of degree and difficult to assess in nonlinear models like ours, we appear to have a moderate level of multicollinearity as measured by the

[6] See Table 1.

Table 4
Nano-publishing and patenting: knowledge stocks by organization type Poisson regressions with random effects[a] for U.S. regions, 1981–2004

Explanatory variables for region and year	Dependent variable			
	Nano-articles	Nano-articles	Nano-patents	Nano-patents
University non-nano articles knowledge stock	0.134*** (0.002)	0.121*** (0.002)	0.087*** (0.003)	0.079*** (0.003)
University non-nano patents knowledge stock	−1.974*** (0.108)	−2.046*** (0.109)	−3.029*** (0.151)	−3.211*** (0.152)
Firm non-nano articles knowledge stock	0.158*** (0.006)	0.180*** (0.006)	0.214*** (0.007)	0.228*** (0.007)
Firm non-nano-patents knowledge stock	0.016*** (0.004)	−0.018*** (0.004)	0.101*** (0.005)	0.080*** (0.005)
Government non-nano-articles knowledge stock	0.187*** (0.009)	0.139*** (0.009)	−0.053*** (0.015)	−0.094*** (0.015)
Government non-nano-patents knowledge stock	−1.893*** (0.098)	−2.251*** (0.099)	−1.103*** (0.156)	−1.382*** (0.158)
Research institute non-nano-articles knowledge stock	−0.164*** (0.018)	−0.112*** (0.019)	−0.157*** (0.028)	−0.094*** (0.028)
Research institute non-nano-patents knowledge stock	−5.711*** (0.355)	−5.451*** (0.351)	8.466*** (0.435)	9.101*** (0.436)
Unclassified non-nano-articles knowledge stock	−0.034*** (0.009)	−0.010 (0.009)	0.120*** (0.014)	0.139*** (0.014)
Unclassified non-nano-patents knowledge stock	0.935*** (0.073)	1.292*** (0.074)	0.900*** (0.103)	1.067*** (0.103)
NSF nano-funding stock		0.425*** (0.008)		0.317*** (0.014)
Constant	1.754*** (0.116)	1.760*** (0.116)	0.817*** (0.108)	0.826*** (0.108)
Log likelihood	−28197.4***	−27040.5***	−14810.4***	−14583.4***
Wald log-likelihood test for NSF nano-funding stock[b]	–	2313.8***	–	454.0***
Log-likelihood test vs. Table 3[c]	4740.0***	4255.8***	1676.4***	1798.6***

Notes—Dependent variables: nano-articles published in year and nano-patents granted in year. Robust standard errors in parentheses below coefficient estimates. $N = 4296$. Significance levels: ˆ0.10, *0.05, **0.01 and ***0.001. Knowledge stocks (KS) are computed as a perpetual inventory of the indicated series with 20% per year depreciation and measured in thousands of cumulated articles of patents.

[a] Group variable for random effects: region (BEA number), 179 groups.

[b] This Wald (1943) log-likelihood test is for the null hypothesis that adding the NSF nano-funding stock does not significantly improve the fit of the regression.

[c] This Wald (1943) log-likelihood test is for the null hypothesis that constraining all the articles-knowledge-stock coefficients to be the same and all the patents-knowledge-stock coefficients to be the same (as is implicit in Table 3) does not significantly reduce the goodness of fit the regression.

variance inflation factor (VIF). When multicollinearity is a serious problem, the estimated coefficients as well as the standard errors are unstable and tend to vary widely with a slight change of the observations. The bootstrap procedure discussed at the end of Section 8 is a standard response to concerns about multicollinearity. In general, the bootstrapped standard errors reported in Zucker et al. (2006) are somewhat larger than those reported in Tables 3–6, and some of the coefficients become insignificant, but the main, quantitative conclusions still hold.

10. The impact of general knowledge flow on the production of knowledge in nanotechnology: empirical results

Table 5 displays the results of more Poisson with random effects regressions, this time run with cumulative counts of *cross-institutional-type co-authored* non-nano-articles in given regions substituted as an explanatory variable instead of the total counts of non-nano-articles.[7] Until the Zucker–Darby articles-patent person matching project is completed, we cannot measure cross-institutional-type collaborations among inventors since their affiliation is available only from articles.

These results are similar to those displayed in Table 3, with the exception that cumulative counts of non-nano-patents also now have a significant positive impact on the rate of production of nano-articles, even in the presence of the NSF funding variable. When the knowledge flows are broken down by sector pairings, as in Table 6, the flows between firms and universities, firms and research institutes, firms and government, and universities and government all show robustly positive impacts on the rate of production both of articles and of patents. Relaxing the equality constraint on the coefficients in moving from Tables 5 and 6 again very significantly improves the explanatory power of these regressions.

However, the explanatory power of these regressions is in every case substantially lower than the corresponding regressions in Tables 3 and 4. Nonetheless, we are encouraged that even the crude flow variables we are able to measure at this time do as well as they do and are encouraged to pursue this concept in future research when data is available to measure collaborations for patents as well as articles.

[7] These counts, unlike those for general knowledge stocks, are not discounted as they represent institutions precedents, not depreciating knowledge.

Table 5
Knowledge flows across organization type boundaries poisson regressions with random effects[a] for U.S. regions, 1981–2004

Explanatory variables for region and year	Nano-scale article and patent			
	Nano-article	Nano-article	Nano-patent	Nano-patent
Non-nano-patents knowledge stock	0.046*** (0.001)	0.021*** (0.001)	0.077*** (0.001)	0.060*** (0.002)
Non-nano-articles knowledge flow	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
NSF nano-funding stock		0.541*** (0.008)		0.348*** (0.013)
Constant	2.684*** (0.130)	2.633*** (0.129)	1.590*** (0.119)	1.576*** (0.119)
Log likelihood	−40540.6***	−38108.5***	−18445.6***	−18113.8***
Wald log-likelihood test for NSF nano-funding stock[b]	–	4864.2***	–	663.6***

Notes—Dependent variables: nano-articles published in year and nano-patents granted in year. Robust standard errors in parentheses below coefficient estimates. $N = 4296$. Significance levels: ˆ0.10, *0.05, **0.01 and ***0.001. Knowledge flows are computed as the number of articles co-authored across organization types within one BEA.

[a] Group variable for random effects: region (BEA number), 179 groups.

[b] This Wald (1943) log-likelihood test is for the null hypothesis that adding the NSF nano-funding stock does not significantly improve the fit of the regression.

Table 6
knowledge flows across differentiated organization type boundaries Poisson regressions with random effects[a] for U.S. regions, 1981–2004

Explanatory variables for region and year	Dependent variable			
	Nano-articles	Nano-articles	Nano-patents	Nano-patents
Non-nano-patents knowledge stock	0.013*** (0.002)	−0.014*** (0.002)	0.109*** (0.003)	0.093*** (0.003)
Firm-university non-nano-articles knowledge flow	0.006*** (0.000)	0.005*** (0.000)	0.005*** (0.000)	0.004*** (0.000)
Firm-research institute non-nano-articles knowledge flow	0.009*** (0.001)	0.015*** (0.001)	0.011*** (0.001)	0.015*** (0.002)
Firm-government non-nano-articles knowledge flow	0.012*** (0.001)	0.013*** (0.001)	0.015*** (0.001)	0.014*** (0.001)
Firm-unclassified non-nano-articles knowledge flow	−0.016*** (0.000)	−0.012*** (0.000)	−0.009*** (0.001)	−0.005*** (0.001)
University-research institute non-nano-articles knowledge flow	−0.003*** (0.000)	−0.003*** (0.000)	−0.004*** (0.000)	−0.004*** (0.000)
University-government non-nano-articles knowledge flow	0.002*** (0.000)	0.002*** (0.000)	−0.005*** (0.000)	−0.006*** (0.000)
University-unclassified non-nano-articles knowledge flow	0.002*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)
Government-research institute non-nano-articles knowledge flow	−0.015*** (0.001)	−0.019*** (0.001)	−0.009*** (0.002)	−0.011*** (0.002)
Government-unclassified non-nano-articles knowledge flow	0.008*** (0.000)	0.006*** (0.000)	0.006*** (0.001)	0.006*** (0.001)
Research institute-unclassified non-nano-articles knowledge flow	−0.006*** (0.001)	−0.010*** (0.001)	−0.001 (0.001)	−0.005*** (0.001)
NSF nano-funding stock		0.530*** (0.008)		0.404*** (0.014)
Constant	2.583*** (0.128)	2.533*** (0.127)	1.417*** (0.115)	1.391*** (0.115)
Log likelihood	−37231.3***	−35133.2***	−16821.1***	−16424.3***
Wald log-likelihood test for NSF nano-funding stock[b]	–	4196.2***	–	793.6***
Log-likelihood test vs. Table 5[c]	6618.6***	5950.6***	3249.0***	3379.0***

Notes—Dependent variables: nano-articles published in year and nano-patents granted in year. Robust standard errors in parentheses below coefficient estimates. $N = 4296$. Significance levels: ˆ0.10, *0.05, **0.01 and ***0.001. Knowledge flows are computed as the number of articles co-authored across organization types within one BEA.

[a] Group variable for random effects: region (BEA number), 179 groups.

[b] This Wald (1943) log-likelihood test is for the null hypothesis that adding the NSF nano-funding stock does not significantly improve the fit of the regression.

[c] This Wald (1943) log-likelihood test is for the null hypothesis that constraining all the articles-knowledge-flow coefficients to be the same (as is implicit in Table 5) does not significantly reduce the goodness of fit of the regression.

11. Conclusions

In this paper, we have described our investigations of the impact of non-nano-knowledge stocks and knowledge flows on knowledge production in nanotechnology. The research question that we sought to answer in the course of these investigations was "What factors influence the rates at which new knowledge is produced in the field of nanotechnology?" We believe that this question is an important one, not simply because answering it improves our understanding of the processes by which new knowledge is produced in a rapidly evolving field, but because scientific knowledge is commonly assumed to be a strong source of economic growth. If we can identify conditions under which the rate of production of knowledge in a given field is observed to increase, we may be able to point to firm-, regional-, and national-level strategies and policies that will encourage the creation of those conditions, and that will ultimately promote economic growth.

The data analyses reported in this paper provide the following headline results:

(i) That the size of the cumulative knowledge stock of articles published in non-nanotechnological fields in a given geographical region has a significant positive effect on the rate of production of nanotechnological articles and patents in that region (see Table 3).
(ii) That the size of the cumulative knowledge stock of patents published in non-nanotechnological fields in a given geographical region has a significant positive effect on the rate of production of nanotechnological patents in that region (see Table 3).
(iii) That the volume of the cumulative knowledge flow of cross-institutionally co-authored articles published in non-nanotechnological fields in a given geographical region has a significant positive effect on the rate of production of nanotechnological articles and patents in that region (see Table 5).
(iv) That the size of the cumulative stock of funding dollars awarded by the NSF to nanotechnology projects initiated by institutions in a given geographical region has a significant positive effect on the rate of production of nanotechnological articles and patents in that region (see Tables 3 and 5).

These results allow us to draw two main conclusions. In the first place, the data provide further support for the general claim that scientific knowledge is strongly cumulative. We suggest that the Kuhnian critique of the cumulative advantage model (summarized in Section 2) is valid only under more limited conditions than are examined here. In the second place, the data supply evidence that ongoing efforts to institutionalize channels for cross-institutional collaboration are worthy of renewed support at organizational, regional, and national levels.

We mentioned in Section 1 our conviction that knowledge production is deeply embedded in a network of social structures and practices. Such structures and practices both constrain action and enable it (see, e.g., Granovetter, 1985). We have found empirically that the production of nanotechnological knowledge is embedded in the wider social context of institutional organization, cross-institutional collaboration, and national structures of incentives and rewards. This embeddedness is constraining, in that the range of possible action is narrowed, but also enabling, as channels for the flow of tacit knowledge deepen, and the flow between organizations of different types becomes more differentiated.

We predict that a productive seam of future data will be found at the level of the individual scientist and/or institution, which is the source both of resistance to, and of support for, nanotechnological knowledge flow. Resistance comes from the large number of scientists trained in non-nanotechnological areas, who face a daunting choice: either continuing to practice as they have in the past and risking devaluation in the labor market, or making the effort to acquire new knowledge and skills, typically at substantial cost. Change comes from a smaller number of scientists who take the positive decision to learn new areas of science (e.g., biology in addition to engineering) and/or to move to a new location to improve their access to new knowledge. Providing a precise specification of the mechanisms by which individual decision-making and localized social context together influence the growth of knowledge in particular fields remains a formidable research challenge.

References

Adams, J.D., 1990. Fundamental stocks of knowledge and productivity growth. Journal of Political Economy 98, 673–702.
Adams, J.D., Jaffe, A.B., 1996. Bounding the effects of R&D: an investigation using matched establishment-firm data. Rand Journal of Economics 27, 700–721.
Antonelli, C., 2001. The Microeconomics of Technological Systems. Oxford University Press, Oxford.
Arrow, K.J., 1962. Economic welfare and the allocation of resources for inventions. In: Nelson, R.R. (Ed.), The Rate and Direction of Economic Activity: Economic and Social Factors. Princeton University Press, Princeton, NJ, pp. 609–626.

Audretsch, D., Feldman, M.P., 1996. R&D spillovers and the geography of innovation and production. American Economic Review 86, 630–640.

Darby, M.R., Zucker, L.G., 2001. Change or die: The adoption of biotechnology in the Japanese and U.S. pharmaceutical industries. Comparative Studies of Technological Evolution 7, 85–125.

DiMaggio, P., 1982a. Cultural entrepreneurship in nineteenth century Boston. I. The creation of an organizational base for high culture in America. Media, Culture and Society 4, 33–50.

DiMaggio, P., 1982b. Cultural entrepreneurship in nineteenth century Boston, II: The classification and framing of American Art. Media, Culture and Society 4, 303–322.

Eccles, R.G., Crane, D.B., 1988. Doing Deals: Investment Banks at Work. Harvard Business School Press, Boston, MA.

Efthimiadis, E.N., 1996. Query expansion. In: Williams, M.E. (Ed.), Annual Review of Information Science and Technology, vol. 31. Learned Information, Medford, NJ, pp. 121–187.

Foray, D., 2004. The Economics of Knowledge. MIT Press, Cambridge, MA.

Gartrell, C.D., 1987. Network approaches to social evaluation. Annual Review of Sociology 13, 49–66.

Granovetter, M., 1985. Economic action, social structure, and embeddedness. American Journal of Sociology 91, 481–510.

Griliches, Z, 1990. Patent statistics as economic indicators: a survey. Journal of Economic Literature 28, 1661–1707.

Gulati, R., 1999. Network location and learning: the influence of network resources and firm capabilities on alliance formation. Strategic Management Journal 20, 397–420.

Henderson, R., Jaffe, A.B., Trajtenberg, M., 1998. Universities as a source of commercial technology: a detailed analysis of university patenting, 1965–1988. Review of Economics and Statistics 80, 119–127.

Hullmann, A., Meyer, M., 2003. Publications and patents in nanotechnology: an overview of previous studies and the state of the art. Scientometrics 58, 507–527.

Jaffe, A.B., 1986. Technological opportunity and spillovers of R&D: evidence from firms' patents, profits, and market value. American Economic Review 76, 984–1001.

Jaffe, A.B., 1989a. Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. Research Policy 18, 87–97.

Jaffe, A.B., 1989b. Real effects of academic research. American Economic Review 79, 957–970.

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

Johnson, K.P., Kort, J.R., 2004. 2004 redefinition of the BEA economic areas. Survey of Current Business, 68–75.

Kennedy, P., 1998. A Guide to Econometrics, 4th ed. MIT Press, Cambridge, MA.

Kuhn, T.S., 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago, IL.

Machlup, F., 1984. Knowledge: Its Creation, Distribution, and Economic Significance, 3. Princeton University Press, Princeton, NJ.

Moed, H.F., Glänzel, W., Schmoch, U. (Eds.), 2004. Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems. Kluwer Academic Publishers, Dordrecht, Netherlands.

Powell, W.W., 1990. Neither market nor hierarchy: network forms of organization. In: Staw, B.B., Cummings, L.L. (Eds.), Research in Organizational Behavior, 12. JAI Press, Greenwich, CT.

Sebastiani, F., 2002. Machine learning in automated text categorization. ACM Computing Surveys 34, 1–47.

Stephan, P.E., 1996. The economics of science. Journal of Economic Literature 34, 1199–1235.

Stigler, G.J., 1983. The process and progress of economics. Journal of Political Economy 91, 529–545.

Tyree, A., Semyonov, M., Hodge, R.W., 1979. Gaps and glissandos: inequality, economic development, and social mobility in 24 countries. American Sociological Review 44, 410–424.

Van Rijsbergen, C.J., 1979. Information Retrieval, 2nd ed. Butterworths, London.

Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society 54, 426–482.

Zucker, L.G., 1977. The role of institutionalization in cultural persistence. American Sociological Review 42, 726–743.

Zucker, L.G., 1991. Markets for bureaucratic authority and control: information quality in professions and services. In: Barley, S., Tolbert, P.S. (Eds.), Research in the Sociology of Organizations, vol. 8. JAI Press, Greenwich, CT, pp. 157–189.

Zucker, L.G., Darby, M.R., 2001. Capturing technological opportunity via Japan's star scientists: evidence from Japanese firms' biotech patents and products. Journal of Technology Transfer 26, 37–58.

Zucker, L.G., Darby, M.R., 2006. Socio-economic impact of nanoscale science: Initial results and NanoBank. In: Roco, M.C., Bainbridge, W.S. (Eds.), Societal Implications of Nanotechnology. Springer Science + Business Media.

Zucker, L.G., Darby, M.R., Armstrong, J., 1998a. Geographically localized knowledge: Spillovers or markets? Economic Inquiry 36, 65–86.

Zucker, L.G., Darby, M.R., Armstrong, J., 2002. Commercializing knowledge: university science, knowledge capture, and firm performance in biotechnology. Management Science 48, 138–153.

Zucker, L.G., Darby, M.R., Brewer, M.B., 1998b. Intellectual human capital and the birth of U.S. Biotechnology Enterprises. American Economic Review 88, 290–306.

Zucker, L.G., Darby, M.R., Brewer, M.B., Peng, Y., 1996. Collaboration structure and information dilemmas in biotechnology: organizational boundaries as trust production. In: Kramer, R.M., Tyler, T.R. (Eds.), Trust in Organizations. Sage, Thousand Oaks, CA, pp. 90–113.

Zucker, L.G., Darby, M.R., Furner, J., Liu, R.C., Ma, H., 2006. Minerva unbound: knowledge stocks, knowledge flows and new knowledge production, National Bureau of Economic Research Working Paper No. 12669.