Econ. Innov. New Techn., 1997, Vol. 5, pp. 19–50
Reprints available directly from the publisher
Photocopying permitted by license only

UNIVERSITY VERSUS CORPORATE PATENTS: A WINDOW ON THE BASICNESS OF INVENTION

MANUEL TRAJTENBERG[a], REBECCA HENDERSON[b] and ADAM JAFFE[c,*]

[a]*Tel Aviv University and NBER;* [b]*Massachusetts Institute of Technology and NBER;* [c]*Brandeis University and NBER*

(Received 20 February 1995; In final form 3 May 1996)

This paper is an attempt to quantify key aspects of innovations, "basicness" and appropriability, and explore the linkages between them. We rely on detailed patent data, particularly on patent citations, thus awarding the proposed measures a very wide coverage. Relying on the prior that universities perform more basic research than corporations, we find that forward-looking measures of "importance" and "generality" capture aspects of the basicness of innovations. Similarly, measures of the degree of reliance on scientific sources, and of the closeness to the origins of innovational paths, appear to reflect the basicness of research. As measures of appropriability we use the fraction of citations coming from patents awarded to the same inventor, and in fact these measures are much higher for corporations than for universities. An examination of a small number of patents that are universally recognized as "basic" provides further support for these measures. We find also evidence of the existence of "technological trajectories".

KEY WORDS: Patents, citations, innovation, basic research
JEL Codes: O30 C81

1. INTRODUCTION

Progress in many areas of economics is often limited by the lack of empirical counterparts to the theoretical constructs that we believe to be important. This problem is particularly severe in the economics of technical change, where it is difficult to find good indicators even for such fundamental notions as the rate of invention or the value of innovations. Many widely used measures, such as simple patent counts or counts of expert identified innovations, are severely limited in that they cannot account for the enormous heterogeneity of research projects and outcomes that characterizes the R&D process.

---

*We gratefully acknowledge support from the National Science Foundation through grant SES91-10516, and from the Ameritech Foundation via the Ameritech Fellows program of the Center for Regional Economic Issues at Case-Western Reserve University.

Two sources of heterogeneity that occupy a prominent place in the economics of technical change are "basicness" and "appropriability". Basicness refers to fundamental features of innovations such as originality, closeness to science, breath, etc. that impinge on the incentives to engage in R&D and on the choice of research projects. Appropriability refers to the ability of inventors to reap the benefits from their own innovations. A great deal of our theoretical understanding of the innovation process rests on these two notions, and on conjectures about the links between them. Thus, although appropriability problems plague all forms of investment in R&D, they are thought to be more severe as we move from applied to more basic research (Arrow, 1962). This view underlies a widespread division of labor whereby public institutions such as universities perform most of the basic research, and private firms do the bulk of the development.

We focus in this paper on the construction of a set of measures that we believe capture key aspects of basicness and of appropriability. These measures are grounded in a view of technical change as a cumulative process, whereby each innovation builds on the body of knowledge that preceded it, and forms in turn a foundation for subsequent advances. We compute the measures using detailed information contained in patents, relying heavily on *citations* to other patents, since these citations provide good evidence of the links between an innovation and its technological "antecedents" and "descendants". We use matched samples of university and corporate patents to exploit our prior belief that university research is more basic than corporate research, and rely on the contrast between them in order to test the hypothesis that these measures are legitimate proxies for the fundamental attributes of innovations that we are after. In so doing we also explore the characteristics of university research and of university patented innovations, and examine the methodological underpinnings of the notion of "indicator" (or proxy variable).

The statistical analysis of these data lends ample support to the hypothesis that most of our measures indeed reflect aspects of basicness and appropriability. In particular, we find that measures of the overall importance of innovations, of generality of research outcomes, and of reliance on scientific sources discriminate well between more and less basic innovations. Additional evidence from a handful of fundamental innovations lends further support to these statistical results. Likewise, a measure of the proportion of follow-up technical developments performed by the organization responsible for the original innovation seems to capture aspects of appropriability. Since patent data cover the great majority of recorded inventions and have become available recently in easily-accessible computerized form, we believe that these measures hold the potential for becoming a standardized tool of wide applicability for research in the economics of technical change.

Section 2 explains in more detail the nature of the patent data used in this study. The measures themselves are motivated and described in section 3. In section 4 we lay out the methodological basis for testing the hypothesis that our measures capture aspects of basicness and appropriability. We discuss in section 5 the research design and data characteristics, and expand on the nature of university research. The statistical analysis and empirical results are in section 6. The last section contains concluding comments and suggestions for future research.

2. THE USE OF PATENT DATA

The measures of basicness and appropriability that we put forward rely exclusively on information contained in patents. We are thus tapping one of the richest sources of data on inventions (over 5 million patents have been granted so far by the US), and certainly the one with the widest coverage. We intend to exploit detailed information that appears on individual patents, and not just patent counts as has been common practice in much of the research in this area.[1]

A patent is a temporary monopoly awarded to inventors for the commercial use of a newly invented device. For a patent to be granted, the innovation must be non-trivial, meaning that it would not appear obvious to a skilled practitioner of the relevant technology, and it must be useful, meaning that it has potential commercial value. If a patent is granted, an extensive public document is created. The front page of a patent contains detailed information about the invention, the inventor, the assignee, and the technological antecedents of the invention, all of which can be accessed in computerized form. An item of particular importance for our purposes is the citations to previous patents. We believe that important aspects of basicness and of appropriability are embodied in the relationship between the innovation and its technological antecedents and descendants, and that patent citations, made and received, provide an effective means for identifying and tracing these relationships.

Patent citations serve an important legal function, since they delimit the scope of the property rights awarded by the patent. Thus, if patent 2 cites patent 1, it implies that patent 1 represents a piece of previously existing knowledge upon which patent 2 builds, and over which 2 cannot have a claim. The applicant has a legal duty to disclose any knowledge of the prior art, but the decision regarding which patents to cite ultimately rests with the patent examiner, who is supposed to be an expert in the area and hence to be able to identify relevant prior art that the applicant misses or conceals.[2] The framework for the examiner's search of previous innova-

---

[1] The use of patent counts as indicators of innovation has had varying degrees of success (see e.g. Griliches, 1984 and 1990; Jaffe, 1986). Related work has shown that patent citations contain information about the value of patents and the links among them (Carpenter et al, 1981; Carpenter and Narin, 1983; Trajtenberg, 1990a and 1990b; Jaffe, Trajtenberg and Henderson, 1993).

[2] Because of the role of the examiner and the legal significance of patent citations, there is reason to believe that patent citations are less likely to be contaminated by extraneous motives in the decision of what to cite than other bibliographic data such as citations in the

tions is the patent classification system, which currently consists of over 100,000 patent subclasses, aggregated into about 400 3-digit patent classes. The combination of citation data, detailed technological classification, and information about each inventor provides a unique mechanism for placing research and research results in their broader technological and economic context.

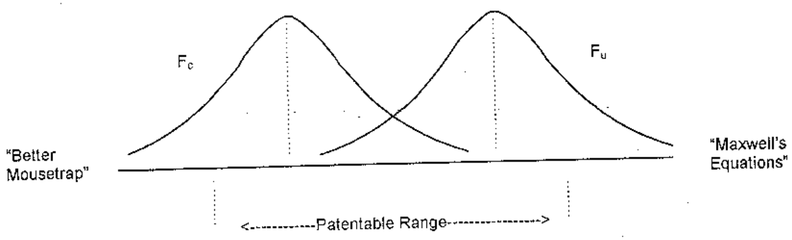
These data have, however, two important limitations: first, the range of *patentable* innovations constitutes just a sub-set of all research outcomes, and second, patenting is a *strategic* decision and hence not all *patentable* innovations are actually *patented*. As to the first limitation, consider Figure 1.a where we depict the "basicness" of research outcomes, ranging from the most applied on the left to the most basic on the right. Clearly, neither end of the continuum is patentable: Maxwell's equations could not be patented since they do not constitute a device (ideas cannot be patented); on the other hand, a marginally better mousetrap is not patentable either, because the innovation has to be non-trivial. Thus, our measures would not capture purely scientific advances devoid of immediate applicability, as well as run-of-the-mill technological improvements that are too trite to pass for discrete, codifiable innovations.

The second limitation is rooted in the fact that it may be optimal for inventors *not* to apply for patents even though their innovations would satisfy the criteria for patentability. For example, until 1980 universities could not collect royalties for the use of patents derived from federally funded research. This limitation greatly reduced the incentive to patent results from such research, which constitutes about 90% of all university research. Firms, on the other hand, may elect not to patent and rely instead on secrecy to protect their property rights (there is a large variance across industries in the reliance on patents versus secrecy: see Levin et al, 1987). Thus, patentability requirements and incentives to refrain from patenting limit the scope of measures built on patent data. It is widely believed that these limitations are not too severe, but that remains an open empirical issue.

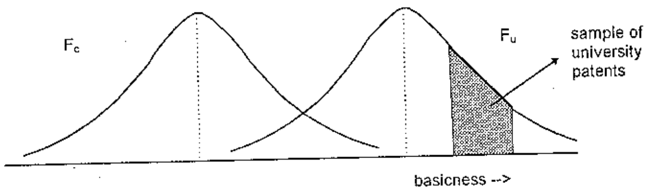
3. DEFINITION OF THE MEASURES

In searching for measurable aspects of basicness, we draw from the views of basic research expressed in the scientific and technological literature. Two parallel notions of basicness are found in this literature: one refers to the nature of research itself (that is, to the process leading to inventions), the other to the nature of research outcomes (Kuznets, 1962, made a similar distinction). Thus, research is regarded as basic if it focuses on scientific rather than on technological questions; if it seeks to elucidate general laws rather than solving particular technical problems; and if it addresses old puzzles with original methods (e.g. Kuhn, 1962; Rosenberg, 1982).[3] Research *outcomes* are held as basic if they have a major impact upon a given field, or a diffused but significant impact across a broad range of fields; if they are fundamental to much later work, and are often

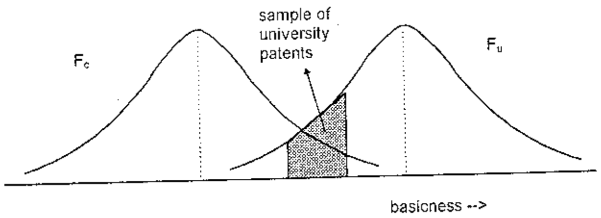
(a) The prior: Universities (F_u) vs. Corporations (F_c)



(b) Sample Selection Bias: Cov(D_i, ξ_i) > 0*



(c) Sample Selection Bias: Cov(D_i, ξ_i) < 0*



*D_i = 1 for Universities, D_i = 0 for Corporations

Figure 1. The Distribution of Basicness: Universities vs. Corporations

---

scientific literature (Van Raan, 1988; Weingart *et al*, 1988). Moreover, bibliometric data are of limited value in tracing the *economic* impact of scientific results, since they are not linked to economic agents or decisions.

---

[3] Thus for example, the research activities of the team headed by William Shockley at the Bell Labs that led to the discovery of the transistor can be seen as basic in this sense (Nelson, 1962).

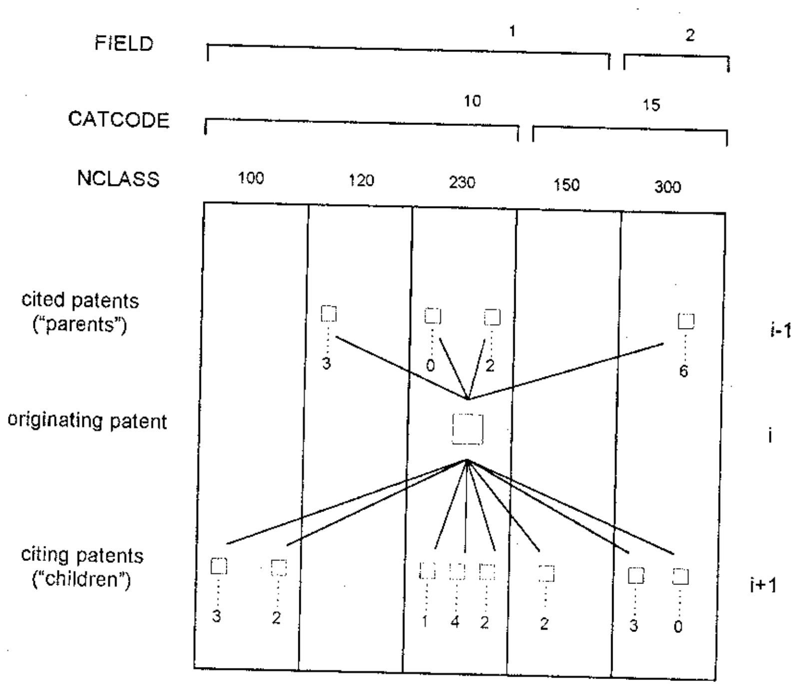
referred to and relied upon by scientists in the same or other fields (e.g. Watson and Crick's discovery of DNA).

We construct, accordingly, two sets of measures. "Backward-looking" measures (B/measures for short) are derived from the relationship between a given patent and the body of knowledge that preceded it (i.e. its antecedents). "Forward-looking" measures (F/measures for short) are derived from the relationship between a patent and subsequent technological developments that built upon it (i.e. its descendants). The presumption is that B/measures would be informative of the nature of research, whereas F/measures would be informative of the subsequent impact of research outcomes.

We use the patent citations made by each patent to identify its antecedents, and the subsequent patents that cite it to identify its descendants. For each of these patents we have information on their technological and temporal location (i.e. their patent class and date of application), the number of citations that they received, and the identity of the assignees. We can thus compute the number, technological diversity, and ownership pattern of patents corresponding to the antecedents and descendants of any patent, and the "distance" in time and technology space between these patents and the originating patent.[4] We can then use this information to construct measures of closeness to science, originality, subsequent impact, generality, etc. For the definition and computation of the measures we use the following notation (see figure 2):[5]

NCITING:	number of patents citing the originating patent ("o-patent").
NCITED:	number of patents cited by the o-patent.
NPCITES:	number of non-patent sources cited by the o-patent.
NCLASS:	3-digit original patent class.
CATCODE:	2-digit technological class (built by aggregating NCLASS).
FIELD:	1-digit classification by main technological fields.
LAG:	difference in years between the application date of a citing or cited patent, and the application date of the o-patent.

Index i corresponds to the patent under consideration, the originating or o-patent, i + 1 to citing patents, and i − 1 to cited patents. All measures but one (SCIENCE) will be defined and computed in equivalent ways backwards and forward; however, their precise meaning and interpretation



$$IMPORTF_i = 8 + 0.5 \cdot 17 = 16.5$$

$$IMPORTB_i = 4 + 0.5 \cdot 11 = 9.5$$

$$GENERAL = 1 - \left[\left(\frac{2}{8}\right)^2 + \left(\frac{3}{8}\right)^2 + \left(\frac{1}{8}\right)^2 + \left(\frac{2}{8}\right)^2\right] = \frac{23}{32}$$

$$ORIGINAL = 1 - \left[\left(\frac{1}{4}\right)^2 + \left(\frac{2}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = \frac{5}{8}$$

$$TECHF = \left[3 \cdot 0 + 2 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} + 2 \cdot 1\right] \Big/ 8 = \frac{5}{12}$$

$$TECHB = \left[2 \cdot 0 + 1 \cdot \frac{1}{3} + 1 \cdot 1\right] \Big/ 4 = \frac{1}{3}$$

Figure 2. Computing the Measures of Basicness: An Illustration

varies in some cases across the forward/backward (F/B) divide. The following table provides an overview of the forward looking measures (the B-measures are defined analogously):

---

[4] Patent information can also be used to characterize the distance in geographic space between an inventor and her descendants or antecedents. See Jaffe, Henderson and Trajtenberg, 1993.

[5] CATCODE is taken from Jaffe (1986). The technological areas in FIELD are, 1: Drugs and Medical; 2: Chemical (except Drugs); 3: Electronics, Optics and Nuclear; 4: Mechanical Arts; 5: Other.

1. Basicness Measures		
1.1 IMPORTF	number of citing patents, including second generation cites	$IMPORTF_i = NCITING_i + \lambda \sum_{j=1}^{nciting_i} NCITING_{i+1,j}$
1.2 GENERAL	Herfindahl index on technological classes of citing patents	$GENERAL_i = 1 - \sum_{k=1}^{N_i} (\frac{NCITING_{ik}}{NCITING_i})^2$
2. Distance Measures		
2.1 TECHF	distance in technology space	$TECHF_i = \sum_{j=1}^{nciting_i} \frac{TECH_j}{NCITING_i}$
2.2 TIMEF	average citation lag	$TIMEF_i = \sum_{j=1}^{nciting_i} \frac{LAG_j}{NCITING_i}$
3. Appropriability		
3.1 PSELF	number of self-citations	

3.1 Forward-Looking Measures

The first, and probably the key aspect of the relationship between a patent and its descendants is what we call the overall "importance" of a patent, denoted IMPORTF (the F for forward). This measure is designed to capture the technological impact of an invention as reflected in the number and importance of its descendants, and hence corresponds to the most intuitively appealing notion of basic innovations. In the words of Kuznets (1962),

> "Some inventions, representing as they do a breakthrough in a major field, have a wide technical potential in the sense that they **provide a base for numerous subsequent technical changes** [our emphasis]...the first steam engine, which initiated a whole series of major technical changes and applications...is vastly different from the invention of the safety match or the pocket lighter. This wide range is for our purposes the major characteristic relevant to the problem of measurement". (p. 26).

Thinking of citations to a patent as coming from follow-up advances that at least in part build upon or stem from the originating patent, we would like IMPORTF to reflect both the number of subsequent citations, and *their* respective importance. Thus we define (see figure 2),[6]

$$IMPORTF_i = NCITING_i + \lambda \sum_{j=1}^{nciting_i} NCITING_{i+1,j}$$

---

[6] Notice that the unavoidable truncation of the citation data at the point in time when the data are collected (T) means that IMPORTF is a lower bound, and should be taken to mean the "importance" of patent i *as revealed—or realized—up to T.*

where $0 < \lambda < 1$ is an arbitrary "discount factor" that is meant to downweight the "second-generation" descendants of a patent relative to the first-generation citing patents. We introduce discounting to alleviate the thorny problem of attribution: suppose that patent A is cited just by patent B, but the latter is cited by many patents. Without discounting, IMPORTF for patent A will be larger than for patent B, but intuition says that patent B is the one that had the largest *direct* impact. In all calculations reported here we have set $\lambda = 0.5$, but none of the results appear to depend upon this choice (we experimented with values of λ in the 0.25–0.75 range). We present in the appendix the two patents with the highest values of IMPORTF in our sample: the first discloses an important innovation in the manufacturing of semiconductor devices, the second in fiber optics sensors and transducers (two cutting-edge technologies). We show for each the titles of a sample of citing patents, thus illustrating the notion that these subsequent patents constitute follow-up developments that build upon the originating patent.

The second measure of F/basicness is "generality" (GENERAL), that is, the extent to which the follow-up technical advances are spread across different technological fields, rather than being concentrated in just a few of them.[7] We compute GENERAL on the basis of the Herfindahl index of concentration, whereby the number of citations in each 3-digit patent class (NCLASS) plays the same role as the sales of each firm in the traditional industrial organization context, that is,

$$GENERAL_i = 1 - \sum_{k=1}^{N_i} \left(\frac{NCITING_{ik}}{NCITING_i}\right)^2$$

where k is the index of patent classes, and N_i the number of different classes to which the citing patents belong. Notice that $0 \leq$ GENERAL ≤ 1, and that higher values represent *less* concentration and hence more generality.

IMPORTF and GENERAL presumably capture important determinants of the *social* returns to innovations: those with many descendants, or with descendants that span a wide range of technical fields, are likely to have high social returns. For example, Trajtenberg (1990a) found that the social value of innovations in Computed Tomography (CT) Scanners is highly correlated with a citations-weighted count of patents in that field (see also section 6.3). On the other hand, high marks of IMPORTF and GENERAL do not necessarily imply high *private* returns, the key intervening variable being of course appropriability. Thus, innovations with high IMPORTF

---

[7] Thus for example, if a patent in solid-state physics is cited by later patents in chemistry, in superconductivity and in medical instrumentation, we would regard it as more general, and hence more basic, than a similar patent that received the same number of citations but all or most of them belong to the same field.

may yield low returns if the follow-up innovations are done by other firms, whereas low-IMPORTF innovations can be highly profitable if they land in a market niche well-protected from competition. High generality may well interfere with appropriation and hence reduce private returns, particularly for small firms, who may find it difficult to assemble the complementary assets necessary to exploit a highly diverse set of market opportunities (Nelson, 1959.)

We also compute measures of "distance" in time and technology space between the innovation and its antecedents, and between the innovation and its offsprings. The presumption is that remoteness in time and technology may be related to aspects of basicness and/or to the conditions of appropriability. The F/looking time distance measure is defined simply as the average forward LAG, that is,

$$TIMEF_i = \sum_{j=1}^{nciting_i} \frac{LAG_j}{NCITING_i}$$

The F/distance in technology space is computed as follows: if the citing patent is in the same 3-digit class (NCLASS) as the originating patent, then the distance between them, TECH, is set to zero; if they are in the same 2-digit class (CATCODE) but not in the same 3-digit class, then TECH = 0.33; if they are in the same 1-digit class (FIELD) and not in the same 2- or 3-digit class, then TECH = 0.66; if they are even in different 1-digit classes then TECH = 1. The average distance for o-patent i is then,[8]

$$TECHF_i = \sum_{j=1}^{nciting_i} \frac{TECH_j}{NCITING_i}$$

We hypothesize that TIMEF will be related to basicness if the technical difficulties encountered in the R&D process are commensurable with the degree of basicness: in that case more basic innovations would take longer to generate offspring. As to TECHF, if basicness implies a higher probability of serendipitous discoveries, and if these tend to occur in remote technological areas, then it is plausible that the *incidence* of far-removed follow-ups would be higher the more basic a patent is. We also hypothesize that the distance measures may be related to the difficulty of appropriating the returns from innovations: the more technically dissimilar are the subsequent developments of an invention, and the more time that passes before they come to be, the more likely it is that other firms will overcome any advantages held by the original inventor and thereby take away a larger share of the economic returns (a similar reasoning applies to the equivalent B/measures, TIMEB and TECHB).

Our final F/looking measure relates to the ownership structure of the innovation's descendants. We propose a measure, PSELFF, that is defined simply as the percentage of citing patents issued to the same assignee as that of the originating patent. The rationale for this measure is that these subsequent patents are likely to reflect follow-up developments of the original invention, and that these developments are the conduit that leads to the appropriation of returns. Thus, the higher the proportion of these later developments that take place "in-house" the larger would be the fraction of the benefits captured by the original inventor.

3.2 Backward-Looking Measures

Turning now to the B/measures (that is, to the basicness of *research*), we define the equivalent to IMPORTF as,

$$IMPORTB_i = NCITED_i + \lambda \sum_{j=1}^{ncited_i} NCITING_{i-1,j}$$

Thus, IMPORTB will be large if the o-patent cites many previous patents, and these cited patents are "important" in the usual sense that they in turn were highly cited (as with IMPORTF we use = 0.5). In other words, IMPORTB reflects the extent to which a given o-patent stands on a wide base of previous innovations that are in themselves important. Our presumption is that more basic patents would have *fewer* and/or *less* important predecessors and therefore lower values of IMPORTB.

The equivalent B/measure for GENERAL, which we label ORIGINAL, is defined as,

$$ORIGINAL_i = 1 - \sum_{k=1}^{N_i} \left(\frac{NCITED_{ik}}{NCITED_i}\right)^2$$

Thus the larger is ORIGINAL the broader are the technological roots of the underlying research. Our notion is that synthesis of divergent ideas is characteristic of research that is highly original and hence basic in that sense. Finally, we define a measure of scientific base which lacks a F/counterpart,

$$SCIENCE_i = \frac{NPCITES_i}{NPCITES_i + NCITED_i}$$

---

[8] Notice that TECHF is related to GENERAL, but it is by no means the same: a patent whose descendants are in a number of different classes, all of which are close in technology space to the originating class, would have a high value of GENERAL but a low value of TECHF. Conversely, a patent that spawns a single rich line of subsequent development would have a low value of GENERAL, but could have a high value of TECHF if those developments are in a field far away from the originating patent.

that is, SCIENCE measures the predominance of scientific sources as proxied by NPCITES, over technological ones (embedded in NCITED). The non-patent references, which appear on the front page of patents under the heading "Other Publications", may include articles in scientific journals, books, abstracts, proceedings, etc. That is, they constitute prior scientific knowledge or ideas to which the patent is related. Our conjecture is that more basic research would tend to draw relatively more from scientific sources than from technology, and hence would be associated with higher values of SCIENCE.

The B/distance measures, TIMEB and TECHB, are defined in an analogous way to the F/distance measures, except that we substitute NCITED for NCITING in the corresponding formula. Their interpretation is straightforward: larger values of TIMEB indicate that the o-patent draws from older sources, large values of TECHB that the innovation has roots in remote technological fields. We define also an equivalent B/measure to PSELFF, PSELFB, which measures the extent to which the o-patent represents appropriation of benefits to its antecedents.

4. THE PROPOSED MEASURES AS INDICATORS OF BASICNESS: A STATISTICAL TEST

The measures put forward above are predicated on the assumption that they capture aspects of basicness and appropriability. The question is how to test this hypothesis relying for that purpose just on patent data, since it is extremely hard to find independent indicators of basicness that could help legitimize our measures (see, however, section 6.3). The test that we suggest here relies on the prior that university research is more basic than corporate research, and exploits the consequent contrast between university and corporate patents.

The attributes of innovation that we are trying to capture with our measures are inherently unobservable, as is education or ability in the context of labor economics, or permanent income in macro. In these latter cases, the starting point of the analysis is usually an equation of the form,

$$y = \beta x^* + \epsilon, \tag{1}$$

where y is observed but x^* is not, and ϵ is an i.i.d. disturbance (for example, y could be the wage rate and x^* education). Given the existence of a proxy variable x (e.g. years of schooling) such that,

$$x = x^* + v \tag{2}$$

where v is an i.i.d. measurement error, the issue is then how to obtain a consistent estimate for β, since if x were just used *in lieu* of x^* the resulting estimate will be biased and inconsistent. We start here a step earlier and ask, how can one establish the connection between a candidate proxy and x^* given that by definition no direct data exist on x^*? In other words, how can we test the hypothesis that x is indeed a proxy for x^* as described in (2)?[9] *A priori* reasoning may suffice in some cases (as in years of schooling as proxy for education), but in areas far removed from common experience that may not be so, and the area of innovation and patenting is certainly of that sort. Moreover, since many of the solutions to the errors in variables problem call for the use of multiple indicators, it is important to be able to assess in advance whether or not the various candidate variables qualify indeed as proxies.[10]

We propose to tackle the issue by resorting to additional information that takes the form of the following *prior*:[11] suppose we knew that there are two groups in the population from which x^* is drawn, S_1 and S_2, such that,

$$E(x^* | x^* \in S_1) > E(x^* | x^* \in S_2) \tag{3}$$

Consider the following hypothetical regression,

$$x_i^* = \alpha_0 + \alpha_1 D_i + \xi_i \tag{4}$$

where $D_i = 1$ if the ith observation belongs to S_1, and $D_i = 0$ otherwise. Clearly, if (3) is true then $\alpha_1 > 0$. Now take a candidate variable x; if it is indeed a proxy for x^* then,

$$x_i = \delta_0 + \delta_1 x_i^* + v_i, \qquad \delta_1 > 0 \tag{5}$$

How do we know that (5) holds? Estimate the following equation,

$$x_i = \zeta_0 + \zeta_1 D_i + w_i \tag{6}$$

and test H_0: $\zeta_1 > 0$. If $\delta_1 > 0$ then it must be that H_0 holds, since $\zeta_1 = \alpha_1 \delta_1$ and we know from (4) that $\alpha_1 > 0$. In words, given the presumed

---

[9] It is interesting to note that this question is rarely asked in economics, but rather (2) is taken for granted, and the problem is confined to the consistent estimation of (1) given the presence of a measurement error.

[10] Obviously, estimating a regression of y on x will not do, since if (2) holds then the estimate of β will be biased towards zero. Moreover, y and x may exhibit some spurious correlation even if x and x^* are uncorrelated, or the model in (1) might be misspecified in the sense that $cor(x, x^*) = 0$ but x belongs in the model in its own right.

[11] We use here "prior" and "maintained hypothesis" interchangeably: the point is that we bring in a statement (such as eq. 3), presumably well- grounded on external information, that we regard as factually true. All subsequent steps rely on this being so, and hence all down-the-line inferences should be regarded as conditional on the prior being true.

differences in the unobservable between the two groups in the population, a finding that x is on average larger for one of the groups (i.e. $\zeta_1 > 0$) implies that x is a proxy of x^* in the sense of (5). This is so provided that $\hat{\zeta}_1$ is unbiased, which in turn requires that $Cov(D_i, w_i) = 0$. Noticing that $w_i = \delta_1 \xi_i + v_i$, this latter condition implies that the measurement error is not systematically higher for one of the groups, meaning that $Cov(D_i, v_i) = 0$, and that there is no sample selection bias, i.e. that $Cov(D_i, \xi_i) = 0$.

4.1 Applying the Test: Selectivity Biases?

The much larger fraction of R&D devoted by universities to basic research (see below) suggests that university patents are more basic than corporate patents. The presumption is thus that university patents correspond to S_1 in (3) above, and corporate patents to S_2. Thus, testing for the statistical significance of differences in the means of the measures between universities and corporations (which is the same as testing for H_0: $\zeta_1 > 0$ in eq. 6) amounts to testing whether or not the measures are in fact indicators of the fundamental attributes of innovations that we are after.

The question is whether one can safely assume that $Cov(D_i, w_i) = 0$, and hence that $\hat{\zeta}_1$ is unbiased. It is hard to think of circumstances whereby $Cov(D_i, v_i) = 0$ would not hold, that is, that the measurement error of say, GENERAL as defined in section 3 (as a proxy for true "generality" of innovations) would be systematically larger for university patents; hence we disregard this possibility. On the other hand, selectivity biases of various types may cause $Cov(D_i, \xi_i) \neq 0$. Suppose for example that university patents were drawn from the upper half of the distribution of research outcomes, as shown in figure 1.b: that amounts to $Cov(D_i, \xi_i) > 0$, causing an upward bias in $\hat{\zeta}_1$, and hence invalidating the suggested test. Conversely, if university patents were drawn from the lower half the distribution as in figure 1.c, the resulting downward bias in $\hat{\zeta}_1$ would make it tougher to accept H_0: $\zeta_1 > 0$, but a finding to that effect would certainly be statistically valid.

Notice that in order for $Cov(D_i, \xi_i) > 0$, those research outcomes that are patented by universities would have to be on average more basic than the university average. This is very unlikely, considering that about 2/3 of university research is defined as basic, and hence is not oriented towards innovations with practical potential. Patents do have to have such potential, and hence it is much more likely that they reflect research outcomes that are, if anything, *less* basic than the university average. The other fact to keep in mind in this context is that, as shown in section 5, university patents do reflect roughly the distribution of research over broad fields, and hence in that sense at least there is no presumption of selectivity bias. In sum, there is good reason to believe that $Cov(D_i, \xi_i) \leq 0$, and hence that the proposed test is statistically sound.

5. RESEARCH DESIGN AND DATA CHARACTERISTICS

5.1 Sample Design and Data Gathering

The choice of our sample of patents was dictated primarily by the requirements of the test outlined above. In addition, we needed to cover sufficiently long sequences of innovations so as to be able to compute measures that rely on backward and forward linkages, and to control for technological areas, since citation practices may vary systematically across them. Thus, we took as our core sample *all* university patents applied for in 1975 (319 patents) and in 1980 (482 patents), which gave us substantial time horizons backwards and forward.[12] We identified and gathered data on each of the (earlier) patents cited by these originating patents, and on each of the (subsequent) patents citing them, thus forming a complete set which encompasses three successive generations of related inventions.[13] We also obtained the number of citations made and received by each of the cited and citing patents, which gave us some information about the "grandparents" and the "grandchildren" as well (see figure 3). Empirical work rarely examines such long stretches of the innovational stream.

We then identified two samples of corporate patents in parallel to the university patents. The first was drawn from the universe of patents granted to the top 200 R&D-performing U.S. firms in 1986, as reported in 10-K reports and coded by the Compustat data service. We expect that at least some of these firms perform appreciable basic research. The other corporate sample was drawn from the universe of patents assigned to all other U.S. corporations, which perform by definition less R&D, and presumably devote a significantly smaller share of their R&D budget to basic research.

In order to control for technological field, each of these samples was drawn to match the university patent cohorts by patent class, application year and grant date. That is, for each originating university patent, we selected a corporate patent from each universe that had the same application year and the same (3-digit) patent class as the university patent, and was granted as close in time as possible. This design allows us to compare averages of our measures across institutional groups, without worrying that the estimates might differ just because universities and corporations

---

[12] In principle there should always be enough of a backwards horizon, but in practice the availability of data declines dramatically as we go back in time, to the point that for the 1975 cohort, for example, a great deal of the data of the **cited** patents are missing, and that created serious problems in computing the measures. There is reason to believe though that this limitation will soon vanish, as more and more patent data become computerized and available as such.

[13] The data on citing patents extend only up to 1989, primarily because we had to rely on "third parties" to obtain the data, and that meant long delays. Again, availability is constantly improving, so that in future research one should be able to obtain much more recent data.

exhibit different distributions of patents across fields. We then collected data on the predecessors and successors of these corporate patents, exactly as we had for the set of university patents. As figure 3 makes clear, this sampling scheme lead to "explosive" data requirements: starting from just 319 university patents applied for in 1975 and 482 in 1980, we ended up collecting data on over 26,000 patents.

5.2 Characteristics of University Research and Patents[14]

Since the core sample is of university patents, it is important to describe the salient features of university R&D and university patents. The R&D performed by academic institutions in the US amounted to 17.2 billion dollars in 1991, which constituted 11.4% of total R&D expenditures in the US for that year (in 1970 the share was of 8.9%, and it has been rising steadily since). However, the role of universities in pushing the frontiers of science and technology go far beyond their share of total R&D, because of the nature of the research done at academic institutions: about 65% of it is defined as *basic* research, 30% as applied research, and just 5% as development (NSF, 1992). Thus, basic research performed at universities accounts for almost 50% of all basic research in the US, whereas academic R&D labeled as development accounts for less than 1% of all development.

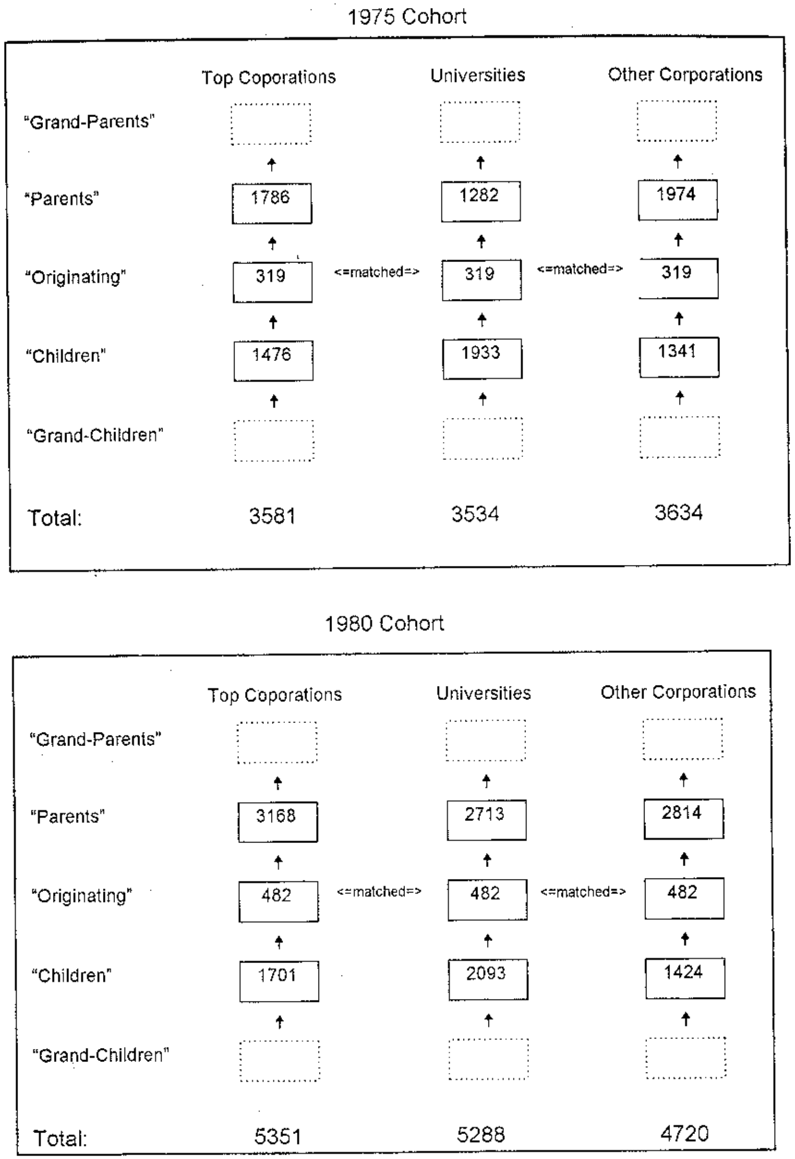
Given the nature of academic research, it is no surprise that most of the research outcomes from universities (at least those that are observable and quantifiable) take the form of publications in scientific journals, and only a few end up as patents.[15] The incentives of universities to take patents were further dampened in the past by a law that precluded them from charging royalties for patents stemming from federally funded research (which accounts for the bulk of academic research). The lifting of this legal restriction in 1980, and the proliferation of collaborative ventures between industry and academic institutions in recent years (biotechnology is a prime example), brought about a steady increase in the absolute and relative number of university patents. Still, university patents account for a very small fraction of all patents granted in the US (1.2% in 1990, or 2.4% of all patents of US *origin* that same year).

Not surprisingly, university patents do not constitute a representative sample of the universe of US patents: during the years examined here they were concentrated in a relatively small number of fields, and their distribution differed greatly from that of all US patents. On the other hand, the

---

[14] For a detailed account of university patenting over time see Henderson, Jaffe and Trajtenberg (1996).

[15] In addition to the fact that most of the university research is "basic" and hence largely non-patentable, the incentives that university researchers face (in terms of promotion, prestige, etc.) encourage primarily publication in the scientific literature, and not patent applications.



Total # of Patents for 1975: 10,749; for 1980: 15,359; Grand Total: 26108

Figure 3. Sample Design

technological mix of university patents do seem to reflect the distribution of academic R&D expenditures over broad scientific fields: at least 25–30% of university patents belong to patent classes related to the Biological and Medical Sciences, which commanded 45% of all academic R&D in 1980, and about 12% of patents belong to the Physical Sciences, to which universities allocated 11% of their R&D budget (not including engineering).

University patents are highly concentrated in the hands of relatively few academic institutions: of the 75 universities that were issued patents in 1975, half received just one patent, and the top ten received over 50% (similar figures apply for 1980). There seems to be also a positive link between the size of the university R&D budget and the number of patents received: 5 of the top 10 universities with the most patents in both 1975 and 1980 belong also to the top ten in terms of R&D expenditures in 1989.

6. STATISTICAL ANALYSIS AND RESULTS

6.1 A First Look at the Measures

Tables 1 and 2 present descriptive statistics of the measures and Pearson correlations between them, and figures 4 and 5 depict the empirical distributions of some of them. Notice first the striking similarity in the shape of the distributions of IMPORTF and IMPORTB, and likewise for GENERAL and ORIGINAL; as it turns out, this is true for *all* equivalent F/B measures, which is an interesting finding that deserves further scrutiny (see section 6.4).

As figure 4 shows, the distribution of IMPORTF is extremely skewed (for 1980 the mean is 12.6 and the median just 5.5). If we interpret IMPORTF as an indicator of the value of patents, the observed skewness would fit nicely with previous findings regarding the distribution of such values (e.g. Pakes 1986, Trajtenberg 1990): most patents turn out to be of very little value (i.e. to have few if any descendants), and only a handful make it big. IMPORTB is similarly skewed, which would mean that most patents come from "humble origins", and few have important technological predecessors. Thus important innovations appear to be in very short supply as one looks either up or down the innovational stream.

The distributions of GENERAL and of ORIGINAL are much closer to being bell-shaped, except for the large mass at zero. Sixty percent of the patents with GENERAL = 0 had just one citing patent, which means that GENERAL could only be zero; an additional 25% of these patents had NCITING = 2, which are very likely to render GENERAL = 0. Still, the mass at zero is not an artifact: patents that "fathered" just one or two further technological developments can claim indeed little generality.[16]

Turning to the correlations between F/measures and between B/measures, notice in table 2(i) that the various measures do capture different aspects of the underlying phenomena (the results for the 1975 sample are very similar and hence we show just those for 1980): none of the correlations exceeds 0.5, most are much smaller. The variables that exhibit the largest pairwise correlations are GENERAL and TECHF (and, in parallel, ORIGINAL and TECHB), which makes sense since both refer

Table 1. Descriptive Statistics*

***(i) Basicness Measures*:**

Variable	*N*	*Mean*	*Std Dev*	*Minimum*	*Maximum*
IMPORTF (1975)	948	12.58	23.21	0	380.50
(1980)	1446	6.96	14.68	0	250.50
GENERAL	806	0.32	0.29	0	0.88
	1127	0.27	0.27	0	0.85
IMPORTB	763	21.34	22.10	0	204.50
	1340	27.76	33.33	0	393.50
ORIGINAL	719	0.22	0.27	0	0.82
	1261	0.27	0.27	0	0.88
SCIENCE	945	0.14	0.25	0	1.00
	1446	0.20	0.30	0	1.00

***(ii) Distance Measures*:**

Variable	*N*	*Mean*	*Std Dev*	*Minimum*	*Maximum*
TECHF	791	0.32	0.30	0	1.00
	1124	0.32	0.31	0	1.00
TIMEF	811	7.26	2.49	0	13.00
	1126	4.38	1.55	0	9.00
TECHB	708	0.31	0.34	0	1.00
	1261	0.30	0.31	0	1.00
TIMEB	909	7.66	3.95	0	17.00
	1367	9.17	4.85	0	22.00

***(iii) Appropriability Measures*:**

Variable	*N*	*Mean*	*Std Dev*	*Minimum*	*Maximum*
PSELFF	783	0.11	0.24	0	1.00
	1061	0.16	0.29	0	1.00
PSELFB	719	0.14	0.30	0	1.00
	1342	0.13	0.25	0	1.00

*The top line of each variable corresponds to the 1975 sample, the bottom one to the 1980 sample.

---

[16] For ORIGINAL = 0 the percentage of patents with NCITED = 1 or NCITED = 2 was just 39% (25% for 1975), thus it would seem that the finding of a large mass at zero is more telling for ORIGINAL than for GENERAL. The only qualification is that there are many missing backwards data, and thus many of the patents for which NCITED = 1 or 2 might have received a value of zero for ORIGINAL (rather than missing) if the data had been available.

TABLE 2. Correlations Between Measures—1980 Sample*

(i) Correlations between Forward Measures	*IMPORTF*	*GENERAL*	*TECHF*	*TIMEF*	*PSELFF*
IMPORTF	1.0				
	0.0				
GENERAL	0.238	1.0			
	0.0001	0.0			
TECHF	0.003	**0.372**	1.0		
	0.927	0.0001	0.0		
TIMEF	−0.051	0.019	0.006	1.0	
	0.089	0.524	0.845	0.0	
PSELFF	−0.019	−0.074	−0.035	**−0.161**	1.0
	0.520	0.015	0.259	0.0001	0.0

(ii) Correlations Across F/B Measures	*IMPORTF*	*GENERAL*	*TECHF*	*TIMEF*	*PSELFF*
IMPORTB	0.255	0.115	−0.047	0.046	0.014
	0.0001	0.0002	0.129	0.132	0.646
ORIGINAL	0.055	**0.246**	0.203	−0.011	0.006
	0.051	0.0001	0.0001	0.721	0.852
TECHB	−0.020	0.187	**0.387**	−0.027	0.051
	0.475	0.0001	0.0001	0.381	0.114
TIMEB	−0.167	−0.114	0.005	0.090	−0.012
	0.0001	0.0002	0.878	0.003	0.692
PSELFB	−0.057	−0.050	−0.013	−0.046	**0.211**
	0.037	0.103	0.657	0.130	0.0001
SCIENCE	−0.005	0.052	0.038	−0.029	0.017
	0.838	0.082	0.204	0.333	0.569

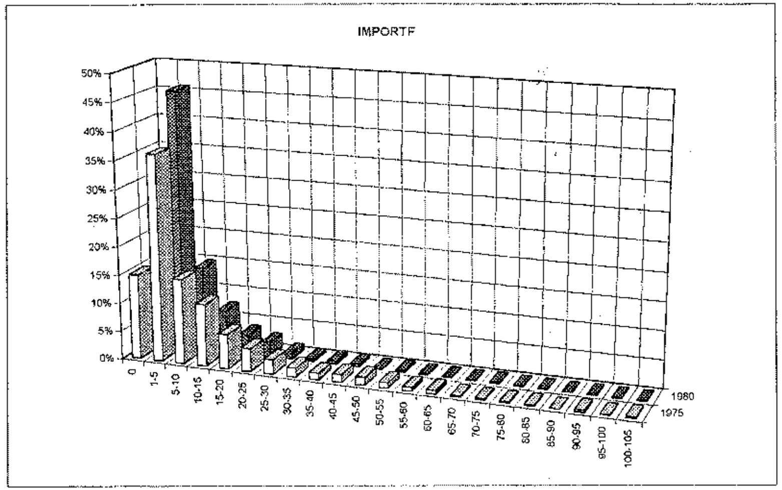
*Pearson correlation coefficients; significance probabilities right under.

to the positioning of patents in technology space. Another pair exhibiting a high correlation is IMPORTF and GENERAL, and likewise IMPORTB and ORIGINAL.

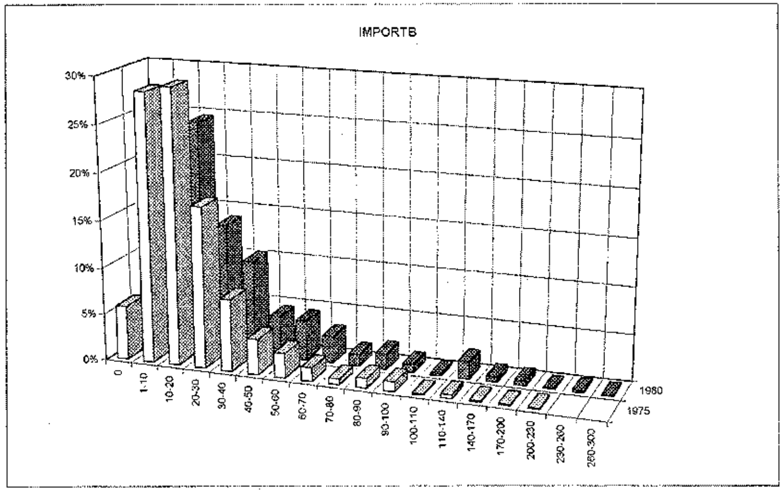
Notice also the large negative correlation between PSELFF and TIMEF, implying that spillovers tend to occur in-house faster than externally. Thus the same R&D organization may be able to recognize earlier the potential for further developments of a given innovation, it may have already in place the requisite competencies needed to develop the successor innovations, etc. For the converse reasons outsiders would take longer in benefiting from spillovers originating in labs other than their own.

6.2 Universities vs. Corporations: Comparing the Means

Table 3 presents the sample means of the proposed measures for university and corporate patents, and t-tests for the significance of the differences between them. To recall, our prior is that research, and research outcomes



Truncated-Values of IMPORTF>105 do not appear



Truncated-Values of IMPORTB>300 do not appear

Figure 4. Frequency Distribution of Basicness Measures

from universities are more basic than those from corporations, and therefore if our measures do capture aspects of basicness university patents should rank higher than corporate patents along those dimensions.

The results for IMPORTF strongly support this hypothesis: university patents received significantly more first- and second generation citations, and the difference seems to increase over time (see the results for 1975 vis a vis 1980). The figures for GENERAL indicate that the follow-up innova-

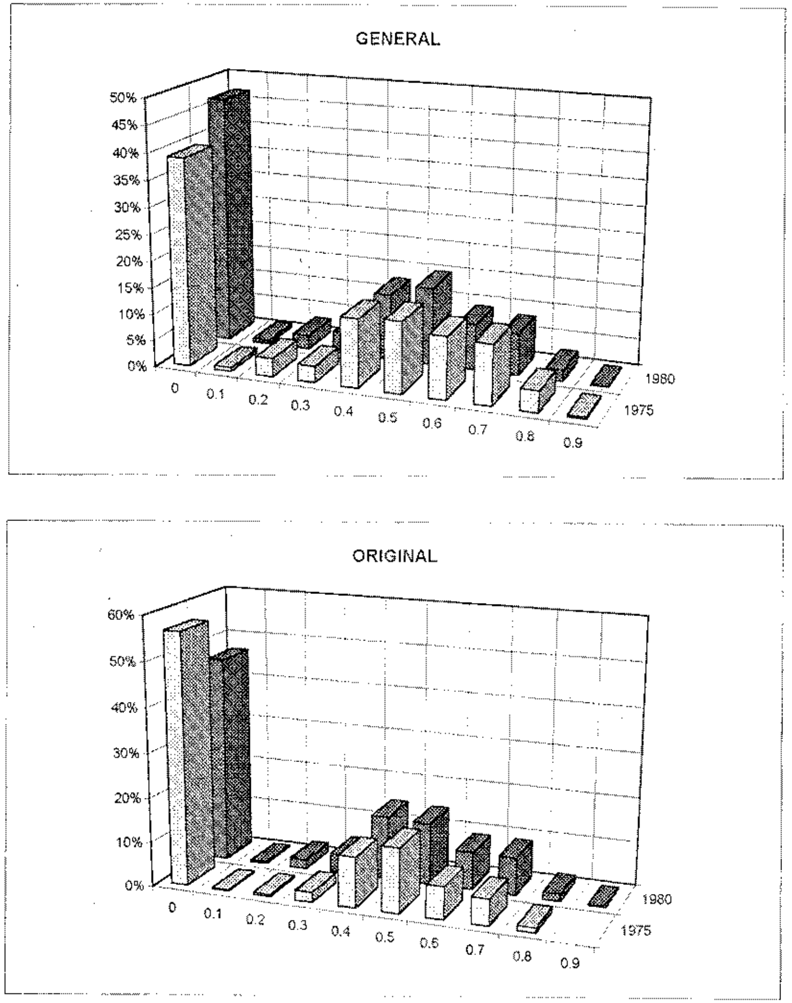


Figure 5. Frequency Distribution of Basicness Measures

tions from universities spread indeed more widely over different technological areas, but the 1975 results suggest that these differences tend to narrow down over time. In fact, we recomputed GENERAL for the 1975 cohort truncating the citation data in 1984 so as to replicate the time span of the 1980 sample, and the results are almost identical to those for 1980. Thus the offsprings from corporate patents tend to be more technologically concentrated in the short term, but eventually these spillovers become more diffused, narrowing the gap between them and university patents. This fits the common wisdom regarding the way spillovers work, and goes in tandem with the finding in our companion paper (Jaffe et al 1993) of diminishing geographic localization of spillovers over time.

TABLE 3. Comparison of Means: Universities vs. Corporations[a]

(i) Basicness Measures: *Variable*	*Universities*	*Corporations*
IMPORTF (1975)	8.80	6.04***
(1980)	16.76	10.49***
TIMEF	0.31	0.25***
	0.34	0.31[b]
IMPORTB	26.83	28.22
	16.31	23.75***
ORIGINAL	0.28	0.27
	0.20	0.24[#]
SCIENCE	0.28	0.16***
	0.20	0.10***
(ii) Distance Measures: *Variable*	*Universities*	*Corporations*
TECHF	0.35	0.30***
	0.32	0.31
TECHB	0.33	0.29**
	0.30	0.32[#]
TIMEF	4.44	4.34
	6.48	7.69[#]
TIMEB	9.08	9.22[#]
	7.50	7.73[#]
(iii) Appropriability Measures: *Variable*	*Universities*	*Corporations*
PSELFF	0.09	0.19***
	0.07	0.13***
PSELFB	0.06	0.16***
	0.14	0.14

[a]The top row of each variable corresponds to 1980, the bottom to 1975. [b]The difference is significant for a "truncated" sample—see text. **,***differences from the mean of university patents statistically significant at the .05 and .01 level, respectively. [#]figures for which the differences from universities have the "wrong" sign (i.e. contrary to the prior).

As conjectured, IMPORTB shows that corporate innovations rely on more numerous and more "important" predecessors than universities (the differences though are significant for 1975 but not for 1980). Thus university research appears to be located nearer the origins of innovational paths, supporting the notion that this may be an aspect of basicness. ORIGINAL does not live up to expectations: there is a slight difference for 1980 but it lacks statistical significance, and for 1975 the difference even has the "wrong" sign. The measure SCIENCE does support the hypothesis, and strongly so: university patents do rely relatively more on non-patent (i.e. scientific) sources than corporate patents (recall that this is cannot be a "field effect", since the samples are matched by technology field).[17]

The results for the distance measures are weak and inconclusive: the measure of technological distance is significantly different in 1980 but not in 1975, whereas in most cases the results for TIME (F and B) run contrary to the prior, that is, the follow-up innovations of corporations appear to take *longer* than those of universities. In sharp contrast to them, the appropriability measures do perform very well: PSELF (F and B) is much larger for corporations than for universities, suggesting that these measures may indeed be indicative of the extent to which inventors succeed in reaping the benefits of their own research.

In sum, with the exception of ORIGINAL, and with some reservation regarding the time span of GENERAL, the contrast between universities and corporations suggest that our basicness measures qualify as legitimate proxies for some fundamental features of innovations. On the other hand, the conceptual ambiguity of the distance measures surfaces also in the empirical results: there is some evidence that the offsprings from university innovations may be more remote in technology space, but certainly not in time. In general F/measures perform somewhat better than B/measures, and the 1980 sample shows crisper results than 1975.

A subsidiary hypothesis was that the basicness measures would exhibit higher values for patents of top corporations (i.e. the 200 corporations with the largest R&D outlays) than for those of other corporations. Presumably the larger corporations can afford to invest more in basic research, since they may be able to appropriate a larger fraction of the benefits. The evidence does not support this conjecture: with the notable exception of PSELF (which turned out to be significantly larger for top corporations), and to a lesser extent IMPORTF, the differences between top and other corporations were not statistically significant.

However, it is quite likely that the lack of contrast between them stems simply from the particular sample chosen, which is by no means representative of corporate research, but rather replicates exactly the composition of university patents. Thus, it may well be that in those particular fields there is little difference between small and large R&D performers, but that those differences do exist in the population of corporate innovations at large. In fact, a large proportion of our sample is in biotechnology, and we know that in this field small firms, both by

themselves and in cooperation with universities, are particularly innovative. That may also explain the fact that we observe a significantly *lower* value of IMPORTB for the smaller corporations, which would imply that they engage in more original and creative research.

6.3 Additional Evidence

One of the limitations of the forgoing analysis is that it is entirely internal to patent data. As said above, though, it is extremely hard to obtain enough independent data on the basic features of innovations, that would allow us to perform direct statistical tests of the proposed measures. Short of that, we present here a few bits of additional evidence that may help illustrate the meaning of the proposed measures, and perhaps also provide further support to their validity as indicators of basicness.

We searched for a small number of innovations that fulfill the following conditions: (*i*) that would be universally recognized as "basic"; (*ii*) that are embedded in one, or in a small number of easily identified patent(s); and (*iii*) that are sufficiently recent so that the citations data would be fairly complete (since the computerized patent citations data starts only in 1975) .[18] We relied for that purpose primarily on Berry (1993), and on various publications of the Patent Office (such as OTAF, 1977). We retrieved for the handful of patents thus selected the corresponding citations data, and computed the main F-measures.[19] The patents selected were,

Innovation	Patent Number	Inventors	Assignee	App. year	Grant year
Recombinant DNA	4,237,224	S.N. Cohen and H.W. Boyer	Stanford University	1979	1980
CT Scanner	3,778,614	G. Hounsfield	EMI Ltd (UK)	1971	1973
Cardiac Pacemakers	3,833,005	Wingrove	Medtronics, Inc.	1971	1974
Fibre Optics	3,659,915	R. Maurer et al	Corning Glass Works (US)	1970	1972
	3,711,262	D.B. Keck et al	Corning Glass Works (US)	1970	1973
	3,737,293	R. Maurer	Corning Glass Works (US)	1972	1973

---

[17] It is possible though that this result reflects to some extent differences in citation *practices* between university and corporate researchers, and not just genuine differences in the *nature* of their research (i.e. more "scientific"). However, the fact that there are symbiotic linkages between universities and corporations in fields such as biotechnology (where most of the NPCITES occur) would suggest that citation practices are actually similar; unfortunately our data cannot discriminate between these effects.

[18] There probably are a large number of patents that conform to these criteria, but it would take a major research effort to identify them, and to make sure that there is indeed consensus about their being "basic" (for example by interviewing experts in the field). The patents selected here are not meant to be a "representative sample" (since it is not clear what the universe is), but just illustrative cases.

[19] We could not obtain data on the *cited* patents, and hence could not compute B-measures.

Innovation	Patent Number	Inventors	Assignee	App. year	Grant year
Recombinant DNA	4,237,224	S.N. Cohen and H.W. Boyer	Stanford University	1979	1980
Ruby Laser	3,353,115	T. Maiman	Hughes Aircraft Co. (US)	1965	1967

The Cohen-Boyer patent is universally acknowledged as one of the fundamental innovations that paved the way for the emergence of biotechnology; CT (Computed Tomography) scanners revolutionized diagnostic medicine; cardiac pacemakers constituted a major improvement in the long-term treatment of cardiovascular disorders; the work by Maurer, Keck and their team at Corning Glass Works laid the foundations for the practical implementation of fibre optics; T. Maiman was the first to implement a ruby laser, which was a milestone in the development of lasers. This latter patent though was granted in 1967, and hence it does not fulfill the third condition stated above. Nevertheless, we present it here to illustrate the limitations of the data in the time dimension.

We computed for these patents the measures IMPORTF and GENERAL in two ways: (1) using all the citations data available at present (i.e. up to 1994), and (2) truncating the citations data in 1990, in order to make it comparable to the measures for the sample of university and corporate patents used in the analysis above; the latter are shown in parenthesis in the following table:

	Citations		**IMPORTF**		**GENERAL**
Mean 1975 sample		(5)		(13)	0.32
Recombinant DNA	130	(107)	545	(273)	0.26
CT Scanner	98	(95)	485	(380)	0.70
Cardiac Pacemakers	35	(29)	166	(118)	0.17
Fibre Optics	15	(13)	88	(60)	0.76
	32	(28)	199	(149)	0.59
	16	(16)	132	(114)	0.48
Ruby Laser	6	(4)	21	(14)	0.67

Except for the ruby laser, all these patents had a value of IMPORTF *5 to 10 times* higher than the mean for the 1975 sample of patents used in the previous analysis (using citations up to 1990). The fact that these innovations, which are surely "basic" by external criteria, show extremely high values of IMPORTF, can thus be seen as providing additional qualitative (or "anecdotal") support to the notion that this measure may indeed capture aspects of basicness. As to the ruby laser, it is remarkable that it received a value of IMPORTF similar to the average for the 1975 sample,

even though it was awarded 10 years earlier.[20] This suggests that this patent had indeed a strong impact, of which we can observe only the end tail.

The results for GENERAL are very much in line with expectations: for the CT scanner, fiber optics and the ruby laser, its value is much higher than the average for the sample. Indeed, the impact of those innovations was very broad in scope, reaching a wide range of technological fields. For the pacemaker the value of GENERAL was much lower than the average, implying that the spillovers from it were strictly "local". The low value for the rDNA patent means that, although this was an extremely important innovation, its impact was confined to Biotechnology, and did not spill over to other fields, at least not as those are identified by 3-digit patent classes in the current patent classification system.

6.4 The Links Between F- and B-Measures

We turn now to a preliminary examination of the linkages between backward and forward measures, that may throw light on issues related to the R&D process. In particular, we would like to examine whether or not the nature of the research efforts (as captured by the B-measures) affect the features of the resulting innovations. As a first step, we present in table 2(ii) the correlations across the F/B divide. Notice that the largest correlations occur along the main diagonal, that is, between equivalent F/B measures.[21] Thus it would seem that "importance breeds importance", originality breeds generality, coming from far away in technology space leads far away as well, etc. In that sense, then, the (*ex post*) characteristics of patented innovations appear to be related to the attributes of the research that lead to them.

Probing further into these links, we run regressions of the two F/measures of basicness, IMPORTF and GENERAL, on the B/measures, dummies for technological fields, and dummies for corporations (see table 4). The purpose was to estimate a sort of production function whereby the attributes of the patented innovations play the role of "outputs" and the features of the underlying research the role of "inputs," and to see whether the differences between universities and corporations remain there after controlling for the characteristics of research and technological fields.

In line with the findings of table 2(ii), the most significant coefficient in the regression of IMPORTF is its equivalent B/measure, IMPORTB, and likewise ORIGINAL is the most significant in the regression of

---

[20] We know from further work done on patent citations that the vast majority of citations are received during the first 5–10 years after a patent is issued.

[21] The one exception is TIMEB which shows a higher correlation with IMPORTF (and slightly higher with GENERAL) than with TIMEF; notice also that SCIENCE does not have an equivalent F/measure.

Table 4. Regressions of IMPORTF and GENERAL on B/variables*

	IMPORTF		*GENERAL*	
	1975	*1980*	*1975*	*1980*
Constant	19.7	8.2	0.3	0.2
	(5.7)	(5.1)	(6.8)	(5.9)
IMPORTB	0.3	0.1	0.0005	0.0005
	(6.7)	(8.8)	(0.9)	(1.9)
ORIGINAL	−4.0	−2.2	0.11	0.18
	(−1.0)	(−1.2)	(2.3)	(4.8)
TECHB	7.0	−0.04	0.17	0.11
	(2.3)	(−0.02)	(4.7)	(3.4)
TIMEB	−0.3	−0.4	−0.005	−0.005
	(−0.8)	(−3.9)	(−1.4)	(−2.2)
SCIENCE	−6.1	1.6	−0.04	0.14
	(−1.0)	(0.8)	(−0.5)	(3.7)
TC	−8.7	−3.1	−0.03	−0.05
	(−3.8)	(−3.0)	(−1.3)	(−2.4)
OC	−13.0	−3.0	−0.05	−0.05
	(−5.6)	(−2.9)	(−1.8)	(−2.3)
R^2	0.12	0.10	0.09	0.11
# obs.	707	1259	612	1002

*t-values in parentheses. All regressions include 4 dummies for technological "fields"

GENERAL. Notice however that more basic innovations should be associated with *smaller* values of IMPORTB, and hence if basicness leads to basicness we would have expected a *negative* sign on the coefficient of IMPORTB. The positive and highly significant coefficient that we obtained instead may be interpreted as follows: very basic research (in the sense of small values of IMPORTB) is likely to exhibit a large variance in terms of its outcomes—some may do well, others may fail badly (i.e. high and low values of IMPORTF). On the other hand, once a research avenue has proven its worth (i.e. high values of IMPORTB), further significant innovations along those lines are very likely to come, showing up in high IMPORTF. If this effect dominates, we will find indeed a positive association between IMPORTF and IMPORTB.

The results for GENERAL suggest that more original research, as well as research that draws from far removed technological areas (high TECHB), lead to innovations of wider technological applicability. More reliance on scientific sources also enhances the generality of the outcomes (this finding does not hold for the 1975 sample). The negative signs on TIMEB in both regressions imply that more important and more general innovations stem from more *recent* (or *up to date*) research sources. It is also clear that there does remain an "institutional effect" after controlling for the type of research, meaning that even if universities and corporations were to engage in research having similar characteristics, universities would still produce on average more basic innovations.

7. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

This paper is a fresh attempt to quantify various aspects of basicness and appropriability of innovations with the aid of detailed patent data, particularly patent citations. Relying on the prior that universities perform more basic research than corporations, we find that the forward-looking measures of importance and generality do seem to capture aspects of the basicness of *innovations*; similarly, the reliance on scientific versus technological sources, and (to a less extent) the closeness to the origins of innovational paths, appear to reflect aspects of the basicness of *research*. On the other hand our measure of originality does not seem to be able to discriminate between more and less basic research.

The fraction of citations coming from patents awarded to the same inventor was found to be much higher for corporations than for universities, supporting the notion that PSELF may indeed be indicative of actual appropriability. The measures of technological distance appear to be related to basicness but the evidence is not clear-cut, whereas distance in time does not fit our conjectures. In all, then, the F/measures of basicness and the indicators of actual appropriability passed the test by ample margins, the measures of B/basicness did somewhat less well, and most distance measures failed. We also find interesting similarities and high correlations between equivalent F and B measures, suggesting that there may be strong "family effects" in successive generations of patents. Further work along these lines would seek to identify and characterize different "technological trajectories," and relate them to conventional economic data.

Having provided initial support for our measures, we plan in future work to use them on a wide scale in tandem with other economic data, and see whether or not they make a difference. An immediate target would be to redo studies that have used simple patent counts as indicators of innovation, usually with disappointing results. In particular, we would like to re-examine the series of studies by Griliches and associates at the NBER (Griliches, 1981; Pakes, 1985; Cockburn and Griliches, 1988), which sought to identify the impact of R&D and of patent counts on the market value of Compustat firms. We hypothesize that if we were to use composite indicators based on our measures instead of simple patent counts, the impact on stock market value would be much more noticeable.

In particular, we expect that IMPORTF would have a very significant effect, and that it will improve even further when adjusting it with PSELFF, since what should influence the worth of the inventor is just the appropriate rents, not the total. Likewise, we expect that GENERAL would reduce the value of patents of small firms, but not of highly diversified corporations. Another hypothesis is that the B/measures of basicness would be more closely related to R&D expenditures than to indicators of performance such as market value. This type of research may pave the way for the wide- scale use of the proposed measures as key variables in empirical studies of innovation.

APPENDIX

Examples of Patents with High IMPORTF

(i) Patent 4,059,461; issued: 12/10/1975; Assignee: MIT
IMPORTF = 380; NCITING = 64
Title: Method for improving the crystallinity of semiconductor films by laser beam scanning and the products thereof.

Titles of sample of citing patents:

1. Process for producing coarse-grain crystalline/mono-crystalline metal and alloy films
2. Method of making Schottky barrier diode by selective beam-crystallized polycrystalline/amorphous layer
3. Semiconductor embedded layer technology including permeable base transistor, fabrication method
4. Polycrystalline semiconductor processing
5. Method for manufacturing a semiconductor device having regions of different thermal conductivity
6. Metal surface modification
7. Process for manufacturing a semiconductor device having a non-porous passivation layer
8. Method of fabricating display with semiconductor circuits on monolithic structure and flat panel display produced thereby
9. Method of fabricating semiconductor devices using laser annealing

(ii) Patent 4,071,753, issued: 31/03/1975; Assignee: GTE Laboratories Inc.
IMPORTF = 212.5; NCITING = 53
Title: Transducer for converting acoustic energy directly into optical energy [using optical fibers].

Titles of sample of citing patents:

1. Fiber optic magnetic sensors
2. Acoustic to optical pulse code modulating transducer
3. Method and sensor device for measuring a physical parameter utilizing an oscillatory, light modulation element

4. Alarm device with a condition sensor element
5. Fiber optic accelerometer
6. Fiber optics transducers for sensing parameter magnitude
7. Fiber optic hydrophone transducers
8. Coupled waveguide acousto-optic hydrophone
9. Fiber optics transducers for sensing parameter magnitude

References

Arrow, K.J.: "Economic Welfare and the Allocation of Resources for Inventions," in R. Nelson (ed.) **The Rate and Direction of Inventive Activity**, Princeton University Press, 1962.

Berry, Clifton F. Jr., **Inventing the Future**. Washington: Brassey's Inc., 1993.

Caballero, R. and Jaffe, A. "How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth." In O. Blanchard and S. Fisher, eds., **NBER Macroeconomic Annual**, Vol. 8, M. I. T. Press, 1993.

Carpenter, M.P., Narin F. and Wolf P., "Citation Rates to Technologically Important Patents." **World Patent Information**, Vol. 3. No. 4., 1981.

Carpenter, M. and F. Narin.: "Validation Study: Patent Citations as Indicators of Science and Foreign Dependence," **World Patent Information**, Vol. 5, No. 3, pp. 180–185, 1983.

Cockburn, I. and Z. Griliches, "Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents". **American Economic Review, Papers and Proceedings**, 1988 (78), pp. 419–423.

Griliches, Z., "Market Value, R&D and Patents". **Economic Letters** (1981) (7), pp. 183–87.

Griliches, Z. (ed.), **R&D, Patents and Productivity**. University of Chicago Press, 1984.

Griliches, Z.: "Productivity, R&D and Basic Research at the Firm Level in the 1970's." **American Economic Review**, 1986

Griliches, Z. "Patent Statistics as Economic Indicators: a Survey," **Journal of Economic Literature**, 1990, *27*, pp. 1661–1707.

Henderson, R., Jaffe, A. and Trajtenberg, M. "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting 1965–1988." In David, P. and E. Steinmueller (eds.), **A Productive Tension: University-Industry Research Collaboration in the Era of Knowledge-Based Economic Growth**. Stanford University Press, forthcoming 1996.

Jaffe, A.: "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value," **American Economic Review**, 1986, *76*(5), pp. 984–1001.

Jaffe, A., Trajtenberg, M. and Henderson, R. "Geographic Localization of Knowledge Spillovers, as Evidence by Patent Citations". **Quarterly Journal of Economics**, 1993, 108(3), pp. 577-598.

Kuhn, Thomas, **The Structure of Scientific Revolutions**. Chicago: University of Chicago Press, 1962.

Kuznets, S. "Inventive Activity: Problems of Definition and Measurement," in R. Nelson (ed.) **The Rate and Direction of Inventive Activity**, Princeton, NJ: Princeton University Press, 1962.

Levin, R., A. Klevorick, R.R. Nelson, and S.G. Winter, "Appropriating the Returns from Industrial Research and Development," **Brookings Papers on Economic Activity**, 1987, *3*, 783–820.

National Science Foundation, **Science and Technology Data Book**, Washington D.C. NSF 1992.

Nelson, R. "The Simple Economics of Basic Scientific Research," **Journal of Political Economy**, 1959

Nelson, R. "The link between science and invention: the case of the transistor," in R. Nelson (ed.) **The Rate and Direction of Inventive Activity**, Princeton University Press, 1962

Office of Technology Assessment and Forecast (OTAF), U.S. Department of Commerce, Patent and Trademark Office. **Seventh Report**. Washington D.C.: U.S. Government Printing Office, 1977.

Pakes, Ariel, "On Patents, R&D, and the Stock Market Rate of Return". **Journal of Political Economy**, 1985, *93*(*2*), pp. 390–409.

Pakes, Ariel, "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks", **Econometrica**, 1986, *54*(*4*), pp. 755–84.

Rosenberg, N. **Inside the Black Box: Technology and Economics**. Cambridge, England: Cambridge University Press, 1982.

Trajtenberg, M. "A Penny for Your Quotes: Patent Citations and the Value of Innovations," **Rand Journal of Economics**, 1990a, *21*(*1*), pp. 172–187.

Trajtenberg, M. **Economic Analysis of Product Innovation: The Case of CT Scanners**. Cambridge, MA: Harvard University Press, 1990b.

Van Raan (ed), **Handbook of Quantitative Studies of Science and Technology**, Amsterdam: North Holland, 1988.

Weingart, P. R. Sehringer and M. Winterhager. "Bibliometric Indicators for Assessing Strengths and Weaknesses of West German Science," Ch. 13 in Van Raan (ed), **Handbook of Quantitative Studies of Science and Technology**, Amsterdam: North Holland, 1988.

Table 3
Comparison of Means: Universities vs. Corporations[*]

(i) Basicness Measures:

Variable		Universities	Corporations
IMPORTF	(1980)	8.80	6.04^{***}
	(1975)	16.76	10.49^{***}
GENERAL		0.31	0.25^{***}
		0.34	0.31^{a}
IMPORTB		26.83	28.22
		16.31	23.75^{***}
ORIGINAL		0.28	0.27
		0.20	$0.24^{\#}$
SCIENCE		0.28	0.16^{***}
		0.20	0.10^{***}

(ii) Distance Measures:

Variable	Universities	Corporations
TECHF	0.35	0.30^{***}
	0.32	0.31
TECHB	0.33	0.29^{**}
	0.30	$0.32^{\#}$
TIMEF	4.44	4.34
	6.48	$7.69^{\#}$
TIMEB	9.08	$9.22^{\#}$
	7.50	$7.73^{\#}$

(iii) Appropriability Measures:

Variable	Universities	Corporations
PSELFF	0.09	0.19^{***}
	0.07	0.13^{***}
PSELFB	0.06	0.16^{***}
	0.14	0.14

---

[*] The top row of each variable corresponds to 1980, the bottom to 1975.

[a] The difference is significant for a "truncated" sample - see text.

[#] figures for which the differences from universities have the "wrong" sign (i.e. contrary to the prior).

[**], [***] differences from the mean of university patents statistically significant at the .05 and .01 level, respectively.