NBER WORKING PAPER SERIES

PATENT STATISTICS AS ECONOMIC INDICATORS: A SURVEY PART I

Zvi Griliches

Working Paper No. 3301

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1990

NBER Working Paper #3301 Part I
March 1990

PATENT STATISTICS AS ECONOMIC INDICATORS: A SURVEY PART I

ABSTRACT

This survey reviews the growing use of patent data in economic analysis. After describing some of the main characteristics of patents and patent data, it focuses on the use of patents as an indicator of technological change. Cross-sectional and time-series studies of the relationship of patents to R&D expenditures are reviewed, as well as scattered estimates of the distribution of patent values and the value of patent rights, the latter being based on recent analyses of European patent renewal data. Time-series trends of patents granted in the U.S. are examined and their decline in the 1970s is found to be an artifact of the budget stringencies at the Patent Office. The longer run downward trend in patents per R&D dollar is interpreted not as an indication of diminishing returns but rather as a reflection of the changing meaning of such data over time. The conclusion is reached that, in spite of many difficulties and reservations, patent data remain a unique resource for the study of technical change.

Zvi Griliches
Harvard University and
NBER
1050 Massachusetts Ave.
Cambridge, MA 02138

Overheard at a Catskills Resort
(one guest to another):

--The food is so terrible here.
--Yes. And the portions are so small.

1. Introduction

Patents and patent statistics have fascinated economists for a long time. Questions about sources of economic growth, the rate of technological change, the competitive position of different firms and countries, the dynamism of alternative industrial structures and arrangements all tend to revolve around notions of differential inventiveness: What has happened to the "underlying" rate of technical and scientific progress? How has it changed over time and across industries and national boundaries? We have, in fact, have almost no good measures on any of this and are thus reduced to pure speculation or to the use of various, only distantly related, "residual" measures and other proxies. In this desert of data patent statistics loom up as a mirage of wonderful plentitude and objectivity. They are available; they are by definition related to inventiveness, and they are based on what appears to be an objective and only slowly changing standard. No wonder that the idea that something interesting might be learned from such data tends to be rediscovered in each generation.

I shall try, in this survey, to show why I think patent statistics are interesting in spite of all the difficulties that arise in their use and interpretation. To do so I shall first describe the nature of patents and the types of data generated by their issuance, their current availability, and some of the major problems that arise when one tries to use them in economic analysis. I shall next review briefly some of the earlier work on this range of issues, focusing particularily on Jacob Schmookler's work and the questions raised by it. This will be followed by a review of the more modern, "computer age," work of the NBER group (Griliches, Hall, Hausman, Jaffe, Pakes, Schankerman and others), and I shall allude also to similar work of others, especially that of Scherer and the Yale group (Levin, Nelson, Klevorick, Winter, Reiss, Cohen, and others), and the SPRU group (Freeman, Pavitt, Soete and others). I will not be able, however, to do justice to all of this work (the work of others, of my collaborators, and even my own) but I hope to put up enough guideposts so that the interested reader can find his own way to and through this literature.[1]

Over all this work hovers the question "what can one use patent statistics for?" Can one use it to interpret longer term trends? If so, did inventiveness really decline in the 1930s and early 1940s, as indicated by such statistics, and again in the mid 1970s? Does the fact that large firms have alower patents per R&D dollar ratio imply diminishing returns to such investments? Can one use such numbers to conclude that demand forces are the stronger determining factors in the evolution of technological progress than supply factors, than the evolution of science, as Schmookler could be interpreted to say? These are the type of substantive questions that I will explore, though not necessarily answer, in this survey.

There is much that will not be covered in this survey. I will not discuss the literature that deals with the social value of the patent system and with alternative lengths of protection and licensing arrangements. Nor will I deal with the recent and rapidly growing theoretical literature on "patent races" and related game-theoretical topics. One has to draw the line somewhere and the task outlined above may be already too large for one article and one person to deal with. Nor will this be a fully "balanced" survey. I shall, per force, concentrate more on topics that I and my research associates have found most interesting, slighting thereby, sometimes unwittingly, some of the work of others in this field.[2]

2. Patents and Patent Statistics.

A patent is a document, issued by an authorized governmental agency, granting the right to exclude anyone else from the production or use of a specific new device, apparatus, or process for a stated number of years (17 in the U.S. currently). The grant is issued to the inventor of this device or process after an examination that focuses on both the novelty of the claimed item and its potential utility. The right embedded in the patent can be assigned by the inventor to somebody else, usually to his employer, a corporation, and/or sold to or licenced for use by somebody else. This right can be enforced only by the potential threat of or an actual suit in the courts for infringement damages. The stated purpose of the patent system is to encourage invention and technical progress both by providing a temporary monopoly for the inventor and by forcing the early disclosure of the information necessary for the production of this item or the operation of the new process.

The standard of novelty and utility imposed on the granting of such a right is not very high. (In this it probably does not differ greatly from the standards imposed in most fields on the publication of scientific journal articles.) In the U.S., for example, about 104,000 applications were filed in 1980 for ("utility") patents, of which about 65,000 were granted by the end of 1984; 1400 more were granted by the end of 1988, with another 300 or so to follow over the next 3 to 5 years. These numbers are typical. In the U.S. the granting success rate fluctuated around 65 percent in the 1970s. Roughly speaking, two out of three applications are eventually granted. The granting rate, the stringency of examination, varies greatly across countries and also somewhat over time. It has been over 90 percent in France (until the mid-1970s), about 80 percent in the U.K., and only 35 percent in Germany (Schankerman and Pakes 1986, Table 1), and has varied in the U.S from a low of 58 percent in 1965 to a high of 72 percent in 1967 (of domestic applications between 1965 and 1980). This variablity is, as I will show later, largely associated with differences in the procedures and resources of the various patent offices, implying, therefore, also differences in the average "quality" of a granted patent across countries and periods.

Of the approximately 62,000 patents granted in 1980, 24,000 or 39 percent were granted to foreign inventors, a ratio that has been rising sharply over the last decades, from 19 percent in the early 1960s to 48 percent in 1988. U.S. corporations have accounted for about 73 percent of the total patents granted to U.S. inventors (in 1988), with two percent being granted to agencies of the U.S. government, and the rest, 25 percent, going to individual inventors. The fraction accounted by foreign corporations of total foreign patenting in the U.S. has risen from 64 percent in the mid 1960s to 82 percent in 1988. The general trends in such numbers are depicted in Figures 1 and 2.

Even though grants can be thought of as a moving average of past applications, it can be seen in these figures that they tend to fluctuate as much or more than the number of patents applied for. It is also clear that economic conditions impinge on the rate at which patents are applied for. Applications were lower during the Great Depression and also during the World War II years, and their growth is retarded in the 1970s. Moreover, patents assigned to U.S. corporations have not grown anywhere near the rate of growth of total R&D expenditures in industry (and hence even less than the rate of growth in company financed R&D in industry). Since I will argue below that patents are a good index of inventive activity, a major aspect of which is also measured by R&D expenditures, this view will need reconciling with the aggregate facts depicted in Figure 2.

Data are also available at the firm level. In 1984 the largest patenters were General Electric, IBM, and Hitachi with 785, 608, and 596 patents granted respectively. Most of the major U.S. patenting firms experienced a declining trend in patents granted during the 1970s with some recovery in the 1980s, while there has been a rapid growth in U.S. patents granted to the major Japanese electronics and motor vehicles firms. (See Figure 5, Griliches 1989.)

What I have done in the above paragraphs is to discuss the information implicit in patent counts, in the number of patents issued at different times, in different countries, and to different types of inventors. This is the type of information that economists have largely focused on, also cross-classifying it by industry and firm, and it is the use of such numbers in economic analysis which will be the main topic of this survey. But a patent document, which is public after it has been granted, contains much more information than that. Besides information on the names of inventors and their addresses and the name of the organization to which the patent right may have been assigned, it also lists one or more patent classes to which it has been assigned by the examiners, cites a number of previous patents and sometimes also scientific articles to which this particular invention may be related, and also finally, but from the social point of view most importantly, provides a reasonably complete description of the invention covered by this particular patent. Thus, there is much more information derivable from the patent documents than just simply their aggregated number in a particular year or for a particular firm. One can study the geographic distribution of particular inventions, one can investigate citation networks and patterns, and one can actually read the detailed text of a series of patents in a particular field as raw material for an economic-technological history of it. Also, in a number of foreign countries, and in the U.S. since 1982, a non-negligable renewal fee, which rises with the age of the patent, has to be paid. This results in a significant abandonment of patents before their statutory expiration date and generates, in passing, a set of potentially very interesting patent mortality statistics.

In the U.S., aggregate patent statistics classified in a variety of ways are released by the Office of Documentation (formerly the Office of Technology Assessment and Forecast) at the U.S. Patent Office. Major series are published in the National Science Foundation's biannual Science Indicators compendium. More detailed tabulations are available from or can be prepared by the Patent Office and summary information on all recent patents is now also available on CD-ROM disks. The full text of the patents can be found in a number of depository libraries in the U.S. and can be now also accessed via several bibliographic computerized data base services, such as Dialog and BSR. Given the advanced search software available on these services it is possible to conduct a variety of specific searches of such data bases, looking for patents in a particular area or those mentioning a particular material, instrument, or a specific earlier patent, and tabulate the results at a reasonable cost. Patent data for other countries are being collected by the International Patent Documentation Center in Vienna, Austria and published annually in World Intellectual Property Annual. Country summaries are published in OECD, Main Science and Technology Indicators, and by various country statistical offices, such as Statistics, Canada. Current information on individual foreign patents is available on line from Dialog.

There are two major problems in using patents for economic analysis: classification and intrinsic variability. The first is primarily a technical problem. How does one allocate patent data organized by firms or by substantive patent classes into economically relevant industry or product groupings? I shall discuss it shortly. The second problem is fundamentally much harder and will be discussed at some length below. It refers to the obvious fact that patents differ greatly in their technical and economic significance. Many of them reflect minor improvements of little economic value. Some of them, however, prove extremely valuable. Unfortunately, we rarely know which are which and do not have yet a good procedure for "weighting" them appropriately. I shall discuss the available scraps of evidence on this topic in Section 5 of this survey.

Patents are awarded for an invention of a chemical formula, a mechanical device, or a process (procedure), and now even a computer program. The Patent Office classifies patents into many (300+ in the mid 1950s) classes and even many more subclasses (over 50,000), based on its need to ease the search for prior art. The resulting classification system is based primarily on technological and functional principles and is only rarely related to economists notions of products or well defined industries (which may be a mirage anyway). A subclass dealing with the dispensing of liquids contains both a patent for a water pistol and for a holy water dispenser. Another subclass relating to the dispensing of solids contains patents on both manure spreaders and toothpaste tubes. (Schmookler 1966, p.20) Nevertheless, with one notable exception (Scherer 1984) and the more recent Canadian data based studies, almost all attempts to relate patent numbers to industrial data use the subclass system as their basic unit of assignment.

Before any classification is attempted one has to face the inherent ambiguity of the task. Do we want to assign the invention to the industry in which it was made ("origin"), to the industry that is likely to produce it, or to the industry which will use the resulting product or process and whose productivity may benefit thereby (destination or industry of "use")? Consider, as an example, the case of a new plow invented in a chemical firm's research laboratory as part of its project on new combined fertilizer and tillage systems. It depends on what question is to be asked of the data. If we want to study the returns to R&D expenditures we may wish to count it in the chemical industry whence the money came to develop it. If we want to analyze the impact of technological change on the rate of investment, on the sale of new equipment, we may wish to count it in the farm equipment industry. If we are interested in its effects on measured productivity we are more likely to count it as being relevant to agriculture. This difference in questions reflects itself also in different classification strategies pursued by different researchers.

Schmookler, in his main work, chose to construct data on capital goods patents relevant to a particular industry by reviewing carefully a set of subclasses, sampling a number of patents in them, and deciding whether most of them were indeed likely to be used in the industry in question. He then aggregated the total number of patents in each of the accepted subclasses into an industry wide total. In this way he constructed time series for capital goods inventions of relevance for the railroad industry, the paper making industry, petroleum refining, and building construction. By focusing on capital goods inventions only and on a few selected and better defined industries, and by not insisting on completness or inclusivity, he made life quite a bit easier for himself. This choice forced him, however, to forego any serious analysis of consumer goods or manufacturing processes patenting. His industrial classification was based on the third type: the locus of potential use for the new or improved capital good.

In the mid 1970s the Patent Office established a research unit, the Office of Technology Assesment and Forecast. One of its first jobs, on a contract from the Science Indicators Unit of the National Science Foundation, was to try and produce patent statistics at the three and two-and-a-half SIC digit level, corresponding roughly to the NSF's classification of applied research and development by product field. This was done by developing a "concordance" between the patent class and sub-class classification and the Standard Industrial Classification (SIC). Where a subclass did not obviously belong into a single SIC industry, it was counted in all of the relevant ones, resulting in significant double counting. The industrial allocation was based primarily on the second notion of the relevant industrial classification: patents were allocated to the industries which were expected to produce the products designed by them or to use the new processes in the manufacture of their products. The new plow patent, in the previous example, would be assigned by the OTAF concordance to the farm equipment manufacturing industry.

The OTAF concordance was criticized early on both because of the arbitrariness in the assignment of some of the subclasses and the misleading inferences that could arise from the pervasive double counting (Scherer 1982a, Soete 1983).[3] The two most glaring examples of problems raised by such procedures was the appearance of significant and fast growing patenting by the Japanese in the aircraft industry, a rather surprising and mysterious development given the rather rudimentary state of the Japanese aircraft industry at that time. It turned out to be the result of allocating the "engines" patents category to both motor vehicles and aircraft. Almost all of the Japanese engine patents were automobile engine patents and since patenting in the engine category was high relative to other kinds of aircraft patents, it came to dominate the aircaraft patents category almost entirely. Another example was provided by the agricultural chemicals and drug industries where the assigned patents overlapped at the rate of 90 percent[!]. That is, only 10 percent of the patents counted in those industries were unique to them. It is doubtful whether such heavily overlapping data can be used in economic analyses which try to learn something about sources of technical progress by examining the contrasting experiences of different industries. The OTAF "industry" data contained too little independent data on the patenting history of actual industries.

As the result of such criticisms the 1985 version of these data has been improved by correcting some of the more obvious errors and by fractionalizing the allocation of dubious subclasses, reducing thereby their overall importance in the final totals. But most of the basic questions of classification remain still to be answered.

One way to get around some of these problems is to have the patent examiner assign the individual patent to one or several SIC industries, based on potential use. This is now being done in the Canadian patent system. One possibility, currently being pursued by Robert Evenson and his students, is to take a sample of U.S. patents also patented in Canada and to cross-tabulate the Canadian SIC assignments against the U.S. patent classification system, deriving thereby an empiricaly based and already naturally fractionized alternative concordance. (See Annex A of Englander, Evenson, and Hanazaki, 1988, Evenson et al, 1988, and Kortum and Putnam, 1989.)

An alternative approach, first pursued by Scherer (1965a and b) and more currently by the NBER group (see Bound et al, 1984), starts from patent totals for particular firms and then groups them into industries according to a firm's primary activity. This is an "origin" classification. It may be useful for the analysis of firm level data, relating patents to the R&D investments and the subsequent fortunes of the firms where they had been originally developed. But it is much less useful for the analysis of industrial data, both because of the conglomerateness of many of the large U.S. corporations and because particular patents may be having an impact far beyond the boundaries of their industry of "origin."

The extensive diversification of many firms and also the various merger waves create severe technical problems in trying to use the patent data even at the individual corporation level. What is noted on the patent is the name of the organization to which it has been assigned. This organization can easily be a subsidiary or a separate division of a larger company. Moreover, a company may change its name and/or may merge. Since the patent office does not employ a consistent company code in its computer record, except for the "top patenting companies" where the list of subsidiaries is checked manually, the company patenting numbers produced by a simple aggregation of its computer records can be seriously incomplete. (See Hall et al, 1988, for additional detail on this range of issues).

Because of such considerations and because he was interested in tracing through the spillover effects of R&D on productivity in industries which were most likely to benefit from them, Scherer (1982, 1984) undertook the large task of examining over 15,000 patents awarded from June 1976 through March 1977 to the 443 largest U.S. manufacturing corporations represented in the FTC's Lines of Business survey in 1974. There are at least two unique aspects to this data construction effort: First, each patent was examined individually, classified as to product or process invention, and assigned to up to three potential industries of "use" or two possible general use categories. In addition, the patent was also assigned to an industry of "origin" on the basis of the information on the location of the inventors within the Lines of Business structure of the particular company. That is, and this is the second unique aspect of these data, the industry of origin was defined "below" the company level, at the more relevant "business" or divisional level and the R&D

expenditures of the companies were similarily subdivided and matched at this more appropriate industrial level. One of the final products of this work was a "technology flow" matrix, using the resulting cross-classification of patents by industry of origin and industry of use to "flow through" the R&D expenditures from industries in which they have been incurred to industries whose productivity growth may reflect the fruits of such expenditures. (Such a matrix was suggested by Schmookler 1966, p. 167.) Unfortunately, this large, one-time data construction effort does not really have a time series dimension to it. Moreover, the FTC has discontinued collecting data at the Lines of Business level in 1979, making it less likely that it could be replicated in the future.

A less ambitious but somewhat more extensive data construction effort was pursued by the NBER group (see Bound et al, 1984, Griliches, Pakes and Hall 1987, and Hall et al, 1988) who tried to match the patent office data on patents issued to all organizations from 1969 through 1982 with income and balance sheet amd stock market value data for all publicly traded manufacturing corporations, defined as of 1976, and also create a consistent historical record for them for the period 1959-1981. The resulting data sets consisted of a cross-section of about 2600 firms in 1976 (with over 1700 firms receiving at least one patent between 1969 and 1979, about a 1000 firms applying for at least one, ultimately granted, patent in 1976, and about 1500 firms reporting R&D expenditures in 1976) and a panel of about a 1000 to 1800 firms with detailed data between 1963 and 1981, with a subset of about 700 firms reporting consistent R&D data between 1972 and 1980. These data sets formed the basis for a number of studies which will be discussed below.

3. Patents as Indicators of What?

There are two ways of asking this question: What aspects of economic activity do patent statistics actually capture? And, what would we like them to measure? Ultimately, only the first question is of relevance but it is useful to spend some time on the second, since it provides some understanding of the research in this field.

Roughly speaking, we would like to measure and understand better the economic processes which lead to the reduction in the cost of producing existing products and the development of new products and services. We would like to measure both the inputs and the outputs of such processes, to understand what determines the allocation of resources to such "technology changing" activities, and also what is happening and why to the efficiency with which they are pursued in different times and in different places. Assuming that different new products can be brought to a common denominator through the use of some meta-hedonic function, one can think of invention as shifting outward the production possibilities frontier for some generalized aggregate of potential human wants. Ideally, we might hope that patent statistics would provide a measure of the output of such an activity, a direct reading on the rate at which the potential production possibilities frontier is shifting outward. The reality, however, is very far from it.

The dream of getting hold of an output indicator of inventive activity is one of the strong motivating forces for economic research in this area. After all a patent does represent a minimal quantum of invention which has passed both the scrutiny of the patent office as to its novelty and the test of the investment of effort and resources by the inventor and his organization into the development of this product or idea, indicating thereby the presence of a non-negligible expectation as to its ultimate utility and marketability. One recognizes, of course, the presence of a whole host of problems: Not all inventions are patentable, not all inventions are patented, and the inventions that are patented differ greatly in "quality," in the magnitude of inventive output associated with them. The first two problems, one thinks, can be taken care of by industry dummy variables, or by limiting the analysis to a particular sector or industry. For the third, one tries to invoke the help of the "Law of Large Numbers": "the economic ... significance of any sampled patent can also be interpreted as a random variable with some probability distribution." (Scherer 1965, p. 1098.) The question whether our samples are large enough, given the underlying heterogeneity in what is measured by a patent, is a topic to which I shall return below.

It is interesting to note that Schmookler started out thinking that he could use patent statistics as an index of inventive output and as an explanation of the growth in the aggregate efficiency of the U.S. economy. Schmookler was the first, as far as I can tell, to publish numbers on aggregate "total factor productivity growth" (Schmookler 1952) (though he never seemed to have claimed much originality for it), and to relate them to patent statistics (Schmookler 1951). Unfortunately, the relationship did not work. There seemed to be little correlation between aggregate total factor productivity and total patenting numbers. Schmookler did not give up on patent statistics but ultimately redefined what he thought they could do. In his hands patents became an index of inventive "activity," primarily an input rather that an output index.

He moved, essentially, in the direction of what patents can measure rather than what we would want them to measure. His interpretation of inventive activity became quite narrow. It excluded research, which he interpreted as a search for new knowledge, an attempt to discover new properties of classes of objects or processes, and it excluded also development, which is largely the development and refinement of already made inventions (even though quite a few patents are likely to be generated also during this phase).. Inventive activity per se is "work specifically directed towards the formulation of the essential properties of a novel product or process" (Schmookler 1966, p.8.) This is an "input" definition, to be thought of as computable in manhour equivalents and corresponds to only a very thin slice, both quantitatively and in the time dimension, of what is usually covered by the notion of R&D and the associated R&D statistics.

One should keep in mind, however, the historical context of most of the earlier work on patents. There were no R&D expenditure statistics of any generality before the late 1950s and only scattered numbers on scientists employed in different industrial laboratories or on the distribution of the technically trained labor force (see Mowery, 1983). Thus, an indicator of input was also valuable. There was almost no substitute for it. Even today, with data much more plentiful, the available detail in the published R&D statistics is still quite limited. Thus, as I shall argue below, showing that patent statistics are a good indicator of inputs into inventive activity is a useful accomplishment on its own merit. It allows us an insight into what is going on in more areas and also in much more detail than is possible to glimpse from the available R&D statistics.

How does one come to know whether patent statistics measure anything interesting? Input or output? One way of doing it is by looking for correlations between patent counts and other variables which are thought to matter: input measures such as R&D expenditures, and output measures such as productivity growth, profitability, or the stock market value of the firm. It is useful, therefore, to introduce here a figure (Fig. 3) from Pakes and Griliches (1984) which essentially restates the previous sentence in graphic terms and allows a more detailed discussion of its underlying assumptions.

In the center of Figure 3 is an unobservable variable, $\dot{K}$, the net accretion of economically valuable knowledge. This is the variable that we would like to measure. It is the measure of "inventive output" which one would hope that patents would be a good indicator of. The diagram indicates that and adds an error v to the determinants of patenting, making them an imperfect, fallible measure of $\dot{K}$. The causal part of this diagram starts in the lower right hand corner with some observable measure of resources invested in inventive activity (R), usually R&D expenditures, or the number of research scientists, which are directed at the production of $\dot{K}$. Because knowledge production is stochastic, the u term is added to reflect its changing efficiency and the impact of other informal and unmeasured sources of $\dot{K}$. The variables that we are ultimately interested in explaining are represented by the Z's. These could be various measures of growth, in the productivity, profitability, or the stock market value of the firm. They are all affected by the unobservable $\dot{K}$, by other measurable variables X, and by additional random components, the e's.

A number of extreme simplifications were made in drawing this figure and in defining the various terms. For example, the relationship between $\dot{K}$ and K should be defined explicitly to allow for the possibility of decay in the private value of knowledge. Also, R is taken as exogeneous. If, as is likely, the u's are correlated over time, then one might expect them to feed back into R in subsequent periods. Nor do patents play an explicit economic role here. They are just an indicator of $\dot{K}$. The assumption being made is that some random fraction of $\dot{K}$ gets patented. It is a statistical descriptive model rather than a "theory" of patenting. A "theory" would have to be more explicit about the conditions (economic, technological, and legal) for the benefits from applying for a patent outweighing the potential monetary costs and technology disclosure consequences and would add more structure to the relationship between P and $\dot{K}$.

Such a theory would start with the underlying notion of a research project whose success depends stochastically on both the amount of resources devoted to it and the amount of time that such resources have been deployed. Each technical success is associated with an expectation of the ultimate economic value of a patent to the inventor or the employer. If this expectation exceeds a certain minimum, the cost of patenting, a patent will be applied for. That is, the number of patents applied for is a count of the number of successful projects (inventions) with the economic value of the patent right exceeding a minimal threshold level. If the distribution of the expected value of patenting successful projects remains stable, and if the level of current and past R&D expenditures shifts the probability that projects will be technically successful, an increase in the number of patents can be taken as an indicator of an upward shift in the distribution of $\dot{K}$. Whether the relationship is proportional will depend on the shape of the assumed distributions and the nature of the underlying shifts in them. What is depicted in Fig. 3 is at best a very crude reduced-form-type relation whose theoretical underpinnings have still to be worked out. But one has to start someplace.[4]

There are also ambiguities in the definition of $\dot{K}$ and K. Are we talking about private or social returns to knowledge? That depends on the Z's available to us and the question we are particularily interested in answering. For an analysis of productivity movements at the level of industries, it is the social value that we care about. For an analysis of the stock market value of different firms, only the private value version makes any sense. One may also wish to distinguish between the value of patent _rights_ and the economic value of a particular patent. It is the latter notion that we might be interested in, though it is the former that is likely to show up in survey responses of patentors or be implicit in the decision whether to pay a fee and renew a particular patent. Nevertheless, Figure 3 does provide a schema for discussing much of the research in this area and in particular the question of the "quality" of patent counts as indicators of economically valuable knowledge.

There are several different ways of rephrasing this question: 1. How good is P as an indicator of $\dot{K}$? 2. If P is an "output" measure and R is an "input" measure, are we better off in having one or the other if we had to, or could, make such a choice? 3. What is the value added of P, above and beyond R, to the explanation of the Z's? Since $\dot{K}$ is intrinsically unobservable, the first question cannot really be answered without embedding it in some model such as is sketched out in this figure. It may be helpful, at this point, to write down the simplest possible model that might correspond to this figure:

$$\dot{K} = R + u$$

$$P = a\dot{K} + v = aR + au + v$$

$$Z = b\dot{K} + e = bR + bu + e$$

where the first equation is the "Knowledge production function" with the unobservable $\dot{K}$ being measured in units of R; the second equation is the indicator function relating P to $\dot{K}$; and third equation represents the influence of $\dot{K}$ on subsequent variables of interest. The important assumption that will be made here is that the various random components u, v, and e are independent of each other. I need not repeat the caveats about the simplicity of this model. It is adequate, however, for making the following points: 1. The "quality" of P as indicator of $\dot{K}$ depends on the size of v, the error in the indicator relationship. If we take its variance as a measure of its error and we substitute R for $\dot{K}$ in this relationship, as in the right-hand part of the second equation above, we see that under the assumptions of the model the "quality" of the relationship between P and R provides a lower bound on the "quality" of P as an indicator of $\dot{K}$. That is $\mathrm{var}(au+v) > \mathrm{var}(v)$. This argument suggests looking at the correlation between P and R and claiming that if $\dot{K}$ is the output of the R process and P is an indicator of its success then the correlation between P and $\dot{K}$ would have been even higher, if it could have been measured. This is the sense in which the correlation coefficient between P and R provides a downward biased measure of the quality of P as a indicator.[5] 2. The comparative qualities of P and R as proxies for $\dot{K}$ depend on the relative size of the variance of v and u. If the error of measurement in P is large relative to the stochastic fluctuation in $\dot{K}$, then R may be the better variable even if it does not reflect u. 3. If the stochastic component of $\dot{K}$ is important and if P actually captures any of it, there should be some value added in P above and beyond R. But if the error of measurement in P is large and the samples are small, we may not really see it in the regressions results when P is included as an additional variable.

4. Patents and R&D.

In the attempt to "validate" patents as an economic indicator their relationship to R&D activity has been investigated repeatedly. Schmookler (1966, Chapter 2) and Scherer (1965) are leading examples of earlier investigations. More recent results can be found in Bound et al (1984), Hall et al (1986), Pakes and Griliches (1984), Scherer (1983) and Acs and Audretch (1989). Several conclusions as well as a number of unresolved questions emerge from this work.

A major conclusion, emphasized by Pakes and Griliches, is that there is quite a strong relationship between R&D and the number of patents received at the cross-sectional level, across firms and industries. The median R-square is on the order of 0.9, indicating that patents may indeed be a good indicator of unobserved inventive output, at least in this dimension. That this relationship is not just due to size differences can be seen in Figure 4 (taken from Bound et al), which plots both patents and R&D per unit of a firm's assets.

The same relationship, though still statistically significant, is much weaker in the within-firms time series dimension. The median R-square here is on the order of 0.3 (in contrast to the 0.9 in the cross-sectional dimension). Nevertheless, the evidence is quite strong that when a firm changes its R&D expenditures, parallel changes occur also in its patent numbers. The relationship is close to contemporaneous with some lag effects which are small and not well estimated (Hall et al). This is consistent with the observation that patents tend to be taken out relatively early in the life of a research project. Since the bulk of R&D expenditures are spend on development, most of the time series variance in this variable must come from the differential success in the further development of existing projects rather than from the initiation of new ones.[6] The relatively low correlations in the time dimension should, therefore, not be all that surprising, but they imply that patent numbers are a much poorer indicator of short-term changes in the output of inventive activity or the "fecundity" of R&D.

The question of "Are there diminishing returns to R&D?" hovers over much of this work. In the cross-sectional dimension it is related to the "Schumpeterian" question whether large firms and large R&D labs are more or less efficient as "engines of innovation." (See Baldwin and Scott, Chap. 3, and Cohen and Levin for more general reviews of this topic.) In the time series dimension one is faced with the declining ratio of patents received per R&D dollar spent and the worry that technological and inventive opportunities are being exhausted. There is also the question how does one reconcile the significantly larger estimates of the elasticity of patenting with respect to R&D in the cross-sectional versus the time-series dimensions.

At the cross-sectional level the story is relatively simple. Small firms appear to be more "efficient," receiving a larger number of patents per R&D dollar. This can be seen most easily in Figure 5 (from Bound et al), which plots the patents per R&D ratio as a function of the size of the R&D program. It shows both the much higher ratio for small firms and the fact that this relationship becomes effectively flat, beyond some minimum size. At the larger firm level, where anti-trust policy might be relevant, there is no strong evidence of diminishing returns to the size of the R&D effort. (This is also the conclusion reached by Scherer, 1983, on the basis of a different and better set of data.) Given the non-linearity and the noisiness in this relation, the finding of "diminishing returns" is quite sensitive to functional form, weighting schemes, and the particular point at which the elasticity is evaluated.

All of this can be seen in Figure 6, also taken from Bound et al, which plots the original data and the results of fitting various different models to the same data. Two of the estimation techniques, Poisson and Non-linear Least Squares, indicate diminishing returns, while the other two techniques, OLS and Negative Binomial, imply increasing returns. A glance at the figure will make it clear how a differential emphasis on parts of the data (large versus small firms and the treatment of zeroes - not visible in the figure) could result in such conflicting estimates. Basically there is a sharp contrast between smaller and larger firms. For larger firms the relationship is close to linear while there is a reasonably large number of smaller firms that exhibit significant patenting while reporting very little R&D. When divided into two samples, small (N=1015) and large(N=483), with $2 million in R&D expenditures as the dividing line, the estimated average elasticities are 0.44 and 1.04 respectively. The latter number falls to 0.8 (0.1) if one allows separately for the zero patents observations. Though this estimate of the elasticity of patenting with respect to R&D for the larger firms is still "significantly" less than unity at conventional test levels, allowing for the possibility that the R&D numbers are themselves subject to error, one cannot really reject the hypothesis of constant returns in this size range, since the "reciprocal" regression of R&D on patents implies increasing returns or decreasing costs of getting a patent. (The estimated elasticity of R&D with respect to patents is 0.76).

.The appearance of diminishing returns at the cross-sectional level is due, I think, primarily to two effects: selectivity and the differential role of formal R&D and patents for small and large firms. Most of the data sets available to us are not based on a random or carefully stratified sample from the relevent underlying population. Rather, they are "opportunity" samples, based on other criteria. For example, the 1976 cross-section of Bound et al, is based on all manufacturing firms listed on the New York and American Stock exchanges and also on the Over-the-Counter market. But while almost all relevant large firms are so listed, only a relatively small number of the smaller firms trade in these markets. To be included in (listed on) the market, a small firm has to be in some sense more "succsessful" than those that are not, more "interesting" to the traders. Thus, it is not surprising that it may also hold more patents than might be expected, given its size and R&D program. How atypical these small firms might be is suggested by the rarity of their selection. Table 1 shows the number of firms by size (employment) in this cross-section and the corresponding numbers in the relevant population. While about two-thirds of the large manufacturing firms are included, the smaller one represent less than one percent of all small firms and are obviously a heavily selected lot. Unfortunately, we have no information on the firms not in the sample and hence cannot make an appropriate sample selectivity adjustment.

Another source of the difference between small and large firms is in the role of formal R&D in them and the differential importance of patents to them. A significant amount of patenting is not the result of formal R&D activities though the relative importance of organized R&D rises with the size of the company. Small firms are likely to be doing relatively more informal R&D, reporting less of it, and hence providing the appearance of more patents per reported R&D dollar.[7] Also, for such firms patents may represent their major hope for ultimate success and hence would lead them to pursue them with more vigor. A well established major firm does not depend as much on current patenting for its viability or the survival of its market position. Thus, even at equal underlying true inventiveness rates, the propensity to patent may be lower for large firms, at least relative to the successful new entrants in their field. But in the major range of the data, from middle size to giant firms, there is little evidence for diminishing returns, at least in terms of patents per R&D dollar.[8] That is not surprising, after all. If there were such diminishing returns, firms could split themselves into divisions or separate enterprises and escape them.

The time series dimension has been examined most extensively by Hall et al (1986) (see also Pakes and Griliches 1984 and Hausman et al 1984). The estimated total elasticity of patents with respect to R&D expenditures is between 0.3 and 0.6, even after allowing for several lagged effects. This finding, in contrast to the cros-sectional results, is robust with respect to differential weighting and alternative estimation methods. It is tempting then to accept the diminishing returns result in the within-time-series dimension and interpret it as reflecting real diminishing returns, in terms of patents received, to the expansion of existing research programs. But this conclusion is unnecessary. The relationship between annual changes in R&D and in patenting is very weak, although "statistically" significant, at the firm level. If one allows for the possibility that much of the annual fluctuations in R&D has little to do with that part of inventive activity which generates patents, being largely the result of fluctuations in and vaguaries of the development portion of the various research projects, then the "relevant" R&D

is measured with error and the estimated coefficients are downward biased. This is not a pure "measurement error" case, since reported R&D may be correctly reported as far as its own definition goes, but not exactly what we want (R&D directed at patentable inventions). This is parallel to the transitory-permanent distinction in consumption theory and is isomorphic to the "errors-in-variables" model. Invoking the latter, we may be able to "bracket" the true returns to scale coefficient by running the regression the other way, R&D on patents, and computing the reciprocal of the resulting coefficient. The low correlation between the two rates of change results also in a very low coefficient in this second dimension, on the order of 0.1-0.2, and an implication of increasing returns. The latter should not be taken seriously either, since it is the result of the great randomness in the patent series themselves. The point of this digression is, however, to remind one that the appearance of diminishing returns in such data could be an artifact of the incompleteness of the underlying data rather then a reflection of the characteristics of the invention process itself. As of the moment, the evidence is suggestive but not conclusive.

Besides differing by size of firm, the R&D to patents relationship differs also across industries. In absolute terms, the industries with the largest numbers of patents are drugs, plastics, other rubber products, and computers (in Scherer's LB based data) and instruments, communication equipment, and industrial chemicals (in the OTAF concordance based data). In terms of the "propensity to patent" (patents per R&D dollar), the differences are less apparent and more difficult to interpret. One can look at the tables (5-9) in Griliches (1989) or the Appendix to Cockburn and Griliches (1988) and observe that "low" propensity to patent industries include obvious cases of large R&D industries with significant governmental research support, such as motor vehicles and aircraft, who patent very much less than would be predicted from their R&D numbers alone. Amongst the "high" propensity to patent industries, besides the expected presence of communication equipment, there are a number of industries (such as Screws, Nuts, and Bolts) whose appearance is due to their doing very little R&D but still taking out some occasional patents. An attempt to explain the dispersion in such numbers across industries using data from the Yale Survey (Levin et al, 1987) on the perceived differential effectiveness of patents as a method of appropriating the benefits from innovation was largely unsuccessful. The patent to R&D ratios appear to be dominated by what may be largely irrelevant fluctuations in the R&D numbers and the Yale Survey responses themselves appear to have little relevant cross industry variability in them (see Griliches, 1987, Cockburn and Griliches 1988 and Cockburn 1989). For example, while the drug industry has the highest rating on the patents provide protection scale, its patents per R&D ratio is much lower than that for firms in the paper industry, where the effectiveness of patents is rated to be somewhat below average (see Cockburn and Griliches, appendix C). Since the effectiveness of patents as an appropriability mechanism will affect also the incentive to do R&D, the resulting impact on the ratio of the two is far from obvious. In drugs it clearly encourages research with the result that even with extensive patenting the observed ratio is not much above average. Thus, it is probably misleading to interpret such numbers as being direct indicators of either the effectiveness of patenting or the efficiency of the R&D processes.

5. Patents Rights and Patent Values

Because the economic significance of individual patents is so variable, there has been continued interest in trying to estimate the average value of patent rights, the average value of the invention represented by a particular patent, and the dispersion in both of these concepts. Looking at patents as indicators of success of the underlying inventive activity or R&D program, we are mainly interested in the second concept. The available data, however, are mostly informative only about the first: the value associated with the differential legal situation created by the possession of the patent.

There are basically three sources of data on this topic: 1. Results of direct surveys of patent owners or assignees about past returns and the potential market value of their rights. 2. The valuation implicit in the decision whether to pay a fee to renew the patent, a decision that had to be made by European patent holders in the past and is now also facing U.S. patent holders. And 3, econometric analyses of the relationship of some other value denominated variable, such as profits or stock market value, to the number of patents. An example is the use of patent numbers as a proxy for "intangible" capital in stock-market-value of the firm regressions.

The most detailed and extensive survey of patent holders was conducted over 30 years ago by Barkev Sanders and associates at the Patent and Trademarks Foundation (see Rossman and Sanders 1957, Sanders et al 1958, and Sanders 1962 and 1964, and the discussion of it in Schmookler's book, pp. 47-55). They conducted a mail survey in 1957 of the owners and assignees of a two percent random sample of all patents issued in 1938, 1948, and 1952. There were two major findings in this survey: 1. A surprisingly large fraction of all sampled patents was reported to have been "used" commercially, either currently or in the past. The actual fraction "used" is sensitive to the treatment of non-response. it is over 55 percent for those responding and about 41 percent if one assumes that non-response is equivalent to non-use. The "use" percentage is higher for "small" companies, but so is also the non-response rate (71 percent used among respondents, 40 percent if adjusted for non-response). Thus, it is not true that most patents are never used and are hence not associated with a significant economic event. This finding is also consistent with the renewal information to be discussed below. In Europe, about 50 percent of all patents granted are still being renewed and a renewal fee is being paid 10 years after they had been applied for. 2. The reported economic gain from the innovations associated with these patents was very dispersed. Among the patents reported to be in current use and with relevant numerical responses and a positive gain (accounting for about 20 percent of all the relevant responses), the mean value was $577,000 per patent, but the median value was only about $25,000 (implying, under the assumption of log-normality, 2.5 as the coefficient of variation and a standard deviation of about $1.5 million). If one includes all the no gain, loss, and not yet used patents, the mean gain falls to about $112,000, and the median is close to zero or below, (computed from the tables in Sanders et al, 1958, p. 355 and 357). Even this lower mean number is quite impressive, roughly equivalent to $473,000 per average patent in 1988 prices (using the GNP deflator to convert it from 1957 prices), but so is also the associated dispersion. Scherer (1965) reports that fitting a Pareto-Levy distribution to these data graphically yielded an estimate of the exponent (alpha) of about 0.5, implying a distribution with no finite mean or variance. If this were truly the case, then even in large samples the mean value of patents would not converge rapidly, if at all, to its underlying population average.

There have been only very few other attempts at such a survey and they all reach rather similar conclusions. Schmookler (1966, p. 54-5) reports on a small mail sample with a mean value of $80,000 and a median of about zero. In 1982 the Chemistry Program of NSF decided to evaluate the economic value of patents attributable to its grants (Cutler 1984). Of the 96 patents surveyed, 52 had been licensed or were deemed licensable with an average "economic value" of about $500,000 per patent. (The concept of "economic value" is unclear in this study. It appears to refer to total potential sales of the product rather than net returns to the owners of the patent.) A related study, done for the NSF by SRI International (1985), examined a sample of patents received by the grantees of the Engineering Program and estimated the royalty potential of each patent, which turned out to be about $73,000 on average, again with a very large dispersion. A more representative and large scale survey of patent holders is both feasible and desirable but nothing has been done in this regard since 1957 and there does not seem to be anything like it in the works either in the U.S. or abroad.

In many countries and recently also in the U.S., holders of patents must pay an annual renewal fee in order to keep their patents in force. If the renewal fee is not paid in any single year the patent is permanently cancelled. Assuming that renewal decisions are based on economic criteria, agents will only renew their patents if the value of holding them over an additional year exceeds the cost of such renewal. Observations on the proportion of patents that are renewed at alternative ages, together with the relevant renewal fee schedules, will then contain information on the distribution of the value of holding patents, and on the evolution of this distribution function over the lifespan of the patents. Since patent rights are seldom marketed, this is one of the few sources of information on their value. In a series of papers Pakes and Schankerman (1984), Pakes (1986), and Schankerman and Pakes (1986) present and estimate models which allow them to recover the distribution of returns from holding patents at each age over their lifespan. Since the renewal decision is based on the value of patent protection to the patentee, the procedure used in these articles directly estimates the private value of the benefits derived from the patent laws.

In Figure 8 typical European data on renewal fees and patent survival proportions are reproduced from Schankerman and Pakes (1986). They indicate several interesting facts that should be kept in mind. About half of all patents are renewed through age 10, indicating a significant expectation of some "usefulness" for the majority of patents for some non-negligible time period. On the other hand, the same data indicate that about half of all patents are not renewed within ten years, indicating that the expected value of the future income stream from these rights has fallen below the rather low renewal cost. This implies that the majority of patents are either of low value, or that their value depreciates (obsoletes) rapidly, or both. About 10 percent of all patents survive and pay the fees for the whole statutory period and obviously contain a smaller number of very valuable patents. Pakes and Schankerman use these facts in their various papers to construct models of the renewal process and estimate both a distribution of the underlying patent right values and also their rate of depreciation. Given the existence of an open-ended class of patents in these data (those paying the renewal fees throughout the whole period) and the rather low and relatively stable renewal fee schedules, serious identification problems arise in such models. The estimates of the mean value of patent rights rest, therefore, on specific assumptions about the functional form of their distribution (how it looks in the unseen tail) and on assumptions about the form of the depreciation process. Some of these assumptions may be testable and some of the more interesting conclusions of their work do not depend on them, but ultimately we have to put some prior notions into such data to have them yield specific numerical answers. The issues of identification and estimation are discussed in much detail in the recent papers by Pakes and Simpson (1989) and Schankerman (1989), together with the presentation of interesting new results on additional countries and on industrial detail, and hence will not be pursued further here. (See also Lanjouw, 1989, Schankerman 1990, and Lanjouw and Schankerman 1989.)

In the United States, patents which were applied for after 1980 have to pay renewal fees 3 1/2, 7 1/2, and 11 1/2 years after the granting date. These fees are currently $450, $890, and $1340 respectively for corporations and somewhat less than that for individuals and "small entities." As of the end of 1988, 16 percent of the 1981-84 patents coming up to the payment of the first maintenace fee "expired," with a slightly higher expiration rate for U.S. (17 percent) than for patents owned by foreign residents (15 percent) implying, possibly, a higher average value or "quality" for the latter. An earlier study of a smaller sample of such data found that individually owned patents were expiring at a much higher rate than assigned patents (39 vs 13 percent for U.S. origin patents) and that "mechanical" patents had the highest and "chemical" patents the lowest rates of expiration (Manchuso et al, 1987). The growing availability of such renewal data in the future will provide us with another very interesting window on the inventive process and its rewards in the U.S.

Returning to the specific results from the work on European patent renewals, using a learning model for the early years of a patent's life, Pakes (1986) finds that patents are applied for at an early stage in the inventive process, a stage in which there is still substantial uncertainty concerning both the returns that will be earned from holding the patents, and the returns that will accrue to the patented ideas. Gradually the patentors uncover more information about the actual value of their patents. Most turn out to be of little value, but the rare "winner" justifies the investments that were made in developing them. His estimates imply also that most of the uncertainty with respect to the value of a patent is resolved during the first three or four years of its life. Using this result, Schankerman and Pakes (1986) examine changes in the distribution of patent values over time and the correlates of these changes. The substantive results from these papers imply that the average value of a patent right is quite small, about $7,000 in the population of patent applications in France and the UK. In Germany, where only about 35 percent of all patent applications are granted (about 93 percent and 83 percent were granted in France and the UK respectively), the average value of a patent right among grants was about $17,000. The distribution of these values, however, is very dispersed and skewed. One percent of patent applications in France and the UK had values in excess of $70,000 while in Germany one percent of patents granted had values in excess of $120,000. Moreover, half of all the estimated value of patent rights accrues to between five and ten percent of all the patents. The annual returns to patent protection decay rather quickly over time, with rates of obsolescence on the order of 10 to 20 percent per year. Since about 35,000 patents were applied for per year in France and the UK and about 60,000 in Germany, these figures imply that though the aggregate value of patent rights is quite large, it is only on the order of 10 to 15 percent of the total national expenditures on R&D. Other means of appropriating the benefits of R&D must be, therefore, quite important.

Schankerman and Pakes used their results to adjust the aggregate patent time series for changes in their average "quality" (value). In their 1986 paper they find that even though the number of patents per scientist fell rather sharply between 1965 and 1975 in the three countries examined by them, the estimated "quality-adjusted" total value of patent rights per scientist and engineer was effectively stable in both Germany and the UK, and dropped only slightly in France (Schankerman and Pakes, 1986, Table 6).[10]

6. Patents and Stock Market Value

Another line of work has used data on the stock market valuation of firms to investigate both the "value" of patents and the information content of the variability in their numbers. The use of stock market values as an "output" indicator of the research process has one major advantage. All other indicators of success, such as profits or productivity are likely to reflect it only slowly and erratically. On the other hand, when an event occurs that causes the market to reevaluate the accumulated output of a firm's research endeavours, its full effect on the expected present value of a firm's future net cash flows should be recorded immediately. This, of course, need not equal what will eventually materialize. The downside of this type of measurement is the large volatility in stock market measures. The needle might be there but the haystack can be very large.

The simplest market value model starts from the market valuation identity, with the market value of the firm proportional to its physical ("tangible") and intangible capital, the latter being in part the product of its past R&D investments and possibly also reflected in its accumulated patent position (Griliches 1981, Ben-Zion 1984, Hirschey 1982, and Cockburn and Griliches 1988, among others). It can be written as follows:

$$V = q(A+gK) = qA(1+gK/A)$$

where V is the market value of the firm, A is the current replacement cost of its tangible assets, K is its level of intangible ("knowledge") capital and g is its relative shadow price, and q is the current premium or discount of market value over the replacement cost of tangible assets.[11] Writing q as exp(a+u), where a represents individual firm differences in average valuation due to the exclusion of other unmeasured capital components or market position variables, taking logarithms, and approximating log(1+x) = x, we can rewrite the estimating equation as:

$$\ln Q = \ln(V/A) = a + gK/A + u$$

where the dependent variable is the logarithm of what has come to be called "Tobin's Q." Using different measures of current and past patents and current and past R&D expenditures as proxies for K, this kind of equation has been estimated by various researchers. Table 3 reproduces a number of results from the Cockburn and Griliches study. It shows that if we look at patents alone the estimated value of a recent patent is about $500,000. This estimate is halved when we put in both past and current R&D expenditures in the equation. By and large, R&D is the "stronger" variable. The evidence for additional information in the patent variables varies from sample to sample (patents were stronger in the Griliches 1981 study which was based on a much smaller sample of firms but also used the panel aspects of the data) and depends on which other variables are included in the equation (see the move of the results from columns 2 to 3 in this table).[12]

A more dynamic point of view is taken by Pakes (1985) in his analysis of the relationship between patents, R&D, and the stock market rate of return. Events occur which affect the market value of a firm's R&D program and what one estimates are the reduced form relationships between the percentage increase in this value and current and subsequent changes in the firm's R&D expenditures, its patent applications, and the market rate of return on its stock. His empirical results indicate that about five percent of the variance in the stock market rate of return is caused by the events which change both R&D and patent applications. This leads to a significant correlation between movements in the stock market rate of return and unpredictable changes in both patents and R&D expenditures, changes which could not be predicted from past values of patents and R&D. On average, an "unexpected" increase in one patent is associated with an increase in the firm's market value of $810,000, while an unexpected increase of $100 of R&D expenditures is, on average, associated with a $1,870 increase in the value of the firm. Patents are estimated to contain a significant noise component (a component whose variance is not related to either the R&D or the stock market rate of return series). This noise component accounts for only a small fraction of the large differences in the number of patent applications of different firms (about 25%), but plays a much larger role among the smaller fluctuations that occur in the patent applications of a given firm over time (about 95%). Similarly, the effect of unexpected increases in patents on market value is highly variable. Nevertheless, there is still some information in the time-series dimension. If we were to observe, for example, a sudden large burst in the patent applications of a given firm, we could be quite sure that events have occurred to cause a large change in the market value of its R&D program; but smaller changes in the patent applications of a given firm are not likely to be very informative.

The timing of the response of patents and R&D to events which change the value of a firm's R&D effort is quite similar. One gets the impression from the estimates that such events cause a chain reaction, inducing an increase in R&D expenditures far into the future, and that firms patent around the links of this chain almost as quickly as they are completed, resulting in a rather close relationship between R&D expenditures and the number of patents applied for. Perhaps surprisingly, Pakes finds no evidence that independent changes in the number of patents applied for (independent of current and earlier R&D

expenditures) produce significant effects on the market's valuation of the firm. Hence it is not possible to distinguish between demand shocks, where demand shocks are loosely defined as events which cause increases in patenting only through the R&D expenditures they induce, and technological or supply shocks which may have a direct effect on patents as well as an indirect effect via induced R&D demand.

It is not obvious whether one can separate "demand" from "supply" factors in this area, even conceptually. One way of defining "demand" factors is to identify them with macro shifts in aggregate demand, population, exchange rates, and relative factor prices that make inventive activity more (or less) profitable at a given level of scientific information, a fixed "innovation possibilities frontier." Changes in technological "opportunity," on the other hand, are those scientific and technological breakthroughs which made additional innovation more profitable or less costly at a fixed aggregate or industry level demand. These distinctions are far from sharp, especially given our inability to measure the contributions of science and technology directly. Moreover, what is a technological opportunity in one industry may spillover as a derived demand effect to another. Nevertheless, there is something distinct in these factors, in their sources of change and dynamics.[13]

Patent data could help here if one were willing to assume that independent, "unanticipated" shifts in the level of patenting by firms, represent shifts in technological opportunities and not responses to changes in economic conditions (demand forces). That is, the identiying assumption is that demand impinges on the level of patenting only through the level of R&D expenditures (and slowly changing trends) and that the "news" component in the patent statistics reflects technological "news," the information that a particular line of research has turned out to be more (or less) fruitful or easier (harder) than expected when the decision to invest in it was made originally. Changes in technological opportunity are thus identified with "abnormal," "unexpected," bursts (or declines) in the number of patents applied for.

Several implications of this formulation are immediate. If patent statistics contain additional information about shifts in technological opportunities, then they should be correlated with current changes in market value above and beyond their current relationship with R&D and they should affect R&D levels in the future, even in the presence of the change in market value variable since the latter variable is measured with much error. Patents should "cause" R&D in the sense of Granger (1969).

The available evidence on this point is not too encouraging: As noted above, Griliches (1981) found a significant independent effect of patents on the market values of firms, above and beyond their R&D expenditures, but Pakes did not detect a significant influence of lagged patents on R&D in the presence of lagged R&D and the stock market rate of return variables. Nor did Hall, Griliches and Hausman (1986) find future R&D affecting current patenting as the "causality" argument might have implied. Griliches, Hall and Pakes (1990) replicate some of Pakes' computations on a larger sample (340 firms) and expand his equation system to add equations for sales, employment, and investment. Their results indicate that the addition of the latter variables is helpful, in the sense that fluctuations in their growth rates are related to fluctuations in both the growth rate of R&D and the stock market rate of return and hence should help in identifying the relationships we are interested in. But the expansion of the sample to include many small firms with low levels of patenting, deteriorates significantly the informational content of this variable, raising its noise to signal ratio, and making it hard to discern a feedback from the independent variablity in patenting to any of the other variables. Thus, at the moment, it does not look as if the data can sustain a model with two separate factors ("market" and "technological" innovations), even though in principle such a model is identifiable.

The difficulties in implementing such models arise to a large extent from the large "noise" component in patents as indicators of R&D output in the short-run within-firm dimension. While the problem may have been obvious from the beginning, it was the work of Pakes and Schankerman (1984) and their estimates of the dispersion and the skewness in patent value which alerted us to its actual magnitude.

To derive quantitative implications of such a skewed distribution of values for the quality of this indicator we can combine what we know about patent counts in both the time series and cross section dimension with estimates of the distribution of their values.

One can write the innovation in the value of the firm (net of its expected dividend and investment policy) as the sum of three components:

$$q_t V_t = w_t + \eta_t + u_t$$

where q_t is the rate of return on stock holding, V_t is the total market-value of the firm's assets, and the three components w_t, η_t, and u_t are defined to be orthogonal to each other; w_t corresponds to the change in the value of a firm's R&D "position" (program) arising from the "news" associated with current patent applications; η_t reflects revaluations of previous achievements associated with past patents (above and beyond their correlation with current patents); while u_t reflects all other sources of fluctuation in the value of the firm, including also possibly the contribution of not patented R&D. Looking first at w_t and the role of patent numbers as an indicator of it we can ask about the possible magnitude of the variance of w_t (relative to the variance of q_tV_t). That is, how large could the contribution of current patents be to the explanation of fluctuations in market value, even if we had a perfect measures of these <u>values</u>?

To decompose the variance of the first component, we write it as

$$w = \sum_{i=1}^{p} y_i$$

and assume that (1) p, the number of patents applied for each year is distributed as a Poisson random variable with a mean, λ, which is a distributed lag of past R&D expenditures (see Hausman, Hall, and Griliches 1985); and (2) y_i is the underlying value of each patent and is distributed as a log-normal random variable with a mean and variance which will be derived from the earlier literature.

The first two moments of w (under independence) are

$$E(w) = E\,[py] = \lambda E[y] \qquad \text{where } \lambda = E[p]$$

$$V(w) = V(\sum_{i=1}^{n} y_i) = \lambda\, V[y] + \lambda(Ey)^2$$

The component of the variance of w which could be accounted for by patent <u>numbers</u> corresponds to the last term

$$\text{Var}\,[p\bar{y}] = \lambda\bar{y}^2$$

and its relative size is given by

$$\text{Var }[py]/\text{Var }[w] = 1/(1+V[y]/E[y]^2) = 1/(1+\tau^2)$$

where τ is the coefficient of variation in the distribution of patent values.

Turning to the literature for some order of magnitude estimates of various parameters, we have estimates of the mean value of the news associated with patents in the U.S. of between $200,000 (Griliches, 1981) and $800,000 (Pakes, 1985) per patent. There is also some information on this point in Griliches, Hall and Pakes: an estimate of $98,000 per unexpected patent at the geometric mean of their data (with a very large standard error). For the drug industry, where patents are more important, they get a larger and somewhat more precise estimate: an $821,000 average increase in the value of the firm per unexpected patent. This, in fact, is very similar to the Pakes estimate which was based on a smaller sample of larger firms and is therefore more comparable to their drug firms subset.

Taking the upper range of these numbers, $800,000 per "unexpected" patent, and using $\lambda = 13$, the average (geometric) number of patents received in the Griliches, Hall, Pakes sample (per year, per firm), the expected contribution of the variance in patent <u>numbers</u> to the average variance in market value is $13(0.8)^2$ $(\text{mil\$})^2$ = $8.3 millions squared. To get an estimate of Var(y), I borrow the estimated coefficent of variation of the distribution of patent values from Pakes (1986) and Schankerman and Pakes (1986).[14] Both these articles produced coefficients of variation on the order of 2 to 3.6. Since we are looking for upper-bound estimates, taking 3.6 and applying it to the "upper" range estimate of Ey = $0.8 million, gives an estimate of the total variance of w as

$$13[(3.6\text{x}0.8)^2 + (0.8)^2] = \$116 \text{ million sq.}$$

This is to be compared to the average variance of q_tV_t. The variance of q in the Griliches, Hall, Pakes sample is 0.133 which, evaluated at the geometric average value of their firms ($276 million), yields a variance of market value changes on the order of $10,000 million squared. Comparing the two variances gives an estimate of the relative importance of fluctuations in the market value of new patented innovations as at about one percent of the total fluctuations in market value.[15] That is, even if one had good estimates of patent values, they would account for little of the fluctuations in market value. Having numbers instead of values makes matters much worse, reducing this fraction even further. The contribution of patent numbers to the variance in their values is only on the order of 7 percent $(1/(1 + (3.6)^2))$, and their contribution to the explanation of the variance in the unexpected changes in the market values of individual firms is much smaller (less than 0.1 percent).[16] One should not, therefore, use data on stock market fluctuations in this fashion to test detailed hypotheses about the information content of patent statistics. On the other hand, whilethe estimated variance components are rather small, they should not be interpreted as implying that the returns to inventive activity are small or that the topic we have been pursuing is not interesting, only that we have been looking for our particular needle in a very large haystack.

7. Spillovers and Other Uses of Patent Data

A major unresolved issue in the area of economics of technology is the identification and measurement of R&D spillovers, the benefits that one company or industry receives from the R&D activity of another. It is difficult to trace such spillovers without having strong a priori notions about who are the potential beneficiaries of whose research. (See Griliches, 1979, pp. 102-5 for additional discussion of these issues.) One way to approach this problem is to use the detailed information on patenting by type of patent (patent class) to cluster firms into common "technological activity" clusters and looking whether a firm's variables are related to the overall activity levels of its cluster.

In his thesis and several recent papers, Adam Jaffe (1983, 1985, 1986, 1988) has used firm level data on patenting by class of patent and on the distribution of sales by 4-digit SIC to cluster firms into 21 distinct technological clusters and 20 industry (sales orientation) clusters. It turns out that these two criteria lead to different clusterings. Using the technological clusters Jaffe constructed a measure of the total R&D "pool" available for spillovers (borrowing or stealing) in a cluster. He then looked at three "outcome" variables: R&D investment ratio for the firm (in 1976), patents received (average number applied for during 1975-77), and output growth, between 1972 and 1977. In each of these cases, his measure of the R&D pool contributed significantly and positively to the explanation of the firm level "outcome" variables even in the presence of industry dummies (based on sales clustering). Not surprisingly, perhaps, firms in technological clusters with large overall R&D "pools" invested more intensively in R&D than would be predicted just from their industrial (SIC) location. More interesting is the finding that firms received more patents per R&D dollar in clusters where more R&D was performed by others, again above and beyond any pure industry differences (based on a classification of their sales). Similarly, his analysis of firm productivitry growth during the 1972-77 period showed that it was related positively to both the average R&D intensity of the individual firms and the change in the size of the R&D pool available to these firms. In terms of profits, or market value, there were, however, both positive and negative effects of neighboring firms' R&D. The net effect was positive for high R&D firms, but firms with R&D about one standard deviation below the mean were made worse off overall by the R&D of others. Here the idea of R&D spillovers is made operational by using the firm's patenting pattern to construct a measure of its location in "technological space" and showing that the R&D of others, weighted inversely to their distance from this location has an observable impact on its own success. More recently, Jaffe (1989) has used regional data on patenting to investigate spillovers from academic research.

Patent documents contain also citations to other, previous, patents. Following the growth of interest in citations in general and the development of computer software which allows the search for all subsequent citations of a particular patent (or article), there has been a growing interest in using citations counts as alternative "indexes" of differential quality. It should be noted here that patent citations differ from usual scientific citations to the work of others in that that they are largely the contribution of patent examiners whose task is to delimit the reach of the new patent and note the context in which it is granted. In that sense, the "objectivity" of such citations is greater and may contribute to the validity of citation counts as indexes of relative importance. But in another sense, they are like citations added at the insistence of the editor; they may reflect the importance that is put in the field on particular papers but are not a valid indicator for channels of influence, for intellectual spillovers. On the other hand, they bring us closer to something that might be interpreted as measuring the social rather than just the private returns to these patents.

The use of patent citations as "indicators" is discussed, largely in a bibliometric style, by Campbell and Nieves (1979), Carpenter et al (1981), Carpenter and Narin (1983), and Narin et al (1987) (see also the more general discussion of bibliometric evidence in OTA, 1986, Chapter 3). An interesting economic application is to be found in Trajtenberg (1987) who shows that citation weighted patent numbers are more closely correlated with his "output" measure, consumer surplus gains from the development and diffusion of CAT-scanners (computed tomography), while unweighted patent counts are more closely related to "input," to R&D expenditures by the various firms in this field. (For another application of citation data see Lieberman, 1987). This way of using patent data is only in its beginnings and we are likely to see a much wider use of it in the future.

A number of studies have tried to "validate" patents as indicators of technical change by connecting them to counts of innovations, new chemical entities, and subsequent measures of profits or growth. One of the earliest and best studies of this kind, Comanor and Scherer (1969), related pharmaceutical patents to the number of new chemical entities and all new products introduced by the different firms in subsequent years and found a closer relationship between patent applications (rather than grants) with all new products (rather than just the number of new chemical entities). I will not consider in detail a number of studies which found varying degrees of relationship between patents, "invention" or "innovation" counts, because the subjectivity and elasticity of such innovation count data makes their results very difficult to interpret. For examples of such work see Achilladelis et al (1987), Basberg (1982), Kleinknecht (1982) and Walsh (1984). Scherer (1965) shows a positive relationship between earlier patenting rates and subsequent profitability and sales growth differences in a cross-section of firms, but I know of no studies which relate "successfully" patenting rates or patenting stocks to subsequent growth of productivity at the firm level.

Patent data have been used by Pavitt and Soete and their associates, to analyze the relative "competitiveness" of various countries, to construct "Revealed Technology Advantage" indexes for various countries, and to describe and contrast the international location of inventive activity in different industries. (Pavitt and Soete 1980, 1981, Pavitt 1982, Pavitt and Patel 1988, and Soete 1987). Patents have been used by economic historians to study regional patterns of economic growth and the externalities of population size and agglomeration (Kelly 1972, Sokoloff 1988, and Sokoloff and Khan 1989, among others). There have been also many other attempts to use patent data in different areas of economic analysis. It is not possible, unfortunately, to do justice to all of them here.

FOOTNOTES

* I am indebted to my friends and collaborators for many ideas and comments. Parts of this survey borrow heavily (often verbatim) from our earlier work on this topic, especially from Griliches, Pakes and Hall (1987), Griliches, Hall and Pakes (1988) and Griliches (1989). I am indebted to the National Science Foundation (PRA85-12758 and SES 82-08006) and the National Bureau of Economic Research Productivity Program for financial support of this work and to B. Hall, A. Pakes, K. Pavitt, M. Schankerman, and F.M. Scherer for their comments on an earlier draft. The first draft of this survey was begun while I was a guest of the Rockefeller Foundation at the Bellagio Study and Conference Center in Italy. An earlier version of this paper was presented as the W.S. Woytinsky Lecture of 1989 at the University of Michigan.

1. There are several other good surveys on this range of topics. See especially Basberg (1987), Pavitt (1978 and 1985), Pakes and Simpson (1989), Schankerman (1989), and the earlier books by Schmookler (1966) and Taylor and Silberston (1973).

2. This is especially true of some of the European work on related topics, since it often asks somewhat different questions in a different intellectual framework.

3. See OTAF 1985, the proceedings of the conference on the concordance, for a more detailed discussion of some of these issues.

4. Of course, one need not start here. It is a particularly American view,

which finds thinking in terms of a "production function of knowledge" congenial and useful, and looks for patents to serve as a proxy for the "output" of this process. Less "neo-classically" oriented economists would deny the usefulness of this view or the uniform direction of causality that it implicitly espouses.

5. This conclusion depends on the additive nature of the error in the indicator function. If $\dot{K}$ were to be looked at just as an aggregation of inventive events, each with a potential value of its own, drawn independently from some value distribution, and P counted only some fraction of such events and was not related to their values, (as in the calculations outlined in Section 6), then the above inequalities would not hold anymore. If, on the other hand, the patenting decision itself were a function of the size of the expected gain from the invention, as noted in the text, then the situation would be somewhere in between.

6. To the extent that some patents arise in the development stage, they would also be related to R&D with only a short lag.

7. Sirrilli (1987) shows that in small firms in Italy (less than 100 employees) over a third of the inventors (36 percent) come from production and quality and control activities, while in the large firms (employees >1000) only 11 percent of the inventors come from this category. The proportion of patents originating in formal R&D rises from 39 percent in small firms to 63 percent in the large ones with the rest (25 and 26 percent) being in the more ambigious "design" category. Similar conclusions can also be infered from Kleinknecht (1989), who reports a significant underestimate of R&D activities

in small firms by the conventional data collection methodology.

8. See Jensen 1987 for similar results using new chemical entities rather than patents. For contradictory evidence, using other measures, see Scherer 1985b, Chapter 11 and Acs and Audretsch, 1989.

9. Based on unpublished tabluations of the Office of Documentation Information at the U.S. Patent Office.

10. See Pakes and Simpson 1989 and Schankerman, 1989 and 1990 for an extension of these results and Trajtenberg 1988 for another approach to the same problem.

11. This equation would hold exactly in a world in which all assets were fully traded in the same market. More generally, such an equation is valid in a multi-capital setting only under very stringent conditions, such as the linear-homogeneity of the profit function. See Wildasin (1984) and Hayashi and Inoue (1989) for more discussion.

12. See Hall 1989, Chapter 2, for similar results.

13. This is, of course, related to Schmookler's distinction between patents classified by industry of origin versus industry of use. "Who Does the invention" depends more on supply considerations. "For whom the invention is done" is more likely to be affected by demand shifts.

14. They estimate the value of the patent rights. I assume that the value of the underlying innovation is proportional to its patent right value and highly correlated to it.

15. There are two major problems in using this procedure to estimate the variance of the news in the economic value of patents held by the firm: the first is that the distribution estimated by Schankerman and Pakes is a distribution of the value of patent rights, which may vary less than proportionally with the true economic value of the associated invention to the firm. The second problem probably goes in the other direction: some of the change in the firm's patent value this year may not be news, and thus may have already been incorporated into the market value at the beginning of the year. Allowing for some predictability of patent numbers would only reduce such fractions futher, multiplying them essentially by $1-R^2$ of the prediction equation. (See Griliches, Hall and Pakes (1990) for a more detailed discussion of this and related issues.)

16. An alternative approach to this question is developed in Griliches, Hall and Pakes by modelling the components of variance in stock market value surprises explicitly as functions of current and past patenting and R&D activity, allowing one to estimate also the contribution of revisions in past patents values to current changes in market value. Though the resulting estimates are rather imprecise, since they are based essentially on fourth moments of the data, they do imply that the variance in the news about the value of patents (current and past) could account for about five percent of the total variance in market value surprises, a number which may look low but is actually a number as high as any that have been found in other studies of market value revisions. Only about one-fifth of this, however, can be attributed to news associated with <u>current</u> patent applications.

Table 1

Selectivity of Firms in 1976 Cross-Section by Size of Employment

Employment	Number of firms in cross-section[a]	Number of firms in Census of Enterprises[b]	Ratio
< 10	24	16,000	.0015
10-99	301	14,300	.021
100-999	952	9,000	.106
1000 +	1267	1,900	.667
Total	2541	41,200	.062

[a] With good employment data. Computed from the data used in Bound et al (1984).

[b] In comparable manufacturing industries. From U.S. Bureau of the Census, Enterprise Statistics 1977, General Report on Industrial Organization, Table 3, pp. 152-198.

Table 2

The Stock Market's Relative Valuation of R&D and Patents

Dependent Variable: log (q)

SP/A	0.493 (0.165)	0.111 (0.094)	0.246 (0.082)
K/A		1.374 (0.182)	0.741 (0.152)
NR/A			11.99 (1.556)
R^2	0.027	0.125	0.258

V - market value of the firm

A - Total net assets at replacement cost

Q - V/A

K - "Stock" of R&D using 15 percent depreciation rate

NR - "News in R&D": Current R&D less depreciation of the R&D stock

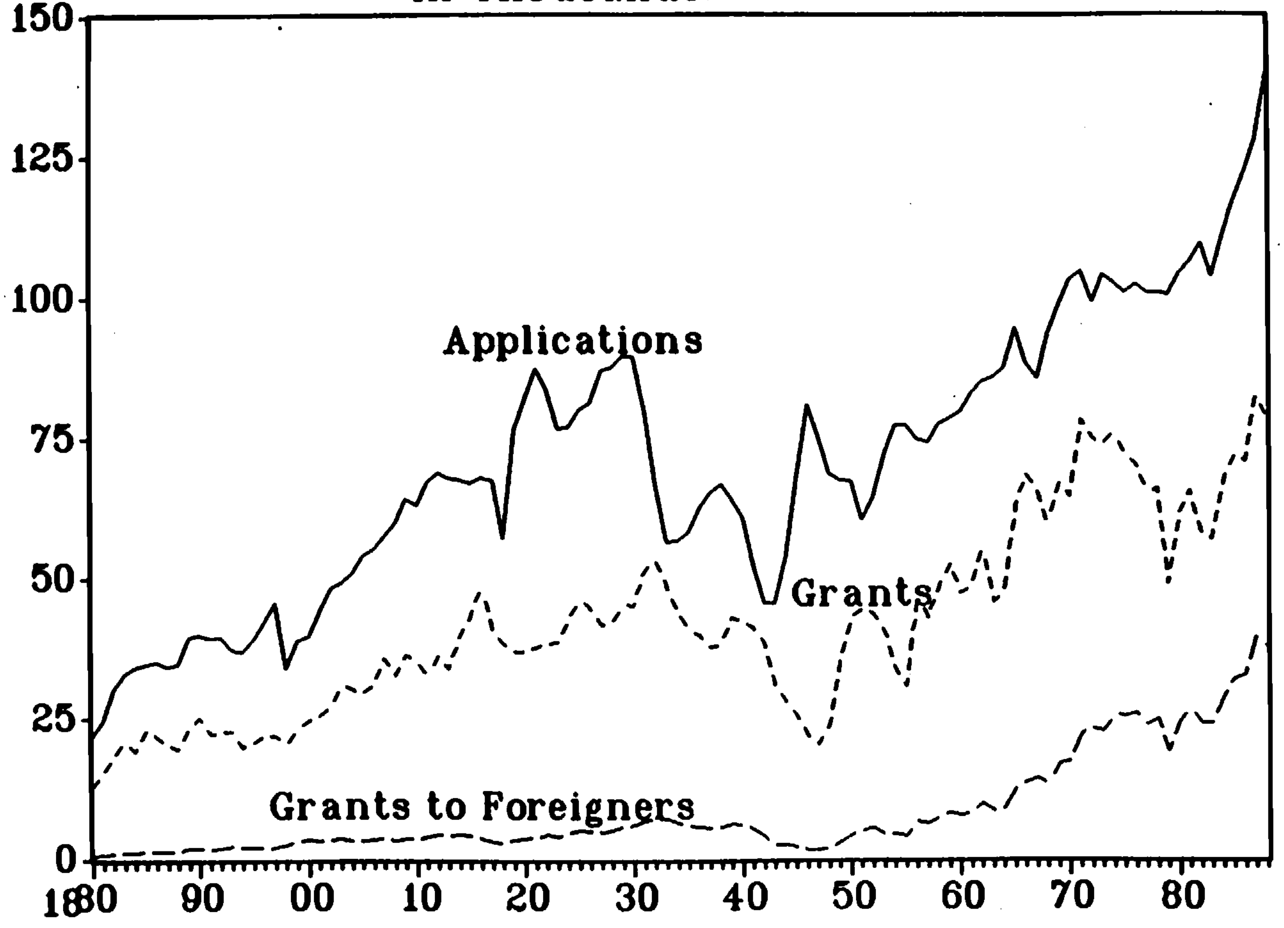
SP - "Stock" of Patents using 30 percent depreciation rate

N - 722. Mean of the dependent variable - -0.272, standard deviation - 0.697.

Heteroscedasticity-consistent standard errors in parentheses.
Matched by IND, 1980 Data.
All equations also contain an intercept term and the logarithm of Assets, whose coefficient was small but consistently significant, on the order of -0.03 (0.01).

From Cockburn and Griliches (1987), Table 3.

Figure 1: U.S. Patent Applications and Grants, 1880–1988, in Thousands.



Sources: U.S. Patent Office. See Notes to Table 2.

Figure 2: U.S. Domestic Patents and R&D, 1953–88, log scale.

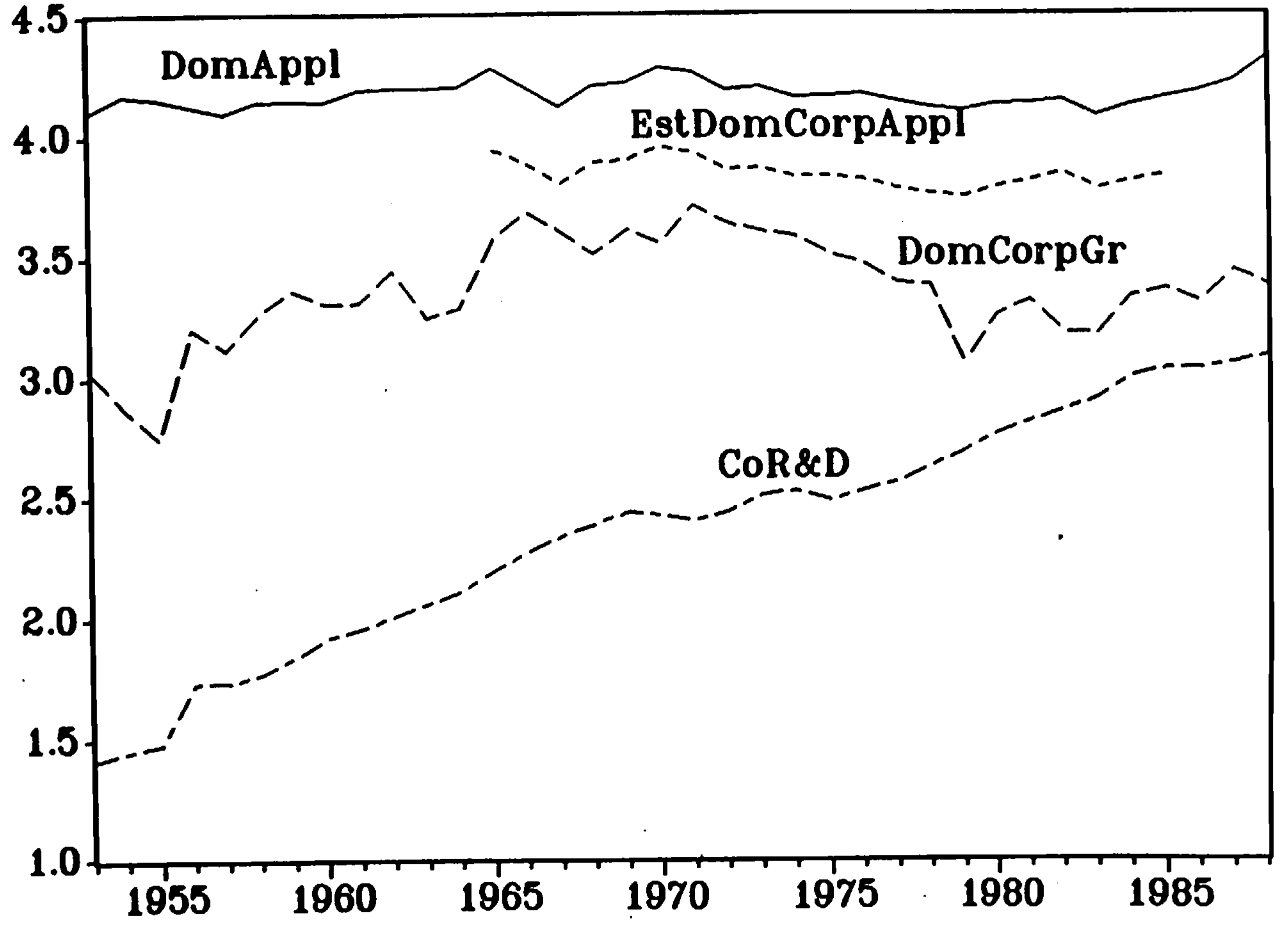
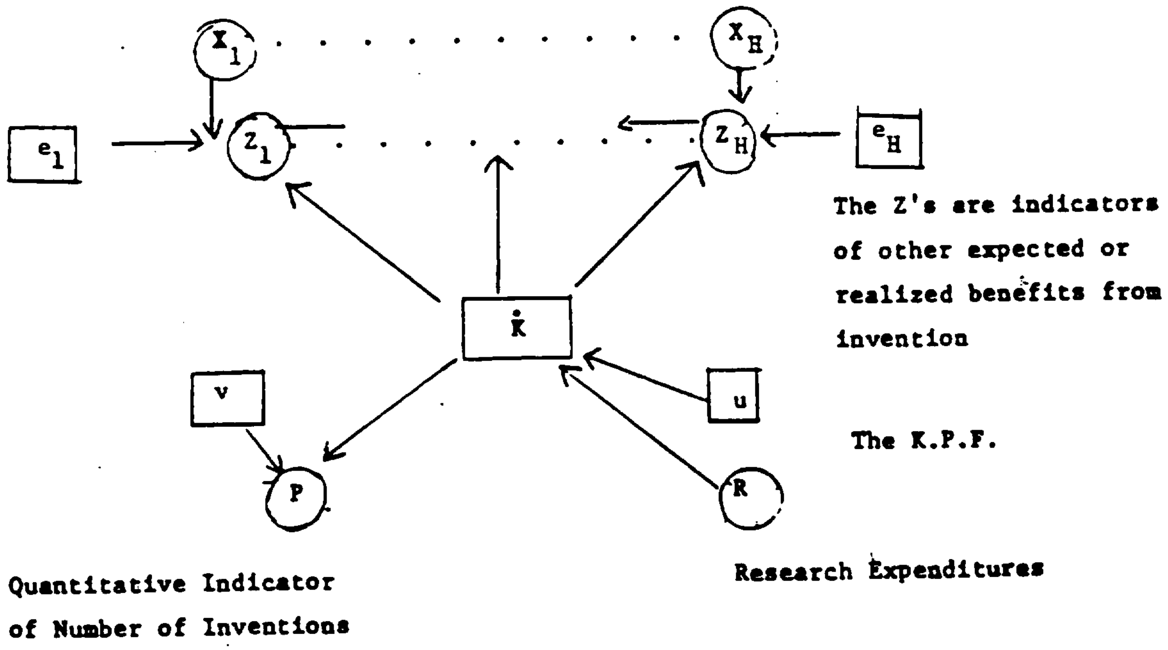


Figure 3

A simplified path analysis diagram of the overall model. Squares denote unobservable magnitudes and circles denote observable ones.



The Z's are indicators of other expected or realized benefits from invention

The K.P.F.

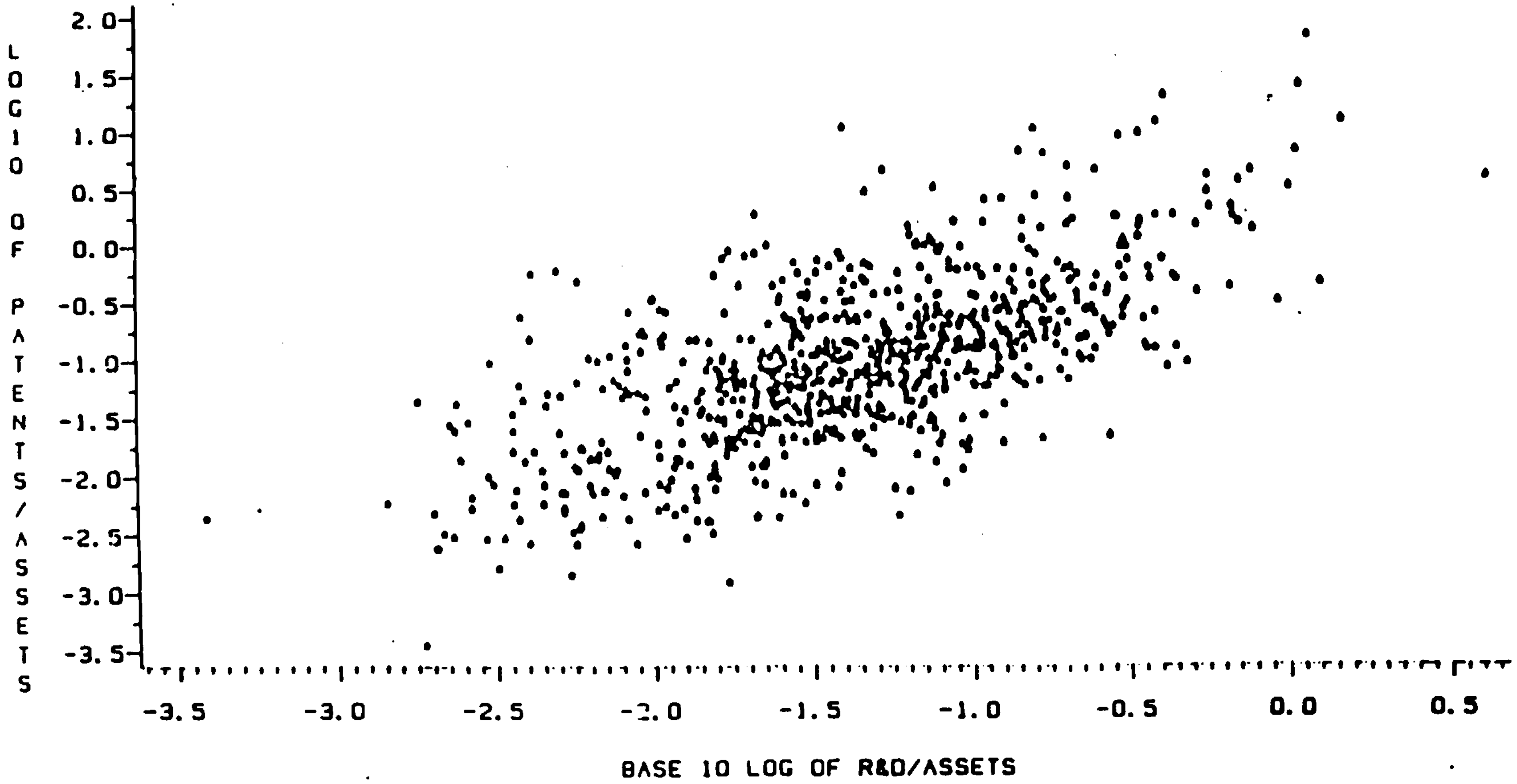
Quantitative Indicator of Number of Inventions

Research Expenditures

From: Pakes, A. and Z. Griliches, 1984, p. 56

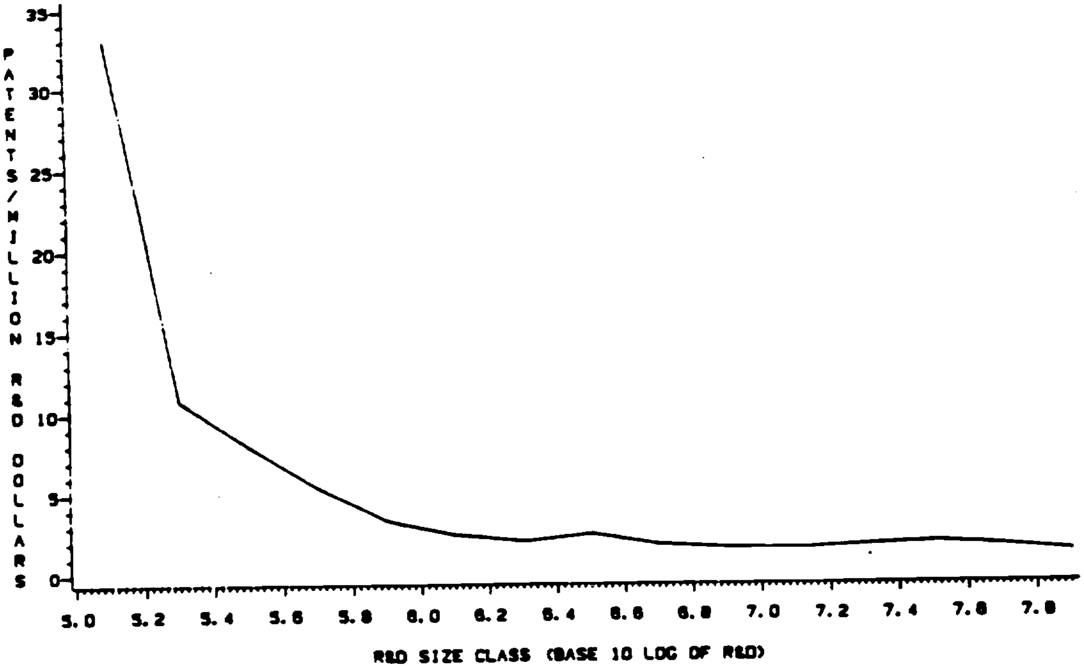
FIGURE 4

LOG OF PATENTS/ASSETS VS LOG OF R&D/ASSETS



From: Bound, J., et al, 1984, Fig. 2.4

FIGURE 5

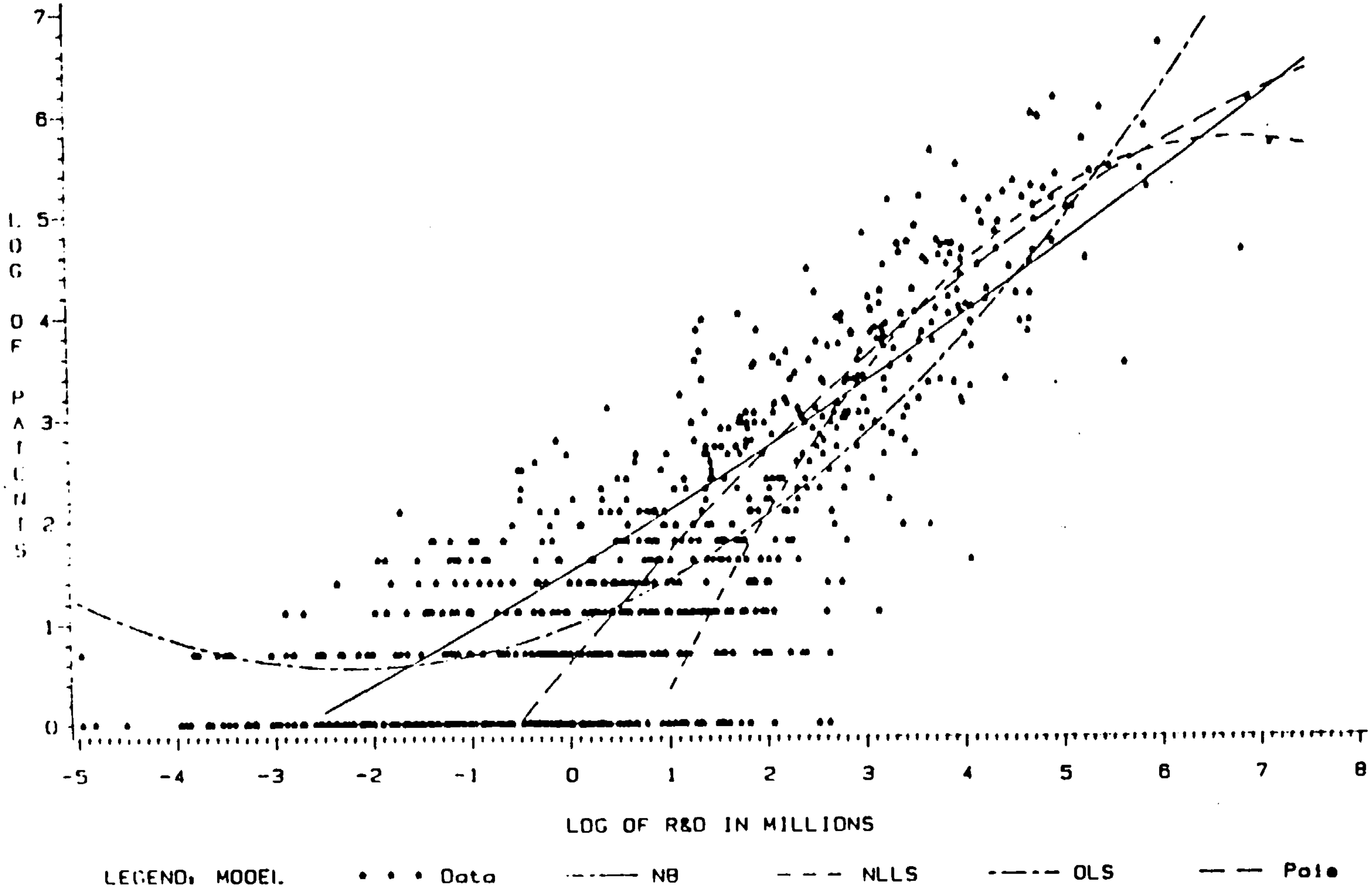


Patents per million R & D dollars by R & D size class for firms with both R & D and patents.

From Bound et al, 1984, Figure 2.6

FIGURE 6

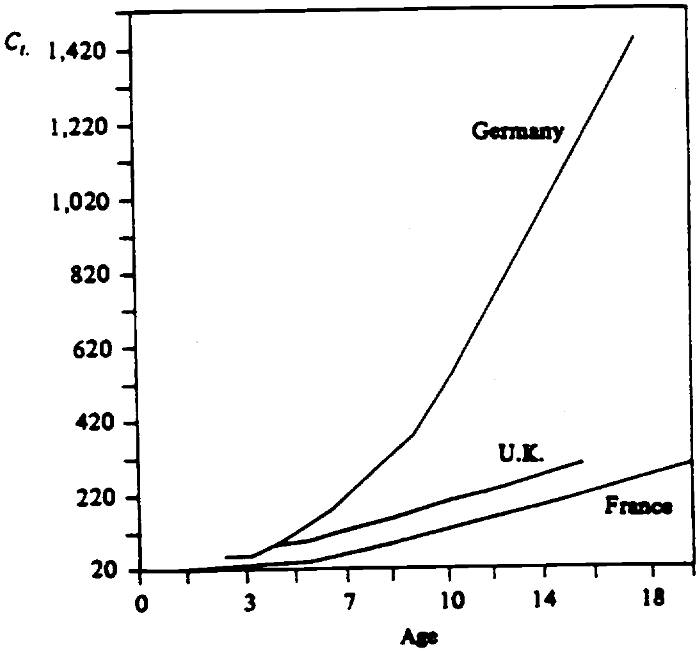
PREDICTIONS OF VARIOUS PATENTS MODELS



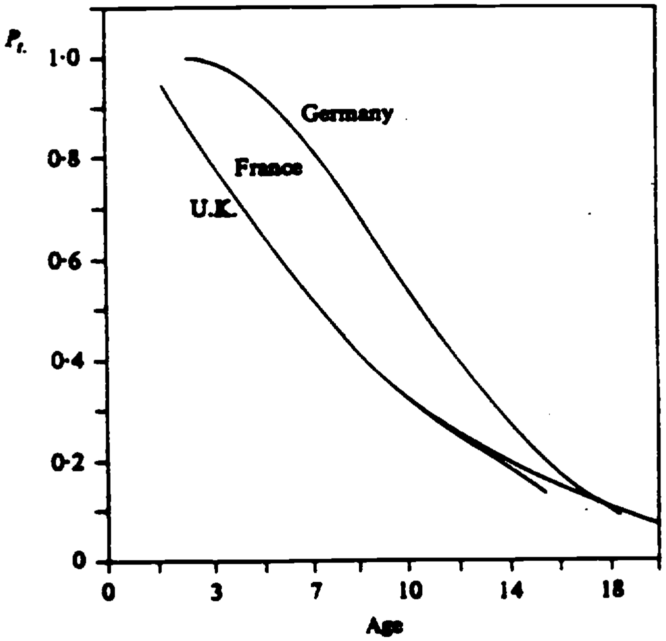
From Bound et al 1984, Figure 2.7

FIGURE 7

Renewal Costs and Renewal Proportions



Age paths of deflated renewal costs.*



Age paths of renewal proportions.

From Schankerman and Pakes, 1986, Figures 2 and 3