THE “SCIENCE OF SCIENCE POLICY”: REFLECTIONS ON THE IMPORTANT QUESTIONS AND THE CHALLENGES THEY PRESENT

Adam B. Jaffe

Fred C. Hecht Professor in Economics and Dean of Arts and Sciences, Brandeis University

Research Associate, National Bureau of Economic Research

ajaffe@brandeis.edu

Keynote Address

NSF Workshop on Advancing Measures of Innovation: Knowledge Flows, Business Metrics, and Measurement Strategies

June 6, 2006

THE “SCIENCE OF SCIENCE POLICY”: REFLECTIONS ON THE IMPORTANT QUESTIONS AND THE CHALLENGES THEY PRESENT

Adam B. Jaffe

Abstract

Developing the “Science of Science Policy” will require data collection and analysis related to the processes of innovation and technological change, and the effects of government policy on those processes. There has been much work on these topics in the last three decades, but there remain difficult problems of finding proxies for subtle concepts, endogeneity, distinguishing private and social returns, untangling cumulative effects, measuring the impact of government programs in a true “but for” sense, and sorting out national and global effects. I offer observations on how to think about these issues.

Keywords: Technological change; innovation; technology policy

JEL Codes: O300; O380

THE "SCIENCE OF SCIENCE POLICY": REFLECTIONS ON THE IMPORTANT QUESTIONS AND THE CHALLENGES THEY PRESENT

Adam B. Jaffe

It is a pleasure and an honor for me to be asked to give this talk. I have been thinking about these issues now for about 25 years. Being allowed to talk about the big picture without having to actually present any results is a perfect assignment for a Dean.

Many of you knew my teacher and mentor, Zvi Griliches. In the great Data Library in the Sky, Zvi is smiling down on us today. For all of the time that I knew him, he devoted significant energy to arguing that the public investment in the data needed to understand the economy in general and the process of innovation in particular were inadequate, and that the social return to investing in a better foundation for economic policymaking would be high. I'm sure he is pleased that people seem finally to be listening, though in typical Zvi fashion he will not grant even partial credit until we demonstrate that we have actually accomplished something.

The starting point for this discussion remains Zvi's 1979 paper, "Issues in Assessing the Contribution of Research and Development to Productivity Growth." I tell my students that this paper is to the economics of innovation what Keynes' General Theory is to macroeconomics—virtually all of the important ideas that people talk about even today were there in some form. If any of you have not read it, you should. (Another useful reference is Zvi's Chapter on "Data Issues" in the Handbook of Econometrics.)

My first foray specifically into the “metrics” question came with my 1998 paper on “Measurement Issues” in Lew Branscomb’s book Investing in Innovation. This paper was written in the immediate aftermath of the passage of the Government Performance and Results Act (GPRA). There was much discussion at that time that science and technology agencies should somehow be exempt from the Act’s requirements of quantitative assessment of the outputs and outcomes of government programs, on the grounds that the relevant outcomes are too intangible to quantify. I argued against that viewpoint, and advocated systematic efforts to develop multiple and diverse quantitative metrics. A number of the points I will make here are ones I made there. I might also note that this paper is among the least highly cited that I have written. This is but one indicator that the noble and notable emphasis we are now trying to put on measurement issues must flow against the current of the higher professional value attributed to theory and model-building.

This is not to say that I believe theory and model-building to be unimportant. Rather, I think that joint consideration of models and data construction strategies is necessary to produce good data and good models. I will return to that theme below.

Overview

I will touch on a hodgepodge of issues that I think we should have in mind in some way as we collectively devise strategies for building the knowledge base for science policy. These are:

1. Our ultimate objectives in devising metrics of innovation;

2. The role of proxies;

3. Endogeneity, unobservables and selection bias;

4. Private versus social rates of return and spillovers;

5. The cumulative nature of innovation and other dynamic effects;

6. The role of government and the special needs of data to support program assessment;

7. The U.S. as part of a global innovation system; and

8. The dangerousness of a little knowledge.

Metrics for What?

Our first objective in devising metrics is to undertake social science research. We want to understand the sources and mechanisms by which society produces new technology, and the economic and social consequences when it does so. To do this, we need measures of the inputs, both human and material, to the process of technology development. For both human and material inputs, these include stocks (human capital and accumulated knowledge; equipment) and flows (hours of scientist/engineer work, chemicals). Since we wish to understand the mechanisms, we also need measures of intermediate products, such as new knowledge that is produced by research and is in turn an input into the development of new products or processes. Next we need measures of the output of the technology development process, i.e. new ways of doing things or new products. Finally, to use the language of the GPRA, we also need measures of outcomes of the innovation process, i.e. the benefits that society derives from having new technologies in use. These include various concepts of productivity, and also more specialized measures of benefits, such as reduced morbidity and mortality in the case of medical technologies.

A second objective is the assessment and evaluation of public policies. I will address this further below, but the primary data need of assessment that goes beyond those identified in the previous paragraph is data on individuals, firms and other institutions tracking their involvement (or lack thereof) with the programs that we wish to study.

Finally, we collect innovation metrics in part for the purpose of spotting trends that may signal changes in the system or emerging issues that require attention. It seems to me that the data on inputs, intermediate and final outputs, and outcomes desired for research purposes are the same data that one would want to monitor for emerging trends.

Proxies, correlates and the like

All social scientists engage in various degrees of looking under the lamppost for the watches they dropped in the street, because the light is better there. The only alternative is to build more lampposts, but we will never cover the whole territory. So our metrics will always be, at best, imperfect. There are some generic strategies we can use to deal with this imperfection (Griliches, 1986). The most obvious is to use multiple metrics, preferably chosen so that there is reason to believe that the errors in the different metrics are uncorrelated with each other. Another strategy is to examine the phenomena we care about at multiple levels of aggregation, or to use "long differences" or other averaging methods to mitigate the variance due to measurement error. Of course, we have to worry when we use these approaches that the measurement error is unsystematic and hence subject to such reduction through averaging.

In my previous paper I listed a number of characteristics that make for “good” proxies. These include:

1. A high signal/noise ratio;

2. Errors that are unbiased and uncorrelated with other phenomena of interest;

3. Linearity (or another known functional relationship) between the proxy and the underlying phenomenon;

4. Stability over time in the relationship between the proxy and the underlying concept;

5. Stability across settings (institutional, geographic) in the relationship between the proxy and the underlying concept (or variation that is itself subject to proxy);

6. Low susceptibility to manipulation; and

7. Subject to consistent measurement at different levels of aggregation (geographic and institutional).

The difficulty of assessing the quality of proxies is easily illustrated with data that I have used extensively, namely the citations or references to earlier patents that appear in patent documents. The number of citations made by the average patent granted in the U.S. has been rising steadily for several decades. If we take such citations as a proxy for the contribution of previous technological developments to the development of new ones, the rising patent rate could be interpreted as an increased “fecundity” of the existing knowledge base in generating new developments. But alternative interpretations of the observed trend include:

1. The number of patents in the pre-existing base has been rising, so they are not each becoming more fecund, there are just more of them.

2. Knowledge is diffusing faster than it used to, so a larger fraction of the pre-existing knowledge base is known and available to the typical inventor at any moment in time.

3. The search capabilities of the patent office have improved, so even if the inventors don't know any more than they used to, the examiners are finding more citations to include.

4. Patent practices have changed so that the average patent embodies a larger "chunk" of new knowledge, leading on average to more citations made per patent but not more per unit of new knowledge.

A conceptually related set of issues arises from the fact that at a given moment in time, U.S. patents make on average more citations to earlier work than do the patents of other jurisdictions around the world. Again, the issue that we need to try to understand is the extent to which this is an artifact of the process that generates the proxy, versus differences in the underlying phenomenon of interest. This can never be resolved absolutely, and typically even partial resolution requires identifying assumptions that may not be testable. For a systematic framework for thinking about these issues specifically with respect to patent citations, see my book with Manuel Trajtenberg (Jaffe and Trajtenberg, 2002).

<u>Endogeneity and All That</u>

Everything is related to everything else, and typically through more than one mechanism. This makes the core task identified above—understanding the sources of innovation and the mechanisms by which it is brought forth—very hard. Conclusions about causation can only be drawn conditional on identifying assumptions, meaning that there is an element of untestable belief behind virtually everything we think we know about the process. A couple of examples illustrate the point.

I first "made my name" as a student of technological change with my dissertation work, which showed that firms that perform R&D in technological areas in which a lot of R&D is performed by other firms enjoy higher R&D productivity ("measured" in terms of patents and productivity improvements), all else equal. I interpreted this as evidence of "spillovers" of knowledge benefiting firms' technological neighbors. But conceptually it could also be evidence that certain regions of technology space are more fruitful at a moment in time; firms are attracted to such regions, so we observe a higher level of R&D activity in such regions, and firms' R&D is more productive precisely because the area is fruitful. (The standard jargon is that variations in "technological opportunity" explain both variations in R&D intensity and in R&D productivity.) Hence the positive correlation between firms' R&D productivity and the amount of R&D in their neighborhood is not causal, but simply a result of the endogeneity of the locus of R&D activity itself. My work dealt with this possibility by including in the regression analysis control variables for location in technology space. The identifying assumption that is necessary to maintain the "spillover" interpretation is that these controls—which were essentially arbitrary in the level of detail at which they captured location—were adequate to capture the variations in research productivity associated with the choice of research topic, so that the observed relationship of productivity to R&D of other firms is over and above the technological opportunity effect and hence attributable to spillovers. Subsequent work using a number of different models and approaches, and different datasets, has confirmed the spillover phenomenon, so I believe that it is real. But "belief" still plays a role in that conclusion.

The second example is the widely observed phenomenon that the private internal rate of return to R&D investment appears to be considerably higher than the private internal rate of return to investment in fixed capital. Now, there are a number of "real" reasons why this might be true: it's riskier (although it would seem to be diversifiable risk, which really shouldn't justify a risk premium), and information asymmetries might explain why this market does reach the equilibrium that would erase the difference in the rates of return. There are also "garden variety" measurement problems with R&D and its return, in particular the problem that investment in equipment that is used for R&D is not properly tracked in the data. But there is also a big endogeneity problem: all kinds of unobservable attributes of firms produce unobservable variation in the likely productivity of R&D. We observe firms that do a lot of R&D making a lot of money, all else equal, but this may be because both the level of R&D and the amount of money made are both being driven by these unobservable variations. (Note that this problem also infects estimates of the *social* rate of return to R&D, but only to the extent that the social rate of return is correlated with the private rate of return. It does not undermine the evidence of a large *gap* between the private and social rates of return, since profit-maximizing firms are not responding to this gap.)

Partial solutions to this problem come in roughly three forms. The most commonly used is structural modeling, in which the determinants of endogenous variables are themselves brought within the model. Assumptions about functional forms and exclusion restrictions "identify" statistically the causal relationships of interest, purified of any effect of reverse causality or left-out variables. Of course, these identifying assumptions are the last refuge of econometric scoundrels. Occasionally, people are very clever and come up with instruments whose exogeneity really cannot be questioned (such as Vietnam-era draft lottery numbers). More commonly, instruments are either based on dubious exclusion restrictions, or are uncorrelated with the phenomena of interest. (I once had the brilliant idea to use dummies for which Big Eight accounting firm was used to certify a firm's annual financial statement as an instrument for measurement error in R&D. Probably exogenous—but also, as it turned out, uncorrelated with measured R&D.)

A second approach is to try to eliminate the effect of unobservables by identifying "control groups" who differ with respect to some endogenous variable, but whom we believe are similar with respect to the important characteristics even though we cannot actually observe the characteristics. One version of this approach includes "fixed effects" and "difference" models, in which we assume that the important unobservables are constant over time for a given firm or individual. Another version is "matched pairs," in which the behavior or performance of individuals with an observable characteristic of interest is compared to the behavior or performance of randomly selected individuals who are identical to the first in all observable respects.

Finally, in some circumstances, it may be possible to reproduce experiment-like conditions so that variables of interest can be thought of as uncorrelated with left-out variables. Randomized trials have not been used to evaluate science and technology policy, but they are considered acceptable in other public policy assessment arenas, such as job training and health insurance. As discussed further below, however, despite having advocated randomized trials for science/technology programs in my 1998 book chapter, I have more recently come to the view that approaches that fall under the "structural modeling" rubric can tell us everything we might hope to learn from randomized trials, with far less political pain.

As already suggested, solutions to this problem are at best partial, in the sense that they will always depend, on some level, on untestable identifying assumptions. But if these assumptions are plausible, and if results tend to be confirmed with different approaches that rely on *different* assumptions, we gradually accumulate understanding.

Private versus Social Rates of Return and Spillovers

The gap between the social and private rates of return to investment in knowledge and technology is the primary reason why innovation is a topic of policy concern. Yet measurement of this gap, and of spillovers, is very difficult.

Spillovers related to innovation come in two flavors: "technological" or "knowledge" spillovers, and "pecuniary" or "rent" spillovers (Jaffe, 1998a). The former corresponds to the phenomenon that your doing research on a topic generates knowledge that I may use to reduce the cost or increase the success rate of my own research. It is a technological externality, meaning that the total cost to society of producing new knowledge is reduced. The latter refers to the likelihood that your economic exploitation of new knowledge that you create is likely to leak some of the economic benefits to your customers (in the form of consumers' surplus that you do not capture) or your competitors (if they copy your innovation and earn profits as a result). It is a pecuniary externality, meaning that the benefit to the spillover recipient is offset by losses to the spillover generator; society as a whole does not gain. Knowledge spillovers are important to endogenous growth models, because they are a source of increasing returns in society as a whole. Pecuniary spillovers do not generate increasing returns, but they are still important from a policy perspective, because they still generate a gap between the private and social returns to investment and hence suggest socially suboptimal investment rates in the absence of policy intervention.

There are three categories of approaches to measuring spillovers. One is to look for correlations, as I did in my dissertation discussed above. If the economic success of one agent or group of agents is correlated with the actions of some other agent or group of agents, we may be willing to infer that spillovers from the latter's actions to the former's success is the explanation for the correlation. Second, we can look for proxies for the spillover flow itself, such as citations. Ideally, we combine these two approaches and show that patterns of citations or some other proxy for the flow of spillovers are consistent with patterns of spillovers inferred from correlations between different agents' actions and performance.

Finally, we can infer spillovers by measure rates of return at different levels of aggregation. As we move from measuring impacts at the level of individual firms, to industries, to society as a whole; or from cities, to states, to countries, to the whole world, we should be "capturing" an ever larger fraction of the spillovers being generated by some act of knowledge creation. This means that if spillovers are important, we should observe higher rates of return (however measured) as we increase the level of aggregation at which the measurement occurs. Systematic exploitation of this phenomenon is hindered by the frequent difficulty of measuring the rates of return in comparable ways at different levels of aggregation.

Dynamics and the cumulative nature of knowledge

There tend to be long and variable lags between when inputs are brought to bear, when outputs are produced, when outcomes are realized. This makes empirical research difficult and makes it hard to understand relationships. It's a bummer, but we have to live with it. Longer time series are the only way that these questions can really be answered.

For many purposes in understanding the innovation process, stocks of knowledge – the accumulated quantity of previous inputs or intermediate products – are often as important as current flows. To estimate stocks, one must account for depreciation or obsolescence of past investments. The appropriate depreciation rate depends on whether you are thinking about private or public knowledge stocks. From a private perspective, if your competitor finds an alternative to your new product, this makes your knowledge obsolete, but from a public perspective, your "obsolete" knowledge may still be productive. Finally stocks, such as human capital, move around over time. If we do not have good information on these movements, it may be impossible to keep track of the relevant quantities for a given institution or region.

Assessment of Government Programs

Measuring the impact of government programs on innovation or knowledge creation raises all of the problems discussed above. In addition, there are several additional issues created by the attempt to identify the impact of government intervention in the economy.

First, we need to worry whether we are increasing knowledge or some other "real" objective, or merely increasing our indicator of it. When we "reward" certain behavior, as with the R&D tax credit, firms will find ways to report that they are doing more of the rewarded activity, whether their real behavior has changed or not. Second, we need to take account of endogenous responses of the economy that mitigate or offset the direct effects of government intervention. After all, economic systems tend to be in some sense in equilibrium; when disturbed they will tend back toward that equilibrium. If, for example, we do succeed in increasing the amount of R&D effort undertaken, this will tend to bid up the price of R&D inputs, so that the increase in real R&D activity will be somewhat less than the increase in R&D spending, at least in the short run (Goolsbee, 1998)

Third, when the government supports a research project or the research program of a firm, and that project or firm is observed to be successful, we still need to worry about whether that success was in some way fostered by the government, or whether the government is simply could at guessing who is going to be successful regardless of support. To know if policy is fostering success, one needs some kind of counterfactual to the historical observation or a baseline against which to compare, and that rarely happens. I've looked at the various publications that NSF puts out, some explicitly responding to the GPRA mandates to measure outputs and outcomes. These publications have lots of good and interesting information, but none of them present this information relative to an explicit counterfactual. As a result, we don't know whether the results of NSF supported research would or would not have come forth otherwise. The reports assume that the wonderful results would not have occurred but for NSF support, and that is probably true to a significant extent. But we really should be trying to actually measure the impact relative to the "but for" scenario. To do so requires understanding how government interacts with the system it tries to effect; we need modeling and statistics to account for that interaction.

Fourth, we need to distinguish between average and marginal effects. Suppose we're willing to ignore the problem of measurement relative to a counterfactual, and accept that if there were no NIH and never had been, we would all be worse off in terms of the state of medical care in the U.S. Even if this is true, however, it doesn't flow logically that doubling the NIH budget would double the results, or even necessarily increase our state of knowledge significantly. That kind of question was never really posed before we decided to double the NIH budget in the late 1990s. From a dean's perspective, it is clear that doubling the NIH budget over a short period of time was a bad idea. It was predictable that after the doubling, there would be a period of flat or declining spending. The result was a large influx of new people and capital into the NIH research "game," all of whom are now scrambling to cope with the now less generous environment. Slower growth would surely have been better. But we think big increases on the margin are a good thing because the enterprise as a whole has historically done good things. Now we are talking about big increases in spending on research in the physical sciences, again without any attention to what the effects will be on the margin, or what the new long-run equilibrium will be.

Finally, we need to distinguish net returns from gross returns: It's easy to look at major technology programs such as NASA or the Internet and say that there have been large social benefits. But the investments have also been large in many cases. Big benefits imply a high social return only if those benefits are big *relative to* the expenditure. Often, that comparison is simply not made.

U.S. as part of global innovation system

The role of innovation and new technology in the global economic system is very much in the news. Consider the following developments:

- China is increasing PhD production greatly.
- The number of foreign PhDs in U.S. schools is dropping.
- Foreign firms continue to open research facilities in the U.S.; and
- U.S. firms continue to open R&D facilities in other countries like China.

For each phenomenon, one can tell a good story about why it is good for U.S. citizens, and one can also tell a story about why it is bad for the U.S. Which is right rests on the relative significance of spillovers on the one hand, and international economic competition on the other. It would be nice to know whether any of these phenomena should be encouraged or discouraged to maximize benefits to the U.S., but we don't know enough to say. In particular, we know very little about what happens to foreigners who get PhDs in the U.S., or about the flow of people into and out of foreign-owned R&D facilities, either here or abroad. If we are going to avoid neo-mercantilist policy responses to these phenomena, we need to study them.

A Little Knowledge is a Dangerous Thing

A few years back, Ashish Arora and Alfonso Gambardella took a pretty careful look at the impact of NSF funding of economics research, using a "differences in differences" approach to see whether people who got funding increased their publishing relative to their historical base more than people who applied for funding and were denied (Arora and

Gambardella, 1996). The results suggested that NSF funding did increase publication for people early in their career, but not for more senior researchers. The paper was never published, but someone on capitol hill got a hold of it anyway and tried to use it to criticize NSF. This is definitely a cautionary tale. No matter how carefully we do our work and qualify the conclusory statements we make, we cannot stop others from misusing our results. I am sympathetic to those who would therefore prefer not to ask the awkward questions. But I would insist that we must do so. If we don't, then we are just another special interest group, pushing for more funding because we "believe" that what we are doing is valuable.

I am truly gratified that the Science of Science Policy initiative is creating a framework for renewed attention to data needs and research related to innovation and new technology. Surely Zvi Griliches is looking down on us from that great seminar room in the sky and smiling. I will confess to some concern that this will be another flash in the pan, that efforts will be geared up and then a new administration or a new problem du jour will shift the focus to something else. We need to try to make this a sustained effort. If we could spend a decade focusing on these issues with our current level of attention, I believe (but cannot actually prove….) that we would know a lot more than we do now about how to maximize the social benefit of public policy in this area.

References

Arora, Ashish and Alfonso Gambardella (1996), "The Impact Of NSF Support For Basic Research In Economics," Heinz School of Public Policy Working Paper, Carnegie-Mellon University.

Goolsbee, Austin (1998), "Does Government R&D Policy Mainly Benefit Scientists and Engineers?" American economic review, 88: 298

Griliches, Zvi (1979), "Issues in Assessing the Contributions of Research and Development to Productivity Growth," Bell Journal of Economics, 10: 92-116

Griliches, Zvi (1986), "Economic Data Issues," in M. Intrilligator and Z. Griliches, eds., Handbook of Econometrics, Volume III, Amsterdam: North Holland

Jaffe, Adam B. (1998), "Measurement Issues," in L. Branscomb and J. Keller, eds., Investing in Innovation: Creating a Research and Innovation Policy that Works, Cambridge: MIT Press

Jaffe, Adam B. (1998a), "The Importance of 'Spillovers' in the Policy Mission of the Advanced Technology Program," ***Journal of Technology Transfer***, Summer 1998

Jaffe, Adam B. and Manuel Trajtenberg (2002), Patents, Citations and Innovations: A Window on the Knowledge Economy, Cambridge: MIT Press