Economic performance measures for evaluating government-sponsored research

By: Albert N. Link

Link, A.N. Economic performance measures for evaluating government-sponsored research. *Scientometrics* **36,** 325–342 (1996). https://doi.org/10.1007/BF02129598

Abstract:

The purpose of this paper is to discuss, in general terms, evaluation issues related to government-sponsored research and to describe and critique the usefulness of economic performance measures for evaluating such activity. Herein is presented an overview of the economic justification for government-sponsored research and the rationale for its evaluation. Also, fundamental evaluation methods are described. The paper ends with a recommendation that benefit-cost analysis may be the most appropriate economic performance measure when evaluating government-sponsored research if used cautiously and with an understanding of its inherent subjectivity.

Keywords: economic performance measures | government-sponsored research | evaluation | benefit-cost analysis

Article:

Introduction

U.S. government initiatives and policies directed toward innovation have historically focused on research activity as the primary investment. For example, the U.S. Navy's sponsored research programs date as far back as 1789, and the U.S. Department of Agriculture's involvement in the land-grant college system dates from the mid-1800s. Since World War II, direct government support of both research and development, as well as other aspects of the innovation process, increased dramatically in response to military needs and to the government's responsibility for enhancing overall research capabilities as outlined in the National Science Foundation Act of 1947. This public support has historically been focused in two areas. One area is basic research, which is an investment in the nation's science base; and the other is applied research and development, which even when it has a defense orientation still enhances the overall research capabilities of individual firms.

Because of the massive economic growth resulting from World War II and the vulnerability of the U.S. economy to a post-war economic downturn, the government developed a number of strategies for harnessing the economic potential of science and technology. Prominent among these was a recommendation by *Vannevar Bush,* then Director of the Office of Science and

Research Development. He urged the government to continue to invest in fundamental science as a cornerstone for maintaining American preeminence. *Bush's* plan called for establishing a national program for both basic research and graduate education that would keep the nation ready to meet future military or economic threats. Combined with plans by the military to maintain its research enterprise, his proposal made possible the establishment of the first U.S. science policy in the late 1940s-a policy separate and distinct from any commercial technology policy and separate and distinct from a science and technology policy dominated by national defense. The *Bush* plan assumed that investments in science would, in a direct and inexorable fashion, result in long-run technological and economic success for the nation.[1]

During the 1940s, in addition to increasing its investment in basic science, the nation also began developing its first systematic technology policy, chiefly in connection with civilian use of nuclear power. Congress established the Joint Committee on Atomic Energy to oversee nuclear policy, and it also established the Atomic Energy Commission to implement that policy. With these actions, the United Stated developed a *de facto* science and technology policy in this one area of future industrial activity. Here was the beginning of a national technology policy, a policy that flowed from national needs at the energy, commerce, and defense nexus.

The traditional assumptions of science policy – *Bush's* notion that economic growth would inevitably spring from basic research as well as the prevailing wisdom that applied research and development are important – vitiated any need for the evaluation of the social impact of research. But, as policy began to move to civilian technology applications, it became clear that the move from laboratory to market was not inexorable and that certain research-based investments would likely have greater payoff than others. This realization of the need both to justify the role of government in support of research, and to evaluate government-sponsored research as an investment in the innovation process that has calculable returns, essentially marked the dawn of public sector research and development evaluation practices.

While historians of technology often claim that our nation's previous military success was due in large part to the technological superiority of American weaponry that developed from its historic commitment to government-sponsored research, and while policy scientists frequently contend that the single most important factor associated with America's competitiveness problems during the late 1970s and 1980s was that it had lost its technological edge as a result of declining industrial research and waning government support for innovation, the fact is that there is little quantitative evidence about the effectiveness of public sector research in support of economic and technological growth, in general, and competitiveness, in particular. As policymakers in the United States and in other industrial nations enter an era where government support of research is being re-evaluated, and in some instances scrutinized, such knowledge becomes increasingly important. This thus raises questions about evaluation methodologies and the applicability and relevance of alternative economic performance measures.

The purpose of this paper is to discuss, in general terms, evaluation issues related to government-sponsored research and to describe and critique the usefulness of economic performance measures for evaluating such activity. Specifically, in Section II the economic justifications for government-sponsored research is set forth and the rationale for its evaluation is discussed. In Section III, fundamental evaluation issues and methods are presented. These issues and methods apply not only to government-sponsored research but also to private sector research and development. In Section IV, the evaluation issues and methods presented in Section III are discussed with emphasis on this relevance to public sector activities. In Section V, some summary observations are offered.

Evaluation issues related to government-sponsored research

While there is *political* history associated with every government-sponsored research program, there is also a set of *economic* justifications for the public sector to be involved both in the funding and conduct of research. This justification rests on the economic principle of market failure.

Market failure occurs when society's costs and benefits are not appropriately balanced. Market failure can arise for a number of reasons. In the case of research, and especially in the case of basic research, market failure is often a result of features intrinsic to the production of knowledge. There is sufficient evidence to suggest that a market economy will underinvest in the production of knowledge because private sector firms that produce knowledge through their research are unable to capture fully all of the benefits that arise from its creation.

If the research that is valuable to society is not being funded or conducted in the private sector, the market failure criterion may indicate the appropriateness of the research being funded by the public sector. However, determining that a research project or program is appropriate to the public sector says little about the subsequent issue of public accountability, that is, whether the project or program is conducted well.

The issue of accountability for public monies is certainly not new. Accountability underlies the motivation for the Competition in Contracting Act of 1984, Public Law 98-369. All of the regulations and guidelines outlined in the Act were done so to ensure that actions taken by government agencies are "in the public interest." Similarly, the Chief Financial Officers Act of 1990, Public Law 101-576, was passed on the basis of Congressional findings that "billions of dollars are lost each year though fraud, waste, abuse, and mismanagement among the hundreds of programs in the Federal Government." The Act's purposes include provisions "for the production of complete, reliable, timely, and consistent financial information for use by the executive branch of the Government and the Congress in the financing, management, and evaluation of Federal programs." Most recently, the Government Performance and Results Act of 1993, Public Law 103-62, was enacted on the basis of the Congressional finding that "waste and inefficiency in Federal programs undermine the confidence of the American people in the Government and reduce the Federal Government's ability to address adequately vital public needs." Accordingly, this legislation was intended to "improve Federal program effectiveness and public accountability."

It is inevitable that research managers, in both the public and private sectors, will become advocates for their own research programs. Watching results on a day-to-day basis and witnessing the benefits of projects and programs to which one is committed understandably lead managers, and other participants in projects and programs, to the *intuitive* conclusion that the projects and program are valuable. Regardless of the veracity of this conclusion, it may not be easily communicated to others. Thus, when political and administrative superiors ask; "But how do you *know* this research is effective?" managers often find themselves either dissembling or simply telling success stories. It is possible, through the systematic application of evaluation methods, to document value and thereby provide a clear, more precise response to the question of accountability.

The results from a research project or program evaluation also provide insights and information to improve both the managerial efficiency and the scientific judgment of those funding or performing research. It was explicitly noted in the motivation for the Government Performance and Results Act of 1993 that "Federal managers are severely disadvantaged in their effort to improve program efficiency and effectiveness, because of insufficient articulation of program goals and inadequate information on program performance."

Through systematic evaluations, research managers in the public sector, in particular, will be better able to identify the research outcomes that are most useful to their constituents, as well as any barriers that might inhibit fuller access to research results. With such information, public sector research managers will increase the credibility of their technological forecasts, which is critical to strategic planning as well as fundamental to the budget setting process.

Fundamental evaluation issues

Evaluation is partly art and partly science. The art of evaluation is in the creative tailoring of available techniques and methods to the problem at hand and to the data available. While there are many standard conventions in evaluation, evaluation studies are seldom routine. The practice of evaluation involves a series of judgments in the application of evaluation methods and the creation of research designs that are only partly standardized. The science of evaluation is in the various techniques and methods developed for application to policy evaluation problems, such as for analyzing data and presenting results. It is a primitive science, but nonetheless it includes some fundamental prerequisites for being a science such as replicability, precision, and testability of propositions.

Characteristics of an evaluation method

When choosing an approach to implement for the evaluation of a research project or program, it is important to specify the objectives of the evaluation and, if possible, to specify guidelines for methods and techniques to meet those objectives.

An economic evaluation emphasizes the assessment of the net benefits to society that are associated with the actual or potential output of the research project or program being evaluated. Several broad guidelines can be inferred from the academic and professional literatures, as well as from the practice of professionals, to assist in designing a method that will be both systematic and flexible enough to accomplish this evaluation objective. These include appropriateness, validity, replicability, ends-focused process, clearly conceptualized design, and comparability.[3-5]

The evaluation method should be appropriate to the evaluation's objectives and available resources. Thus, there is not an ideal evaluation approach apart from such considerations. The selection of an evaluation method should therefore be made only after the research being evaluated is fully understood. Appropriateness also includes consideration of what is well accepted in the evaluation community.

The evaluation method should be valid, especially internally. Evaluation can usually be thought of as a trade-off among various validity types, including not only internal validity (causal relations), but also statistical conclusion validity (veracity of statistical assumptions), external validity (generalizability), and construct validity (the extent to which concepts measured are the appropriate ones). Sometimes the strengthening of one type of validity is realized only at the expense of another. Thus, increased statistical power may require more simple constructs. In making these tradeoffs, one should expect that internal validity will, in general, be at the top of the hierarchy in evaluation objectives. If the causal explanation underlying an evaluation is incorrect, more precision and greater generalizability provide no advantage.

The evaluation method should be replicable. Quite often, research projects and programs continue in time· beyond the date of evaluation. To obtain a more complete evaluation, it is often desirable to replicate at a future point in time the method used in the earlier evaluation.

The evaluation method and attendant techniques should be ends-focused rather than means-focused. In any evaluation task, one of the more thorny issues is ensuring sharp distinctions between means and ends. Many evaluations have been undermined by the failure to stay focused on ends or by the inability to separate valued outcomes from the instruments for achieving those outcomes. The ability to separate means from ends is important for a wide variety of reasons, but one of the most important is that suppositions about the relation of means to ends may be erroneous. If only the means outcomes are evaluated, then one may mistakenly infer a success when in fact either there may be no success or the problem of concern may have been exacerbated.

The evaluation method should have a clear and consistent level of conceptualization. One level of conceptualization frequently used refers to the breadth of the unit under consideration. Thus, society is a broad level of conceptualization and household is a narrower one. In between these extremes are such levels as industrial sector and firm. Comparability as a criterion implies that the evaluation should not mix levels of conceptualization, making comparison between the units problematic.

The evaluation method should be comparable. Many evaluation approaches do not provide an outcome metric that is comparable across project and program types. Thus, it is difficult to compare research in different technical areas. When a comparable metric is used in the evaluation, the decision maker can more effectively evaluate diverse options.

Alternative evaluation methods

Appropriateness is an important guideline criterion for selection of an evaluation method. Not only is it important that the evaluation method be appropriate to the task; it is also important that the evaluation method be accepted within the community of evaluators.

The economics literature on evaluation can be used as a basis for developing a taxonomy for classifying alternative evaluation methods. The taxonomy used here draws from research related to alternative decision frameworks for the economic evaluation of social policies - environmental regulations in particular.[3,6] It is important to remember that the methods discussed herein are for the evaluation of completed research, as opposed to on-going or proposed research projects or programs.

Evaluation frameworks, or methods, can broadly be divided into those that are single criterion and those that are multiple criteria. Typically, single criterion approaches seek to maximize the value of a pre-defined criterion (e.g., technological advance), whereas multiple criteria approaches often involve optimization of multiple objectives as well as a consideration of opportunity costs - a term used in economics to refer to the highest valued alternative use of a resource.

Single criterion approaches. Single criterion approaches are based on the assumption that one criterion, usually amenable to objective measurement, can be applied in the evaluation of the impacts associated with a research project or program. In some cases, the measurement of the single criterion is sophisticated; and in others it is extremely simple, such as the presence or absence of an attribute (e.g., the technology does or does not fit the performance standard). Generally, there is less room for normative interpretation of the underlying information when conducting a single criterion evaluation. Two of the more common single criterion approaches for evaluating government-sponsored research outcomes are referred to as technology-based standards and a cost-effective standards.

Technology-based standards are often applied in a comparative way.[7] The question asked is; Has the best technological result been obtained from a research project or program? Under this evaluation standard, "best" is generally defined in an engineering sense of state-of-the-art. If one research project or program has resulted in more path breaking findings than another, it is evaluated more highly. The primary advantage in this case is that neither quantifiable information on benefits nor quantifiable information on costs is needed to conduct the evaluation. All that is required is consistent engineering judgment. The primary disadvantage is that best is a subjective concept, even to engineers who deal on a regular basis with the technological dimensions of research. A more objective approach to single-criterion technology-based evaluation is to evaluate according to a specific, pre-determined technical standard. For example, if the goal of the research project or program is to develop a technology that, say, doubles the thermal efficiency of a solar collector, either the technical goal has been met or it has not; the determination of success is by objective measures.

Cost-effective standards can also be used in evaluating one research project or program against another.[7] The question asked is; Have the technical objectives of the research been achieved in the most cost-effective way? Because there is no consideration of opportunity costs or other objectives, the answer is straightforward; the single most cost-effective outcome can be identified using an objective criterion. Two research projects or programs can achieve the same technical result, but under this standard the more cost-effective one is presumed to be the more highly valued. One advantage of this cost-effective standard evaluation approach is that there is no need to measure benefits (this eliminates an element of subjectivity and also saves on available resources). Benefits are defined in terms of attainment of the technical objective. Given technical completion, the evaluation criterion is the single dimension of cost effectiveness. The primary disadvantage is that consideration is not given to the usefulness or applicability of the research results. The evaluation decision is thus void of any dimension that characterizes the applicability of the research, and thus is void of any measurable element of social impact.

Multiple criteria approaches. Multiple criteria approaches are based on the assumption that there are trade-offs among the evaluation-related dimensions of a research project or program. Three such multiple criteria approaches are multiple programming analysis, economic impact analysis, and benefit-cost analysis (or cost-benefit analysis).

Multiple programming analysis ranks projects or programs on the basis of a preestablished set of criteria, such as objectives of the research, cost constraints on the research, and outcomes of the research.[8] Given either the evaluator's criteria or the criteria of the government agency funding the research, the mathematical programming model developed will be capable of weighting the *a priori* criteria and solving for that project that satisfies most completely the various criteria. A strength of this method is its objectivity; however, the objectivity is related to the presumed relevance of the *a priori* criteria. Stated alternatively, a weakness is that the objectivity of the model is only as useful as the information programmed into the model.

Economic impact analysis incorporates a number of methods for associating research results with economic variables using a variety of statistical methods. The objective of this evaluation method is to identify quantitatively a pre-determined set of economic variables that is affected by the research (e.g., employment in a particular sector of the economy). The main advantage of this approach is that it focuses on generally accepted macroeconomic variables. The primary disadvantage of this approach is that the cost of conducting the initial research is not considered in the analysis.

Benefit-cost analysis relates the measurable benefits resulting from a research project or program to the cost of conducting the research. The primary advantage of benefit-cost analysis is that it is an easily understood evaluation tool. One important disadvantage of the method is that it requires a significant amount of resources to implement, and all too often many of the identified benefit categories cannot be easily measured, even for completed research projects or programs.

Measurement issues related to benefit-cost analysis

Measurement of Benefits. The benefits that result from completed research projects or programs fall into two broad categories: tangible benefits and intangible benefits. By definition, tangible benefits are quantifiable, while intangible benefits are not. Within each category, benefits can be further disaggregated into sub-categories of direct and indirect.

Direct benefits are those that flow directly to primary organizations – the government-sponsoring organization or the initial private-sector user of the research results. Indirect benefits are those that flow to a secondary organization – the customers of the primary organizations.

There are two general approaches to the measurement of tangible direct benefits and tangible indirect benefits as related to benefit-cost analysis: the revealed preference approach and the expressed preference approach.

The fundamental concept underlying the revealed preference approach is that consumers value the benefits from an activity in terms of their willingness to pay for goods or services related to that activity. In other words, consumers reveal their preferences or values through their participation in market transactions. When the goods and services that have. resulted from the research being evaluated are exchanged in perfectly competitive markets, then market price is the appropriate measure of willingness to pay. However, it is often the case that the goods and services that result from the research being evaluated are not part· of an organized competitive market. In fact, one of the justifications for public sector involvement is that the market has failed to provide a sufficient quantity or quality of such goods and services. Thus, willingness to pay is often applauded in theory, but in practice it may not be implementable.

Another approach to measuring benefits relies on expressed preferences.[9] By the way that individuals express their preferences for benefits, evaluators can directly assess a value for the benefits. While there have been numerous so-called expressed preference assessment methods illustrated in various literatures, most methods are based on the assumption that respondents are knowledgeable about the economic aspects of the benefits being evaluated and will express their preferences consistently as well as truthfully. Two of the more commonly used preference approaches are contingent valuation methods and contingent ranking methods.[10]

A third assessment approach is called the hedonic assessment method. Here, benefits are measured in terms of certain quantifiable characteristics or attributes of the output of the research being evaluated, and then some value is placed upon those characteristics. Such attribute measurements have been based on bibliometric tools (publication counts, citation counts, co-word analysis) or patent analyses.[11-13] In any event, the final assessment of the value of these non-market goods and services must rely on peer evaluation or informed opinion.

Measurement of costs. On the cost side, there are tangible and intangible costs, and both have direct and indirect aspects. The measurement of costs is generally regarded as more straightforward than the measurement of benefits. The reason for this is that tangible costs can be divided into easily identifiable categories related to labor, capital (plant and equipment), and research; and tangible costs can generally be measured through traditional accounting procedures. As with benefits, intangible costs are not quantifiable.

There are two categories of tangible costs that must generally be quantified. One category includes what may be called "push costs," and the other includes what may be called "pull costs." Push costs, in the case of government-sponsored research, are all public sector costs associated with both conducting the research and transferring it from the laboratory to society. In situations where the research being evaluated is part of a larger undertaking, push costs must be disaggregated or separated from the unit's larger budget. In situations where there are multiple performers of the government-sponsored research or multiple research activities that collectively provide benefits, push costs must be aggregated or summed across activities. Pull costs are the costs that society (meaning the sum of all users of the technology) expends to identify, acquire, and implement the government-sponsored technology.

Benefit-cost frameworks

The two more commonly used frameworks for comparing benefits to costs are benefit-cost ratios and internal rate of return calculations.[14]

Benefit-Cost Ratios. Because benefits and costs occur at different points in time, with costs often occurring before any benefits are realized, benefits and costs should be compared in present value terms:

$$B/C = [\Sigma_{0\text{ to }n}\, B_t/(1 + r)^t]/[\Sigma_{0\text{ to }n}\, C_t/(1 + r)^t]$$

where B_t represents all tangible benefits and C_t represents all tangible costs, where the time frame over which costs are incurred and benefits are received is represented by t, and where the relevant rate of discount to equate future values to the present is represented by r. It is often the case that $C_t > B_t$ in the early years of research, and then $B_t > C_t$ in the latter years or after the research is completed (when C_t may be zero).

Internal rate of return calculations. The internal rate of return (IRR) is defined as that rate i that equates the net present value (NPV) of a project to zero:

$$NPV = [(B_0 - C_0)/(1 + i)^0]+\ldots+[(B_n - C_n)/(1 + i)^n]$$

where $(C_t - B_t)$ represents net benefits.

Both benefit-cost ratios and internal rates of return have been widely used when evaluating public sector research.[16] In fact, the two frameworks are related in the sense that if the discount rate on benefits is equal to the discount rate on costs within a benefit-cost ratio, then when NPV=0, B/C= 1.

The B/C ratio is often used when characterizing the returns to one particular government-sponsored research project or program, and the IRR calculation is often used when comparing the returns across government-sponsored research activities or organizations (or when deciding to undertake a particular project given some threshold acceptable rate of return). Benefit-cost ratios may be inappropriate for comparing across projects or programs because, for example, project A may yield less total net benefits than project B, but project A may have the larger benefit-cost ratio. As well, critics of the use of benefit-cost ratios are quick to point out that the meaningfulness of the results depends not only on the accuracy with which both benefits and costs are measured and the time frame horizon over which they are measured, but also on the selection of a discount rate that truly represents the opportunity cost for public monies.

As a hypothetical illustration, consider a government-sponsored research project or program that began in year t=0 and continued through year t=5. After year t=5, tangible push costs were $0. In years t=6 and t=7, firms in the private sector incurred tangible pull costs as they identified, acquired, and implemented the technology results from the government-sponsored research. In years t=8 through t=12, the private sector realized benefits.

Tangible direct benefits were realized by the acquiring/implementing firms and tangible indirect benefits were realized by the customers of these firms through, say, improved product quality. Assuming that the government-sponsoring organization can indeed quantify C_0 ... C_5 accurately (although in practice these costs may be understated because the government-sponsoring organization may be unable to impute a cost to its acquisition of the fundamental scientific information that motivated the research project or program, even if the origin of such fundamental research can be identified), and assuming that private-sector participants can quantify, net of pull costs, benefits B_8 ... B_{12} accurately (although these benefits may also be understated because there may be no verifiable way to identify the specific private-sector firms that utilize the technology results coming from the government-sponsored research, owing to its public-good nature, or among those that can be identified, to determine how long or how completely they have used the technology results; or these benefits may be overstated because few if any private-sector firms are likely to have accounting systems designed to quantify pull costs in years t=6 and t=7 at all, much less quantify them n-years after the fact), then, other fundamental concerns still remain.

It is these other fundamental concerns that underscore the importance of using benefit-cost analysis cautiously and understanding its inherent subjectivity. One such concern relates to the discount rate to use and at what point in time the analysis should be referenced. Unfortunately, there is simply no agreement among economists or practitioners about the appropriate value of the discount rate.[18] That fact aside, there is some agreement about the origin of the analysis. In most analyses, benefits and costs are discounted to year t=0, although some evaluators discount benefits to the year that they began and then inflate costs to that year. Unfortunately, this latter practice will produce a higher B/C because the discount rate is generally larger than the annual rate of cost increase. A second concern relates to the treatment of intangible benefits, possibly ignoring them on the grounds that they are not quantifiable. If one is interested in a unqualified B/C number, then these benefits must by definition be ignored. However, when the intangible benefits are intuitively large and long lived, such as environmental or health benefits, then a qualified B/C ratio may be preferred, where the qualifications include qualitative data about the intangible benefits.

Economic performance measures – Theory and practice

As stated above, an economic evaluation emphasizes the evaluation of the net benefits to society that are associated with the actual or potential output of the research project or program being evaluated. Generally, this means that economic evaluations emphasize the myriad entities that utilize the technical knowledge or technology that results from the completed government-sponsored research being evaluated.

While economic theory stresses the value of expressed preference data for measuring economic benefits – such as market information that the price of a product fell a given percent as a result of technology derived from a government-sponsored research project or program – relevant data are difficult to obtain in practice. The reason for this difficulty is generally associated with the fact that there is a long lag between research, research results, implementation of research results, resulting market effects from having the research results, and the measurement of such market effects. In fact, because most evaluations of government-sponsored research have been conducted by third parties as part of contracted research or academic research, the institutional knowledge that is so vital in constructing the linkages necessary to identity benefits is at best wanting, and more likely missing altogether.

When a government-sponsored research project or program has not been completed, the only method for measuring benefits is through expressed preference data of the anticipated or expected benefits and of the affected entities. This problem is confounded when the government-sponsored research that is being evaluated is basic research. The accepted National Science Foundation definition is: basic research represents original investigation, for the advancement of scientific knowledge, that does not have a specific commercial objective.[20] Accordingly, even anticipated or expected benefits are often nothing more than a guess. When completed basic research projects or programs are evaluated, the time lag is much greater than for applied research and thus benefits can only be estimated by tracing back from technologies to their basic research foundation. This too is wanting in the sense that no revealed preference data are available and more likely "recalled" preference anecdotes become the benchmark for measuring benefits.[21]

Given these apparent shortcomings, economists have tended to rely alternatively on a production function approach to evaluating the economic returns to research. The logic of the production function approach is that if government-sponsored research, basic or applied, is associated with economic benefits in the private sector then those private sector firms utilizing the resulting technology will eventually realize productivity improvements from the technology. Obviously, there are a number of implicit assumptions embedded in this logic. One assumption is that all, or at least the most important, benefits associated with the government-sponsored research occur in the private sector. Another assumption is that in the private sector all economic benefits associated with implementing technology emanating from government-sponsored research are manifested in measurable productivity improvements. Cost-reducing improvements may be so measured, but product improvements or even the development of a new product may not be. And finally, a third and very critical assumption is that when productivity improvements are measured, all economic benefits have been fully realized. This last assumption of timing is, of course, germane to any calculation of economic benefits.

Because the production function approach is fundamental to most economic evaluations of research, especially of basic research, the complete model is derived below to emphasize the assumptions that underlie it. What generally appears in the academic literature when economists implement this model begins with an abbreviated version of equation (7).

The production function approach to estimating the returns to research activity is based on a generalized multi-input and multi-output production function, conceptualized at the level of a firm as:[25]

$$H(q_{1t}, \ldots, q_{mt}; x_{1t}, \ldots, x_{nt}; t) = H(\tilde{q}; \tilde{x}; t) = 0 \tag{1}$$

where the symbol ˜ denotes the vector of m-numbered outputs (q's) and n-numbered inputs (x's), and t indexes time. If the function H is homothetic and weakly separable as defined by H*, then, by definition, equation (1) can be written as:

$$H(\tilde{q};\tilde{x};t) = H*(G*(\tilde{q});F*(\tilde{x});t) = 0 \tag{2}$$

and if the separability of the function is additive, then:

$$G(\tilde{q}) = f(\tilde{x};t) \tag{3}$$

And, if the multi-output vector is replaced by a single composite vector q, the multi-input vector replaced by a single composite vector x, and if time-related technology is neutral and disembodied, then, by definition, equation (3) becomes:

$$q = A(t)f(x) \tag{4}$$

From equation (4), the concept of total factor productivity (TFP), meaning the ratio of output to the combination of all inputs used in the production process, is:

$$TFP = A(t) = q/f(x) \tag{5}$$

and then technological change can be defined in terms of the percentage change in TFP over time, TFP′ (TFP′ =dTFP/dt), as:

$$TFP'/TFP = A(t)'/A(t) = A'/A \tag{6}$$

When economists estimate a rate of return to research, they generally begin with equation (4) and assume that the firm's production function has four inputs: capital (K), labor (L), own technology (OT) coming from the results of accumulated self-financed research, and government technology (GT) coming from the results of accumulated government-sponsored research. They also generally assume that the mathematical form of the production function is Cobb-Douglas:[27]

$$q = A(t)K^{\alpha}L^{(1-\alpha)}OT^{\beta}GT^{\gamma} \tag{7}$$

where α, β, and γ are assumed to be within the positive unit interval. If all other technology is neutral and grows at a disembodied rate of λ, then:

$$q = Ae^{\lambda t}K^{\alpha}L^{(1-\alpha)}OT^{\beta}GT^{\gamma} \tag{8}$$

Using logarithmic transformations and differentiating equation (8) with respect to t:

$$q'/q = \lambda + \alpha(K'/K) + (1-\alpha)(L'/L) + \beta(OT'/OT) + \gamma(GT'/GT) \tag{9}$$

And, recalling the definition of total factor productivity growth from equation (6) and rearranging terms:

$$\mathrm{TFP'/TFP} = (q'/q) - \alpha\,(K'/K) - (1-\alpha)(L'/L) = \lambda + \beta(\mathrm{OT'/OT}) + \gamma(\mathrm{GT'/GT}) \qquad (10)$$

Equation (10) is interpreted to mean that changes in a firm's output over time that are not explained by changes in the stock of capital (K) or labor (L) are linearly related to disembodied technological growth (λ), growth in the stock of own technology (OT), and growth in the stock of government technology (GT).

For empirical purposes, equation (10) becomes the fundamental equation for estimating the returns to government-sponsored research when a random error term, ϵ, is added:

$$\mathrm{TFP'/TFP} = \beta(\mathrm{OT'/OT}) + \gamma(\mathrm{GT'/GT}) + \epsilon \qquad (11)$$

It follows from equation (8) that $\beta = (\delta q/\delta OT)/(OT/q)$ and that $\gamma = (\delta q/\delta GT)/(GT/q)$ so that equation (11) can be rewritten as a linear regression model of the form:

$$\mathrm{TFP'/TFP} = \rho_0 + \rho_1(\mathrm{SFR/q}) + \rho_2(\mathrm{GSR/q}) + \epsilon \qquad (12)$$

where the intercept ρ_0 is the disembodied rate of growth parameter; $\rho_1(\rho_1 = \delta q/\delta OT)$ is the rate of return to self-financed research (SFR) under the assumption that self-financed research expenditures are the relevant flow into the stock of own technology (δOT=SFR); and $\rho_2(\rho_2 = \delta q/\delta OT)$ is the rate of return to government-sponsored research (GSR) under the assumption that government-sponsored research expenditures are the relevant flow into the stock of government technology (δGT=GSR). Therefore, economists have looked at empirical estimates of ρ_2 based on cross-sectional firm (or industry) data to determine the return to government-sponsored research and have concluded that the return is measurably positive but small compared to the return to self-financed research.[28-30]

This production function approach, while dominant among economists, has a number of significant limitations which render its usefulness to an evaluation of any one particular completed government-sponsored research project or program questionable. First, the implicit assumption that all technological improvements result in measurable productivity gains is questionable even if timing is not considered an issue. Second, the production function models are limited in their ability to capture spillover benefits that accrue to either second-level private sector users of the technology, to other public sector users, or even to universities. Third, absent the above problems, it is often the case that government-sponsored research, basic research in particular, has an indirect effect on private sector productivity. That is, government-sponsored basic research can leverage the efficiency of private sector basic or applied research.[31] And fourth, inferring causation from a statistical correlation may not be valid in the case of government-sponsored research, especially if private-sector productivity growth is itself a determinant of the level of public-sector resources allocated to government-sponsored research.

Concluding observations

While the political climate is such that government-sponsored research is being scrutinized with an eye toward budget cutting, it is incumbent upon public research managers to identify measurable outcomes from research and assess their economic importance. Toward this end, some economic performance measures are more appropriate than others.[32] Especially appropriate are performance measures based on benefit-cost analyses, where economic benefits have been measured in such a way as to capture the impacts occurring throughout both the public and private sectors. To accomplish this goal, evaluators will have to rely upon expressed preference data, especially in situations where the government-sponsored research is not yet completed or when its impacts are not fully realized. In so doing, evaluators must understand the limitations placed upon the results of their analysis owing to the fact that expressed preference data are inherently subjective. Albeit subjective, such economic-based performance measures are more complete in scope than traditional production function models and are more generalizable than case studies.

Acknowledgements

This paper benefitted greatly from comments and suggestions by *Ron Kostoff.*

References

1. *Crow*[2] provides an excellent and more complete discussion of the history of U.S. science and technology policy.
2. M. M. CROW, Science and technology policy in the United States: Trading in the 1950 model, *Science and Public Policy,* 21 (1994) 202-212.
3. L. B. LAVE, *The Strategy of Social Regulations: Decision Frameworks for Policy,* The Brookings Institution, Washington, D.C., 1981.
4. E. ORMALA, Nordic experiences of the evaluation of technical research and development, *Research Policy,* 18 (1989) 333-342.
5. E. ORMALA, Impact assessment: European experience of qualitative methods and practices, *Evaluation Review,* 18 (1994) 41-51.
6. V. K. SMITH, A conceptual overview of the foundation of benefit-cost analysis. In: J. D. BENTKOVER, V. T. COVELLO, J. MUMPOWER (Eds), *Benefit Assessment: The State of the Art,* Kluwer Academic Publishers, Dordrecht, Holland, 1986, pp. 13-34.
7. D. W. PEARCE, C. A. NASH, *The Social Appraisal of Projects: A Text in Cost-Benefit Analysis,* John Wiley, New York, 1981.
8. M. ZELNEY, *Multiple Criteria Decisionmaking,* John Wiley, New York, 1982.
9. L. A. Cox, Jr., Theory of regulatory benefits assessment: Econometric and expressed preference approaches. In: J. D. BENTKOVER, V. T. COVELLO, J. MUMPOWER (Eds), *Benefit Assessment: The State of the Art,* Kluwer Academic Publishers, Dordrecht, Holland, 1986, pp. 85-159.
10. R. G. CUMMINGS, L.A. Cox, Jr., A. M. Freemen III, General methods for benefits assessments. In: J. D. Bentkover, V. T. Covello, J. Mumpower (Eds), *Benefit Assessment: The State of the Art,* Kluwer Academic Publishers, Dordrecht, Holland, 1986, pp. 161-191.
11. R. N. Kostoff, Co-word analysis. In: B. BOZEMAN, J. MELKERS (Eds), *Evaluating R&D Impacts: Methods and Practice,* Kluwer Academic Publishers, Boston, 1993, pp. 63-78.
12. J. MELKERS, Bibliometrics as a tool for analysis of R&D impacts. In: B. BOZEMAN, *I.* MELKERS (Eds), *Evaluating R&D Impacts: Methods and Practice,* Kluwer Academic Publishers, Boston, 1993, pp. 43-61.

13. M. PAPADAKIS, Patents and the evaluation of R&D. In: B. BOZEMAN, *I.* MELKERS (Eds), *Evaluating R&D Impacts: Methods and Practice,* Kluwer Academic Publishers, Boston, 1993, pp. 99-121.
14. *Campen*[15] provides an excellent synopsis of the history of benefit-cost analysis, beginning with its first application to implement the requirements of the River and Harbor Act of 1902.
15. J. T. CAMPEN, *Benefit, Costs, and Beyond: The Political Economy of Benefit-Cost Analysis,* Ballinger Publishing, Cambridge, Mass., 1986.
16. See *Link*[17] for examples of the application of benefit-cost analysis to research programs sponsored by the National Institute of Standards and Technology within the U.S. Department of Commerce.
17. A. N. LINK, *Evaluating Public Sector Research and Development,* Praeger Publishers, Westport, Conn., 1996.
18. *Mikesell*[19] suggested using the long-term rate on government bonds plus 2 percentage points, and *Link's*[17] case studies follow this guideline for illustrative purposes. However, it remains subjective, even if this guideline is correct, as to what long-term rate to use-the rate in effect in year t=0, the rate in effect when benefits first appear and are first realized in year t=8, the rate in effect when all benefits are thought to be realized at the end of year t= 12, or the rate in effect when the evaluation is conducted. Thus, the origin in time for the analysis may be very important.
19. R. F. MIKESELL, *The Rate of Discount for Evaluating Public Projects,* American Enterprise Institute, Washington, D.C., 1977.
20. A. N. LINK, On the classification of R&D, *Research Policy,* 25 (1996) 397-401.
21. Examples include Refs 22-24.
22. A. N. LINK, Payoffs from basic research: Dead ends, detours, and redirections. In: J. SOMMERS (Ed), *An Agenda for Science Policy Research,* National Science Foundation, Washington, D.C., 1987, pp. 1-14.
23. A. N. LINK, D. J. TEECE, W. F. FINAN, Estimating the benefits from collaboration: The case of SEMATECH, *Review of Industrial Organization,* forthcoming, 1996.
24. M. PAPADAKIS, A. N. LINK, Measuring the *unmeasurable:* Cost-benefit analysis for new business startups and basic research transfers, mimeographed, 1996.
25. This derivation is taken from *Link.*[26] The references related to the fundamental theory underlying the derivation are noted therein.
26. A. N. LINK, *Technological Change and Productivity Growth,* Harwood Academic Publishers, London, 1987.
27. *Link*[26] discusses applications of other mathematical forms for production functions.
28. D. P. LEYDEN, A. N. LINK, *Government's Role in Innovation,* Kluwer Academic Publishers, Boston, 1992.
29. A. N. LINK, Basic research and productivity increase in manufacturing: Additional evidence, *American Economic Review,* 71 (1981) 1111-1112.
30. E. MANSFIELD, Basic research and productivity increase in manufacturing, *American Economic Review,* 70 (1980) 863-873.
31. G. TASSEY, *Technological Infrastructure and Competitive Position,* Kluwer Academic Publishers, Boston, 1992.
32. Not discussed in this paper are evaluation approaches based on the calculation of consumer and producer surplus. While theoretically sound and clearly important, as carefully discussed by *Averch,*[33] the application of this evaluation approach generally retreats to a case study

analysis that is based on expressed preference data. In this sense, the topic is subsumed in the discussions above.

33. H. A. AVERCH, Economic approaches to the evaluation of research, *Evaluation Review,* 18 (1994) 77-88.