MEASURING THE PERFORMANCE OF GOVERNMENT TECHNOLOGY PROGRAMS: LESSONS FROM MANUFACTURING EXTENSION

By

Ronald S. Jarmin*
U.S. Bureau of the Census
Center for Economic Studies
4700 Silver Hill Road, Stop 6300, WPII,Room 211
Washington, DC 20233-6300
Voice:(301) 457-1858   Fax:(301) 457-1235
email: rjarmin@census.gov

and

J. Bradford Jensen*
Heinz School
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
Voice:(412) 268-8368   Fax:(412) 268-7036
email: jbjensen@andrew.cmu.edu.

CES 97-18   December 1997

Washington Plaza II, Room 211, Bureau of the Census, Washington, DC 20233-6101, (301-457-1882) or INTERNET address snguyen@info.census.gov.

Abstract

Managers of government technology programs are under increasing pressure to demonstrate the effectiveness of their programs. In this paper we examine the issues involved in credibly evaluating such programs in the context of recent efforts to evaluate manufacturing extension programs in the U.S. We provide a stylized model of the dynamic competitive environment in which the plants and firms targeted by these programs operate and discuss its implications for evaluation. We compare and contrast the various methodologies and data sets used to evaluate manufacturing extension. We conclude that the best currently available method for measuring the overall effectiveness of programs such as manufacturing extension is to combine program administrative data with existing panel data sets.

Keywords: Evaluation, Manufacturing Extension, Technology Policy

***An earlier version of this paper was presented at the OECD "Conference on Policy Evaluation Practices in Innovation and Technology." We thank Bob McGuckin for helpful comments. Any findings, conclusions or opinions expressed here are those of the**

I. INTRODUCTION

In 1994, federal and state governments in the United States spent just under $3 billion[1] on programs designed to encourage businesses to discover, develop and adopt improved technologies and business practices[2]. Taxpayers and policymakers want to know if these programs have the desired impact and that this money is well spent. This is especially true after congress passed the Government Performance and Results Act (GPRA) in 1993. The GPRA will require all federal executive branch agencies to provide annual performance reviews by the year 2000.

Program evaluators and managers face many challenges as they try to effectively fulfill the requirements of this mandate in a tight fiscal environment. There are many data and methodological issues that must be addressed in order to credibly measure the performance of these programs. GAO (1997) provides an overview of some of the hurdles faced by a handful of agencies that participated in a pilot implementation of the GPRA. These include i) a lack of well defined and consistent program objectives ii) the difficulty in controlling for non-program influences on performance measures, iii) a lack of appropriate performance information, iv) the difficulty in altering the management culture at agencies to

[1] This estimate combines numbers from Berglund and Coburn (1995) and from a survey of state programs by the State Science & Technology Institute.

[2] Examples of such programs include, in addition to Manufacturing Extension, the Advanced Technology Project (ATP), SEMATECH and the Small Business Innovation Research (SBIR) program.

stress measurable performance improvement and v) the difficulty of integrating performance information into the budget process.

Given these challenges, we believe that it is instructive to review the recent efforts to evaluate a particular technology program: Manufacturing Extension. While important differences exist between it and other technology programs, we believe that many of the lessons learned evaluating manufacturing extension can be applied elsewhere.

A. Contributions of This Paper

Our main goal in this paper is to examine some of the more important recent studies that try to evaluate the effectiveness of manufacturing extension programs[3]. We compare and contrast the different methodologies and data sets employed and discuss the extent to which these studies satisfy the information needs of the wide spectrum of program stakeholders. We also discuss the costs and difficulties associated with the various methodologies.

Although researchers have not yet arrived at a final assessment of the effectiveness of manufacturing extension programs, they have learned much about the technical and practical issues involved in evaluating these programs. This knowledge will help us to improve ongoing analyses of manufacturing extension and

[3] Previous reviews of this literature include Shapira, Youtie and Roessner, 1996 and Feller, Glasmier and Mark, 1996. These studies provide excellent overviews of manufacturing extension programs and some of the issues that arise in trying to evaluate them. However, they predate many of the more systematic and rigorous studies that have come out in the last couple of years and that we discuss here.

will aid researchers trying to understand the issues involved in evaluating other technology programs.

As we began to take stock of what has been learned from the many studies evaluating manufacturing extension programs and trying to compare the different methods researchers have used, we noted that we lacked a coherent framework within which we could judge the merits of the various studies. Poorly defined or inconsistent performance metrics plague many studies. Often the link between program services and performance measures is not clear.

We believe that these difficulties arise from the fact that there is no general model of how manufacturing extension services impact client plants and firms. Most evaluation studies and, indeed, most discussions about manufacturing extension, in general, ignore the competitive environment within which client plants and firms operate. It is in this environment that extension services are intended to improve client performance and where the data we use for program evaluation are generated. Thus, it is important to have an understanding of this environment and how programs like manufacturing extension fit into it in order to optimally design, provide and evaluate program services.

Thus, we provide a stylized model of the environment in which client plants and firms operate. We then show how manufacturing extension fits into this stylized model. The model provides an excellent framework for our discussion of recent evaluation studies.

B. Brief Description of Manufacturing Extension

Before characterizing this framework, let us briefly describe manufacturing extension.[4] Manufacturing extension is the term describing the collection of organizations that provide industrial modernization services to small and medium sized manufacturers (SMEs). At the federal level, manufacturing extension is administered by the National Institute of Standards and Technology's (NIST) Manufacturing Extension Partnership (MEP) as part of their effort to improve the competitiveness of U.S. manufacturing industries. The MEP supports several manufacturing technology centers (MTCs) around the country that provide technical and business assistance to SMEs, much as agricultural extension agents do for farmers. This assistance often consists of providing "off the shelf" solutions to technical problems. However, these centers can also channel more recent innovations generated in government and university laboratories to smaller U.S. manufacturing concerns that may not have access to such information. The idea is that extension services will help these firms become more productive and compete more effectively in the international marketplace.

[4] More detailed discussions of the design and justification for manufacturing extension programs can be found in National Research Council (1993) and Feller (1997).

II. STYLIZED MODEL OF COMPETITION

The purpose if this section is to help us understand how extension services might impact plant performance and to provide a logical framework that can be used to interpret data for evaluation purposes. We begin by providing a highly stylized model of the competitive environment U.S. firms operate in, and in which these policies and programs are intended to work. Next, we characterize the dynamics of firm decision making within this competitive environment. Finally, we discuss how manufacturing extension fits into this dynamic framework.

A. Basic Framework for the Model Stylized Model

Our starting point is to note that firms and their constituent plants are very heterogeneous, even within industries. Important observable dimensions of this heterogeneity include but are not limited to size, age, capital intensity, productivity and wages. Jensen and McGuckin (1996) review an extensive literature that documents the pervasiveness and persistence of plant and firm heterogeneity along these and other dimensions. It is equally important to note that this heterogeneity extends to unobserved dimensions, such as managerial ability.

The lesson to take from this growing literature is that, at any given time, plants and firms have very different characteristics, are pursuing different strategies and are subject to different idiosyncratic shocks[5]. This heterogeneity causes firms to respond differently to common shocks[6] and to government policies and programs. In the case of programs, such as manufacturing extension, plants and firms will respond differently to the availability of program services. Some will choose to take advantage of the program's services and others will not. Even within the group of plants and firms that become clients, there will be a variety of strategies pursued (including making no changes) as a result of their participation in the program. The different strategies pursued by heterogeneous firm cause their observable (and unobservable) characteristics to evolve differently over time. The implication for program evaluation is that studies need to be performed at the plant or firm level, since aggregate data will mask much of the variation that is of interest to the evaluator.

It is within this dynamic and heterogeneous environment that manufacturing extension services are provided and the data available for evaluating the impacts of these services are generated. Thus, our stylized model should incorporate as many features of this environment as possible.

---

[5] Examples of the characteristics we refer to are size, product mix, plant age, workforce composition, capital intensity, technology usage, productivity and so on. Examples of the strategies we refer to include price and output levels, capital investments, adopting new technologies and so on. Idiosyncratic shocks are unforseen events such as machinery break downs, strikes and the like that affect an individual plant or firm.

[6] Common shocks include unforseen changes in consumer tastes, the business cycle, wars and so on.

Figure 1 provides a flow chart of the basic components of our stylized model. It shows what takes place during a single decision making period[7] for an arbitrary firm. At the beginning of the period (indexed by t in figure 1), the firm (indexed by i) has a set of information available to aid it in making decisions regarding price, output, wages, investment, participation in manufacturing extension and so on. Once decisions are made, the firm takes the specified actions (i.e., sets prices, produces output, sets wages, makes investments, takes part in a manufacturing extension project or whatever). Firm i's choices, the choices of other firms, the realization of the state of the world (exogenous events or shocks that impact outcomes for firm i and its competitors) and the current state of technological knowledge then determine the set of outcomes (e.g., profits, productivity, growth, etc) for firm i in period t. Firm i then updates what it knows about other firms, the state of the world, the stock of technological knowledge and the mechanisms through which they, and the actions firm i itself has taken, combine to yield outcomes. The process then repeats itself.

Before describing how manufacturing extension fits into this framework, let us discuss some of the basic components in more detail. The component we will focus on is the information

[7] Firms may differ in the length of their decision making periods and an individual firm may use different planning horizons for different types decisions. Since most of the data used in the studies we review here are annual, we will use a year as the decision making period.

available to the firm. At the beginning of each decision making period, firm i has an information set, $\mathbf{\mathit{I}}_{it} = \{Y_{it},\ Y^{i}_{jt}, A_{it-1},\ A^{i}_{jt-1}, S^{i}_{t},\ K^{i}_{it}\}$, containing six elements that it uses to plan its operations.

The first element is a vector of characteristics, Y_{it}, for plant i at time t. These characteristics can include things such as profits, sales, number of employees, age of the firm, capital intensity, physical location and so on. Recall from the discussion above that, even within the same industry, these characteristics can differ greatly across plants. We assume that plant i observes its own characteristics without error.

The second element includes the characteristics, Y^{i}_{jt}, of i’s competitors that are observable by i. A key thing to note here is that firms do not observe all the relevant characteristics of their rivals and what they do observe, they observe with error. Thus, Y^{i}_{jt} is just an approximation of the true set of characteristics, Y_{jt}.

Also, note that a subset (Y^{g}_{it} d Y_{it}) of any firm’s characteristics is observed by organizations, such as trade associations and government statistical agencies. Generally, econometricians and policy analysts exploit these sources of data to analyze firm behavior and performance and how government programs and policies affect them.

The third element of firm i’s information includes the actions taken (or decisions made) by the plant in the previous decision making period. These can include things such as price and output levels, decisions about variable inputs (e.g., labor and materials), entry and exit decisions, capital investments and so on. The fourth element, A^i_{jt-1}, includes the observable actions taken by plant i's rivals in the previous period. Again we assume that plants know perfectly what actions they have taken in the past, but that they observe the past actions of other plants and firms with error.

The fifth element of firm i's information set describes what the firm knows about the state of the world, S^i_t. By state of the world, we mean things that influence outcomes in firm i's industry but that have their origin outside the industry (i.e., S_t is exogenous to all firms in the industry). This includes things such as consumer preferences, scientific discoveries, government regulations and so on. Firms do not observe past states of the world perfectly and their observations about it (or their interpretations of it) may differ greatly.

The final element, K^i_{it}, is the portion of the stock of technological knowledge, K_t, that firm i can observe. The stock of knowledge represents the information the economy possesses about how to transform inputs (e.g., labor, capital and materials) into outputs. Thus, K_t contains information on the quantity and quality of both fixed and variable inputs required to produce a given quantity of output. Typically, a plant will have a large menu of different input combinations that will yield a given quantity of output. Given market prices for both inputs and outputs, some combinations will be more economically efficient than others.

The stock of knowledge can be increased through both formal knowledge generating activities, such as research and development, and informal channels, such as learning by doing[8]. Clearly, most of the information contained in K_t will not be relevant to a given firm. However, no individual firm possesses all the knowledge that pertains to its operations. Also, note that unlike S_t, K_t can be influenced by past actions of firms in the industry (e.g., through R&D expenditures and learning by doing).

B. The Dynamics of Firm Decision Making and Competition

The objective of each firm i is to choose actions, A_{it}, given information, I_{it}, to maximize its discounted profit stream,

$$\begin{aligned} V_{it} &= \pi_{it}(A_{it}, A_{jt}, S_t, Y_{it}; I_{it}) + \delta_i V_{it+1} \\ s.t. \quad Y_{it} &= f(Y_{it-1}, A_{it-1}, A_{jt-1}, S_{t-1}, K_{it}) \\ I_{it} &= g(I_{it-1}, Y^i_{jt}, A^i_{jt-1}, Y_{it}, A_{jt-1}, S^i_{it}, K^i_{it}) \\ \text{and} \quad K_t &= h(K_{t-1}, A_{it-1}, A_{jt-1}, S_{t-1}) \end{aligned} \tag{1}$$

where V_{it} is the value of the profit stream in period t. The functions *f* and *g,* respectively, describe how the firm's characteristics and information set evolve over time. The function *h* describes how the stock of technological knowledge evolves over

[8] Jaffe (1989) discusses of the knowledge production function. Jarmin (1996) examines the relative contributions of formal and informal knowledge towards productivity growth.

time. The firm’s single period profit function is given by B_{it} and it depends on the current actions of it and its rivals, the realized state of the world and its beginning of period characteristics. Discounted future profits are captured in $*_i V_{it+1}$ where $*_i$ is a firm specific discount factor.

To get a feel for how competition proceeds over time, assume that there is a technological innovation in period t-1. That is, all plants are hit by the same common technology shock to K_t. Examples of such shocks might include the introduction of numerically controlled machine tools, the discovery of new materials or the development of the Internet. Importantly, there are three sources of heterogeneity that might cause firms to optimally respond to shocks, such as these, differently.

First, firms may differ in their observed and unobserved characteristics, Y_{it}. Namely, firms of different sizes, ages and so on, may choose different actions upon observing the shock. Second, firms may differ in their ability to observe the shock. Some firms may not observe the shock at all, while others observe it with varying levels of error (i.e., the K^i_{it} differ for different firms). Finally, firms will also differ in the cognitive ability of their decision makers to process the information about the shock and formulate the optimal response[9].

---

[9] These different sources of heterogeneity have different implications for econometricians and policy analysts. Namely, much of the first and all of the last two sources are typically unobserved. Thus, analysts can only explicitly control for a small portion of the differences that can lead firms to respond to

Given this heterogeneity, firms will choose different courses of action in response to common shocks. Further, firms are also subject to firm specific shocks. Thus, at any given time, individual firms will be pursuing heterogeneous strategies, even within industries, in their attempts to maximize profits.

Plants and firms that consistently choose actions, A_{it}, that lead to profitable outcomes will tend to grow and prosper. That is, their characteristics, Y_{it+1}, will evolve in positive ways. The econometrician or policy analyst will observe this as the continued operation of the firm and perhaps as movements of the firm up the industry sales, employment and productivity distributions. On the other hand, plants and firms that consistently choose actions that lead to unprofitable outcomes will tend to contract and fail. That is, their characteristics will evolve in a negative fashion. The econometrician and policy analyst might observe this as movements of the firm down the industry sales, employment and productivity distributions, and possibly by the firm ceasing operations.

C. The Role of Manufacturing Extension in this Framework

Having characterized the stylized model, we now would like to describe how we see manufacturing extension fitting into this framework. The premise behind manufacturing extension is that small and medium sized manufacturers have systematically less access to technological information than do their larger

---

shocks differently.

counterparts[10]. In the example above, SMEs either do not observe the technology shock, do so with more error, or may be less able to process information about the shock and formulate the most profitable response than large firms.

Proponents of manufacturing extension claim that this information disadvantage is largely responsible for the large and growing performance gap between large and small manufacturers. Through education and outreach, manufacturing extension attempts to increase the content of the information sets of client firms[11]. With this additional information, extension clients can then choose actions that lead to better outcomes.

We assume that policy makers have some subset, Y^{p}_{it} d Y^{g}_{it} (for all client plants i), of observable characteristics that they hope extension services will improve. For illustrative purposes, assume that this policy variable is value added per worker (i.e., $Y^{p}_{it} = (VA_{it}/L_{it})$). Now say that plant i participates in manufacturing extension in period t. The extension center provides i with additional information on the stock of technological knowledge. The

---

[10] The National Research Council (1993) cites five barriers to improved performance at SMEs: i) disproportionate impact of regulation, ii) lack of awareness (of better technologies), iii) isolation (from peers), iv) (don't know) where to seek advice, and v) scarcity of capital. The second, third and fourth barriers are directly related to the information gap between large and small manufacturers, and the first one is indirectly related.

[11] There are a host of private sector sources for such information as well. These include, but are not limited to suppliers, customers, consultants and professional organizations. However, proponents of manufacturing extension argue that the private sector provides an inefficiently small quantity of industrial modernization services to SMEs (see National Research Council, 1993 and McGuckin and Redman, 1995). Thus, they argue that there is a role for state and federal support for MTCs.

firm now chooses a different set of actions than it would have in the absence of extension services. These different actions then affect how the firm’s characteristics evolve over time. The policy maker hopes that value added per worker is higher at client firms in period t+1 than it would have been if the firms had not participated in extension.

III. EVALUATION METHODOLOGIES

The question we are interested in is; how can policy makers know with reasonable confidence that improvements in the performance of client plants or firms are due to participation in the program? In this section, we contrast the methodologies used in various studies that seek to evaluate manufacturing extension[12].

Like other government programs, manufacturing extension possesses a diverse set of program stakeholders including client plants and firms, the manufacturing technology centers, NIST, state and local governments and Congress. This diversity poses several challenges for evaluation since different program stakeholders have different concepts of and needs for evaluation. To meet this diversity of needs, evaluators have produced a diverse set of studies employing several methodologies, since no one evaluation methodology is suitable to meeting the needs of all program stakeholders.

---

[12] See McGuckin and Redman (1995) for a discussion on the characteristics of a “good” evaluation of manufacturing extension.

Although, our discussion below is primarily from the point of view of trying measure the overall performance of manufacturing extension programs, it is important to recognize that various stakeholder groups have different needs concerning program evaluation. For example, the management and staffs of the extension centers want to know if their services encourage client firms to take actions that will lead to better performance. Thus, they would be very interested in monitoring short term, intermediate outcomes, such as whether a client bought a new machine as a result of extension services. Also, to help them make decisions regarding the operations of the program, center managers might be interested in things like evaluating the relative effectiveness of particular program services or monitoring the performance of individual field agents. We refer to this type of evaluation as program monitoring.

In contrast, federal and state policy makers want to know whether extension services lead to improvements in some policy variable, such as value added per worker or exports. Thus, the policy maker is less concerned with what actions clients took that led to improvements, than she is with whether or not the actions ultimately led to the desired policy outcome. We refer to this type of evaluation as measuring program performance.

Another important thing to note, in regards to evaluating manufacturing extension, is that there may be a mismatch between the objectives of the client SMEs and those of policymakers (not to mention inconsistencies between different policy goals). Firms make decisions (e.g., whether to adopt modern technologies and business practices) they believe will lead to more profitable outcomes. These decisions can often be at odds with policy goals, as in the case where firms adopt new technologies that allow them to do the same amount of work with fewer employees.

In this section we review five methods by which technology programs can be evaluated. The methods differ in both the data and empirical methodologies required to carry them out. They are i) experimental methods, ii) studies using administrative data, iii) case studies, iv) client surveys and v) econometric analyses of nonexperimental data. The last four have all been exploited in studies attempting to assess the effectiveness of manufacturing programs in the U.S. Table 1 summarizes the major evaluation methodologies and lists some of their characteristics.

A. Experimental Methods

Recall that the goal of any effort to evaluate technology programs is to determine whether observed changes in the behavior and performance of client firms can be attributed to their association with the program. The most reliable method available is to randomly assign plants and firms to treatment (those that receive services) and control groups. One then analyzes the experimental data to see if the behavior and/or performance of the treatment plants differs from that of the control group. Again assume the evaluator is interested in the policy variable value added per worker ($Y^p_{it} = VA_{it}/L_{it}$). To obtain a measure of the program's impact, she could simply compute the mean value added per worker for both the treatment and control groups and compare them. Because of random assignment, she can be confident that any observed difference in value added per worker is attributable to program participation and is not due to some unobserved characteristics of treatment and control plants.

However, this type of evaluation is impractical for a number of reasons. We are, therefore, obliged to use the less scientifically rigorous methods described below. The choice of the appropriate evaluation methodology depends on what one hopes to learn from the evaluation and how one intends to use the results.

B. Administrative Data

Many program evaluation systems use "administrative data" as a primary input into analyses. "Administrative data" is data on activity levels and the types of services provided that are typically collected through the course of providing services and administering contracts. These data generally are recorded at the project level (or the client level) and describe the amount of time spent, the mix of services provided, ancillary investments made, and the cost of any fee-for-service activities. For evaluation, these data have the advantage that they can be collected with little additional cost.

Nexus Associates (1996) use administrative data from the New York Manufacturing Extension Partnership to provide a rich description of the services provided and the characteristics of client firms. However, the utility of administrative data for evaluation is limited. Administrative data tend to be more an artifact of the record-keeping systems at service providers, and less the result of careful planning by an evaluation team as to the types of data needed in program evaluation. Further, these data are too often used in evaluations that simply report activity levels (number of clients served, hours of consulting services provided, revenue generated, investments made, etc.) with no attempt to ascertain whether the activities had any impact.

Administrative data bases also vary considerably across individual manufacturing extension centers which limits cross center comparisons. Perhaps more important, administrative data sets do not contain control groups. Nor do they typically contain data on policy outcomes. Thus, they are not sufficient by themselves to provide the basis for a rigorous evaluation of the impact of manufacturing extension or other technology programs.

Nevertheless, recording activity levels, the types of services provided and other administrative data is crucial to program evaluation. As will be shown below, combining administrative data with longitudinal establishment data sets can produce a rich and comprehensive resource for evaluation. Because of the usefulness of administrative data, proper design of administrative data collection systems to ensure that we capture the appropriate data, while minimizing reporting burden, is critical to subsequent evaluation efforts and essential for MTC and client cooperation.

C. Case Studies

Case studies are detailed analyses of the interaction between MTCs and individual clients. As such, they provide a wealth of information concerning the characteristics, Y_{it}, of the client firms that are studied, the actions, A_{it}, the clients take as a result of extension services and the subsequent changes in client firm characteristics, Y_{it+1}, that result from these actions. This type of detailed information is especially important to MTC administrators to help them monitor what types of services and service delivery methods are most effective.

Examples of case studies can be found in Oldsman (1996) and in the collection of "Success Stories" maintained by NIST-MEP[13]. These provide descriptions of the problems faced by client firms, the solutions suggested by the individual MTCs and the final impact of the project. Problems faced by clients include excessive scrap and rework rates, poor quality control, outdated machines, safety issues and so on. Solutions included redesigning shop floors,

[13] These are accessible on the NIST-MEP world wide web page at http://vehicles.nist.gov/internet/success.nsf/$about.

implementation of statistical process control and on the job training programs, among other things. Reported impacts included reduced scrap rates and production lead times, increased sales, productivity and employment, and enhanced safety.

However, case studies are costly in terms of time and resources to the evaluator, the MTC and the client firm. For this reason, only a tiny fraction of all extension projects can be evaluated through case studies. Further, because of their expense and intrusive nature, case studies are not typically performed for nonclient firms. Finally, case studies are often criticized for relying too heavily on "success stories." That is, case studies are rarely performed and even more rarely reported for projects with minimal or negative impact. For these reasons, the case study is of limited use to the researcher or policy maker trying to measure overall program performance.

D. Client Surveys

Another method used to collect information and evaluate programs is the client follow-up survey. Client follow-up surveys typically survey all or some subset of the client population[14]. Follow-up surveys have the advantage that the questionnaires can be customized to collect detailed information on both intermediate outcomes (efforts at job training, investment in technology, etc.) and final outcomes (increased sales, increased employment, etc.).

---

[14] Note that client follow-up surveys rely critically on administrative data to create sampling frames.

Client follow-up surveys can be performed by service providers or collected by independent third-parties[15].

Client follow-up surveys have the advantage that they can elicit specific information about the impact of services on both intermediate and policy outcomes. Follow up surveys can be invaluable to center managers by helping them identify what mix of services clients respond to and to measure field agent performance.

However, client follow-up surveys are constrained by the willingness of the clients to report data, the reporting burden associated with the follow-up survey, and the burden associated with other service provider data collection efforts. Another disadvantage of client follow-up surveys is that only clients are surveyed. As a result, it is difficult to identify treatment effects because no control group is included in the analysis.

A possible remedy to the problem of no control group is to include a random sample of nonclients in the survey. However, this increases the cost of the survey. Another is to structure the questionnaire to provide time-series information on client behavior. This approach might be constrained by the willingness of clients to report data, as it would require clients to respond more than once. A third alternative is to construct questions that pose hypothetical scenarios to identify the change due to the services. This type of question is particularly subject to recall bias, since

[15] An obvious advantage of having a third-party conduct the survey is objectivity.

it requires respondents to remember details of their operations prior to receiving services. Thus, even when carefully constructed, these questions can produce responses of dubious value.

Follow up surveys have been commissioned and performed by several state programs. For example, clients of the New York Manufacturing Extension Partnership were surveyed by Nexus Associates (see Oldsman, 1996 and Nexus Associates, 1996). The results of the survey suggest that interactions with MTCs in New York resulted in many clients taking efforts to improve their performance (e.g., reconfiguring plant layout, new software, implementing TQM, etc.). Further, many clients reported that they experienced productivity increases, reduced scrap rates and lead times, and increased sales to name a few impacts. The majority of respondents reported that their interaction with MTCs increased their awareness of technologies and improved their ability to compete.

NIST-MEP currently sponsors a large scale telephone follow up survey conducted by the Census Bureau. This is the first survey to employ a common survey questionnaire for clients of many MTCs from different states. The analysis of the data has just begun and, although no firm conclusions can be made at this early stage, the results suggest that clients report that extension services have led to modest employment impacts (either through job retention or creation).

In the near future, this survey should provide a wealth of information useful for conducting cross center comparisons. This should aid both NIST-MEP and the individual MTCs identify what services are effective and which are not.

E. Econometric Analyses using Non-Experimental Data

The final methodology used to evaluate manufacturing extension comprises studies that specify and estimate econometric models with plant or firm level data from nonexperimental data sets. Jarmin (1995) discusses the modeling and data issues associated with this type of study. To perform these studies, researchers want a plant or firm level data set that contains measures of as many characteristics, Y_{it}, as possible. By appealing to an economic theory that suggests how these variables are related within the framework given in section II, the researcher then specifies an econometric model and estimates the impact of program participation on the relevant characteristic or policy variable. For example, the researcher may estimate a regression model like the following

$$Y^{p}_{it} = \beta^{o}X_{it}^{o} + \beta^{u}X_{it}^{u} + \gamma Z_{it} + \varepsilon_{it} \tag{2}$$

where Y^{p}_{it} is the policy variable (e.g., growth in value added per worker), X^{o}_{it} is a vector of observable plant characteristics (i.e., in the notation from above, X^{o}_{it} f Y^{g}_{it}, where Y^{g}_{it} is the set of characteristics for plant i that the evaluator can observe), X^{u}_{it} is a vector of unobservable plant characteristics (i.e., X^{u}_{it} é Y^{g}_{it}), Z_{it} is a program participation variable, $\$^{o}$ and $\$^{u}$ are vectors of parameters, (is the parameter that measures program impact and g_{it} is an error term. In many cases, Z_{it} is just an indicator variable for whether plant i was a client in period t. In this case, (measures the mean difference in Y^{p} between clients and nonclients controlling for the characteristics measured in X. If the researcher controlled for all the differences between clients and nonclients, other than client status, then (will be an unbiased estimate of program impact.

The principal constraint to performing such analyses is the availability of appropriate data. Jarmin (1995) discusses what properties an evaluation data set should have and provides a “wish list” of variables for evaluating manufacturing extension.

There are two ways to construct such a data set. First, one could randomly select a large number[16] of plants or firms and track them over time. One would collect data on several measures including information about program participation from both clients and nonclients. Surveys such as this are very expensive, typically exceeding the evaluation budgets of most programs.

Second, one can use pre-existing data sets. While these may not be designed and maintained specifically for program evaluation,

---

[16] This would ensure that a sufficient number of clients and a larger number of non-clients were in the final data set.

they are much less costly than special surveys. In the case of manufacturing extension, two such data sets have been utilized: 1) the Performance Benchmarking Database maintained by the Industrial Technology Institute (ITI), and 2) the Census Bureau's Longitudinal Research Database (LRD).

ITI collects data on a number of measures of interest for evaluation through its Performance Benchmarking Service (PBS). For a fee, plants and firms provide ITI with data and in return ITI provides them with a benchmarking report that compares their performance to similar plants and firms. NIST/MEP subsidizes the participation of a number of clients in the PBS. The data set, therefore, has observations on both extension clients and a control group of nonclients. A key advantage of the PBS data set is that participating plants and firms
provide data on an impressive variety of measures. This makes the data set extremely valuable for tracking intermediate, as well as policy outcomes.

To date, two studies have emerged that exploit these data. First, Luria and Wiarda (1996) examine the impact of services provided by the Michigan Manufacturing Technology Center during the period from 1991 to 1993. The results suggest that clients enjoyed greater sales and employment growth and reduced scrap rates faster than nonclients[17]. However, they found no link between program participation and productivity growth.

Nexus Associates (1996) estimate the impact of the New York Manufacturing Extension Partnership on client performance between 1992 and 1994. They regress the change in value added, within a modified production function framework, on several measures of services and find that extension services have a positive impact. However, it is not clear that the nonclient PBS data constitute an appropriate control group for this study. The PBS data are mostly for plants and firms in Michigan and other Midwestern states and, therefore, are likely subject to different shocks than plants and firms in New York.

These two studies are restricted to a small number of client and nonclient observations. Further, as in many evaluation studies, the estimates of program impact may be biased due to self selection[18]. Self selection arises when plants are not randomly assigned to client (treatment) and nonclient (control) groups. If there is some unobserved and uncontrolled for characteristic that determines whether plants become clients that is also related to performance, then standard estimation procedures produce biased

---

[17] Due to the small sample (68 clients and 354 non-clients), many of the results had marginal statistical significance.

[18] The econometric issues that arise as a result of self selection are well known (see Maddala, 1983). Stromsdorfer (1987) and Moffitt (1991) provide reviews of the literature on evaluating job training programs and discuss the issue of self selection.

estimates of the impact of the program on performance[19]. That is, Z_{it} and X^u_{it} in equation (2) are correlated and, because X^u_{it} is unobserved, Z_{it} and g_{it} will be correlated and the estimate of program impact, (, will be biased.

Jarmin (1997) specifies an econometric model that attempts to control for self selection and estimates it with data from the LRD. Administrative data for projects that occurred between 1987 and 1992 at eight MTCs in two states were matched to the LRD. All variables used in the analysis are taken from the LRD to ensure comparability across clients and nonclients, across time and across physical space. An important advantage of the LRD is that it contains data for all manufacturing plants in the US. When linked to the administrative data, it becomes a powerful tool for program evaluation.

Importantly, the two stage model explicitly controls for the client selection process. Estimates of the first stage client selection model suggest that plants located near MTCs, single unit plants, and plants that experienced high sales growth between 1982 and 1987 and had low productivity in 1987 were more likely to participate in manufacturing extension during the 1987 to 1992 period.

The second stage of the model examined the impact of extension on the growth in value added per worker between 1987 and 1992. A

---

[19] Using program service measures that vary across clients, as in Nexus Associates (1996) can help mitigate, but not alleviate this problem.

growth rate specification was used to mitigate the effects of unobserved plant heterogeneity. The second stage model is a simple production function with a dummy indicator for client status. The results suggest that clients experienced between 4 and 16% higher growth in value added per worker over this period than did nonclients.

Although the results of the econometric studies suggest that manufacturing extension has a measurable impact on client performance, one should view them with caution. What the results say is that being an extension client is "associated" with better performance. No study to date has demonstrated that extension services cause improved performance.

F. Comparisons of Evaluation Methodologies

When comparing the methodologies used to evaluate manufacturing extension, it is crucial to consider the stakeholder group for which a particular evaluation study is intended. In this section we examine the efficacy of the methodologies discussed above for both program monitoring and for measuring program performance. This discussion is summarized in Figure 2 which provides charts that plot the relative efficacies of the various methodologies versus the cost of employing them.

The horizontal axis measures the relative efficacy of the different methodologies for program monitoring and measuring program performance in the top and bottom panels, respectively.

Program monitoring requires timely and detailed information about individual projects. Administrative data are well suited to this. Case studies can also yield very detailed information about projects. Nonexperimental data and follow-up surveys provide less detailed information and the data are often only available after a considerable lag. Thus, they are not very useful for program monitoring.

Experimental data clearly provide the most reliable way to measure program performance. Unfortunately the prohibitive cost of setting up controlled experiments precludes this option in most cases. Of the other methods discussed in this paper and actually used to evaluate manufacturing extension, studies using nonexperimental data come the closest to replicating an actual experiment in that they explicitly compare the performance of extension clients to nonclient control groups. We believe these studies offer researchers and policy makers the best opportunity to assess the overall performance of programs like manufacturing extension. Follow-up surveys and case studies can also provide useful insights but the lack of a control group is a serious concern. Administrative data alone are of little use for evaluation, but are crucial for identifying client plants and measuring the type and levels of services provided.

The costs of employing the various methodologies is given on the vertical axis in figure 2. Setting up a randomized controlled experiment and collecting data is the most costly method for evaluating programs. Case studies are also quite expensive, whereas using only administrative data for evaluation is the least costly alternative. Given the limited resources most programs have available for program evaluation the combination of administrative data and pre-existing nonexperimental panel data sets appears to be an attractive evaluation strategy. These two methodologies when used together provide the best combination of functionality and cost effectiveness.

IV. Conclusions

Recent increases in the budgets for government technology programs and the Government Performance and Results Act have both increased the need to credibly measure program performance in a cost effective manner. In this paper, we discussed some of the issues involved with evaluating such programs, in the context of reviewing recent efforts to measure the performance of manufacturing extension programs. Although programs differ, we believe that many of the lessons learned from these efforts are applicable to the evaluation of any program intended to impact firm behavior and performance.

In order to more effectively design, implement and evaluate technology programs, we need to be aware of the dynamic competitive environment, in which plants and firms operate and, in which these programs are intended to function. In the case of manufacturing extension, discussions about the program, in general, and evaluation studies, in particular, have not demonstrated this awareness. To address this shortcoming, we offered a stylized model of this competitive environment and described how manufacturing extension fits into it.

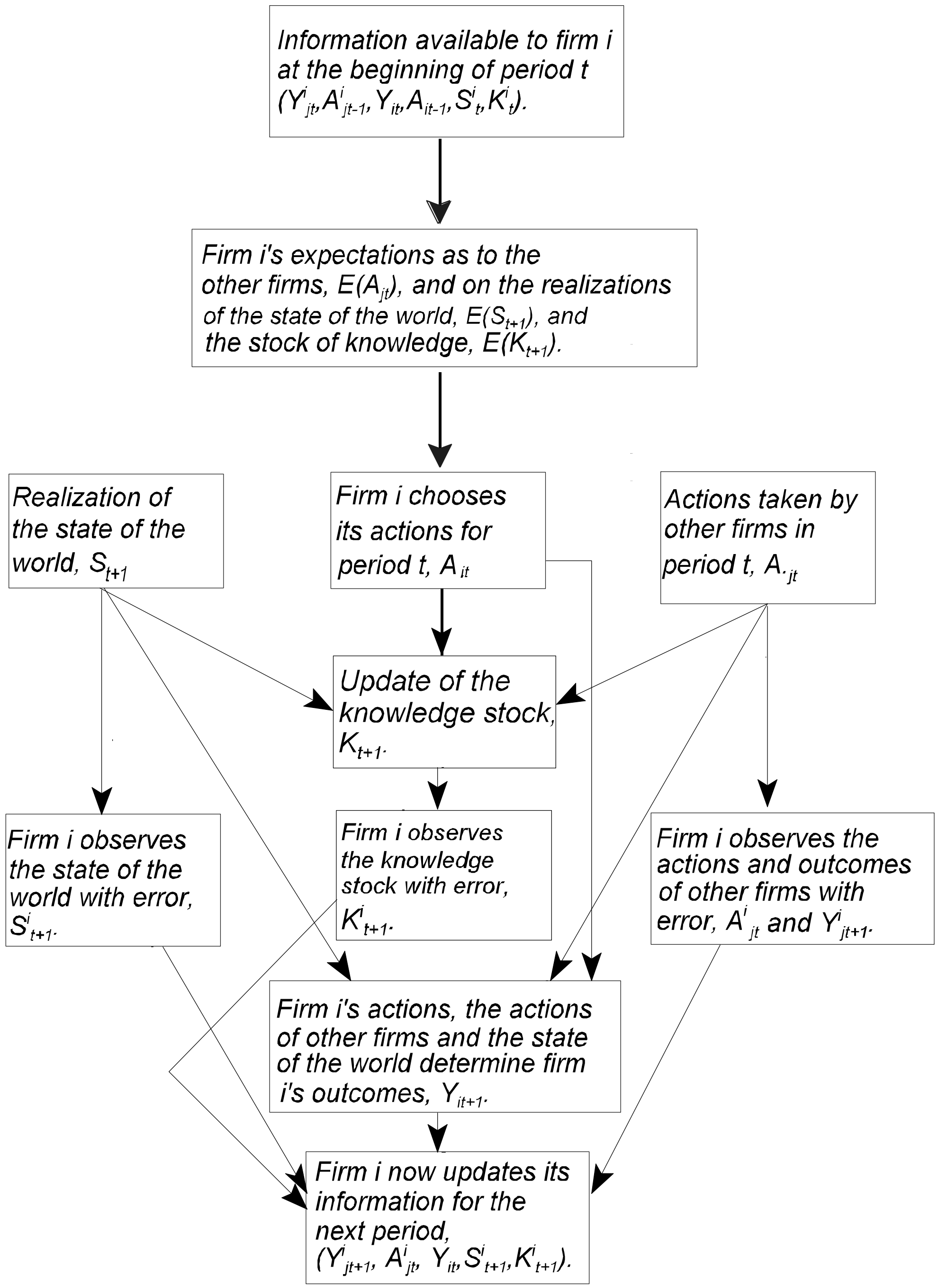
An important feature of the model is that heterogeneous plants and firms do not possess all the information relevant to their operations. One component of this information is knowledge about production technologies. Proponents of manufacturing extension argue that SMEs have less access to this type of information than do larger plants and firms. This leads SMEs to adopt new technologies more slowly than their larger counterparts which may be, at least partially, responsible for the performance gap between them. By providing information on modern production techniques and business practices, manufacturing extension programs seek to improve the ability of SMEs to make wiser choices, which hopefully leads to improved SME performance.

The stylized model has important implications for program evaluation. First, heterogenous plants and firms have differing needs for the services provided by MTCs and will rationally respond differently to their availability. This suggests that evaluation be done with plant or firm level data.

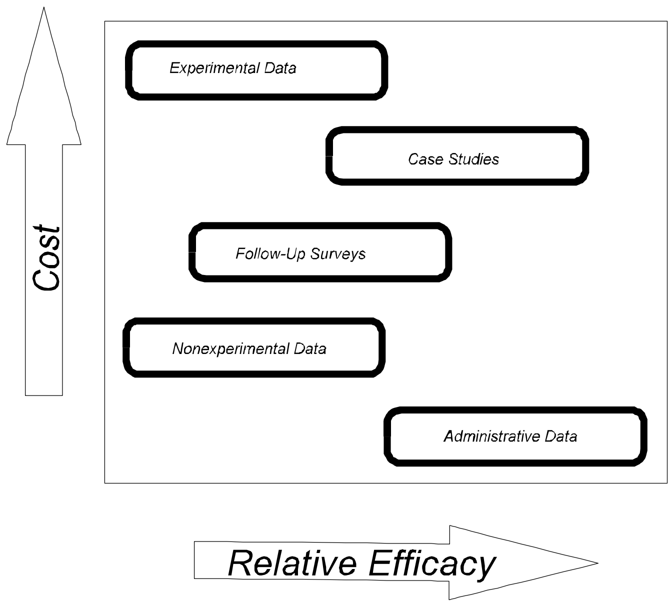
Second, performing controlled experiments is not an option for evaluating manufacturing extension. Therefore, the only way to obtain valid measures of program impact is to compare the performance of client SMEs to control groups. Of the methods used to date for evaluating manufacturing extension, only the econometric analyses have explicitly incorporated control groups. Further, much of the plant heterogeneity is unobserved by the evaluator and, thus, cannot be explicitly controlled for. This suggests the use of panel data to control for the unobserved heterogeneity. Finally, if better performing plants and firms self select into the program, then estimates of program impact may be biased. Only one study (Jarmin, 1997) has dealt explicitly with the issue of selection bias.

We conclude that the best method currently available to evaluate the overall effectiveness of technology programs, such as manufacturing extension, is to match administrative data to existing plant level panel data sets like the Census Bureau's Longitudinal Research Database. This methodology provides the means to compare the performance of client SMEs to non clients while controlling for both observed and, because of the panel feature of the data, unobserved plant characteristics. As such, the results from carefully done studies of this type of study are more credible than case studies and client surveys where no control groups are employed.

Figure 1
Stylized Model


Information available to firm i at the beginning of period t ($Y^i_{jt}, A^i_{jt-1}, Y_{it}, A_{it-1}, S^i_t, K^i_t$).
Firm i's expectations as to the other firms, $E(A_{jt})$, and on the realizations of the state of the world, $E(S_{t+1})$, and the stock of knowledge, $E(K_{t+1})$.
Realization of the state of the world, S_{t+1}
Firm i chooses its actions for period t, A_{it}
Actions taken by other firms in period t, $A_{\cdot jt}$
Update of the knowledge stock, K_{t+1}.
Firm i observes the state of the world with error, S^i_{t+1}.
Firm i observes the knowledge stock with error, K^i_{t+1}.
Firm i observes the actions and outcomes of other firms with error, A^i_{jt} and Y^i_{jt+1}.
Firm i's actions, the actions of other firms and the state of the world determine firm i's outcomes, Y_{it+1}.
Firm i now updates its information for the next period, (Y^i_{jt+1}, A^i_{jt}, Y_{it}, S^i_{t+1}, K^i_{t+1}).

Efficacy for Program Monitoring



Efficacy for Measuring Program Performance

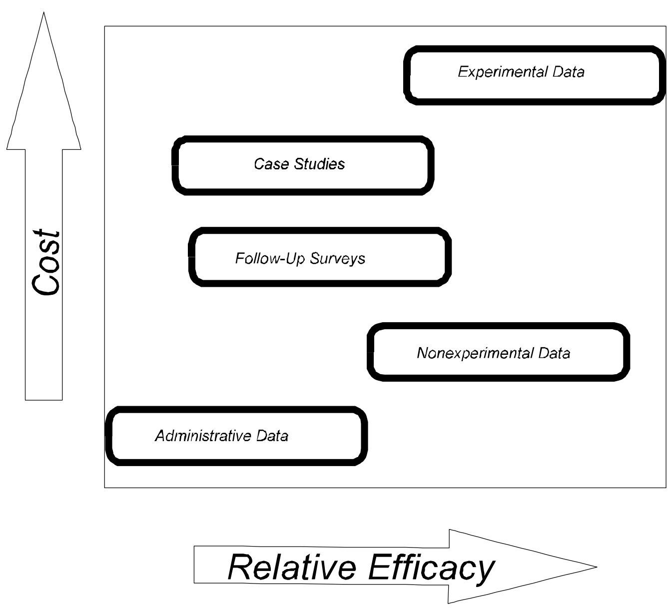


Figure 2
Relative Efficacy of Different Evaluation Methodologies

Table 1
Review of Evaluation Methodologies

Methodology	Advantages	Disadvantages	Studies
Administrative Data	**1.** Provide detailed information on the type and amounts of services provided, typically on a project by project basis. **2.** Provide the ability to identify clients in third party databases such as the LRD. **3.** Provide a sample frame for client follow-up surveys.	**1.** No control group. **2.** Typically limited to describing activity levels. **3.** Data are often not comparable across different extension centers.	Nexus Associated (1996), Centers and NIST/MEP
Case Studies	**1.** Provides very detailed information about the mechanisms through which program services affect client decisions and performance.	**1.** No control group. **2.** Relies too heavily on "success stories."	Oldsman (1996), NIST/MEP collection of "Success Stories" on the WWW
Client Follow up Surveys	**1.** Allows questions to be tailored to the interests of MEP stakeholders. **2.** Provides valuable feedback to MEP field staff.	**1.** No control group. **2.** Recall bias. **3.** Expensive.	Oldsman (1996) and Nexus Associates (1996), Youtie and Shapira (1997) and Census Survey
Econometric Studies with Non-Experimental Data	**1.** Control group. **2.** Can utilize existing data sources (economical). **3.** System wide evaluation.	**1.** Selection bias. **2.** Can only examine variables collected from both client and non-clients.	Jarmin (1995 and 1997), Luria and Wiarda (1996), Nexus Associates 1996)

Methodology	Advantages	Disadvantage	
Experimental Methods	**1.** Control group. **2.** Random assignment avoids the problem of selection bias.	**1.** Not a practical option for evaluating manufacturing extension.	None

REFERENCES

Berglund, D., and C. Coburn, (1995), Partnerships: A Compendium of State and Federal Cooperative Technology Programs, Battelle Press, Columbus, OH.

Feller, I., Glasmeier, A., and M. Mark, (1996), “Issues and Perspectives on Evaluating Manufacturing Modernization Programs,” Research Policy, 25 pp. 309-319.

Feller, I., (1997), “Manufacturing Technology Centers as Components of Regional Technology Infrastructures,” Regional Science & Urban Economics, 27, pp. 181-197.

Jaffe, A., (1989), “Real Effects of Academic Research,” American Economic Review, 79, pp. 957-970.

Jarmin, R.S., (1995), “Using Matched Client and Census Data to Evaluate the Performance of the Manufacturing Extension Partnership,” Center for Economic Studies Working Paper CES 95-7.

Jarmin, R.S., (1996), “Leaning by Doing and Plant Characteristics,” Center for Economics Studies Working Paper CES 96-5.

Jarmin, R.S. (1997), “Evaluating the Impact of Manufacturing Extension on Productivity Growth,” Journal of Policy Analysis and Management, (forthcoming).

Jensen, J.B., and R.H. McGuckin, (1997), “Firm Performance and Evolution: Empirical Regularities in the U.S. Microdata, Industrial and Corporate Change, 6(1), pp. 25-47.

Luria, D., and E. Wiarda, (1996), “Performance Benchmarking and Measuring Program Impacts on Customers: Lessons from the Midwest Manufacturing Technology Center,” Research Policy, 25, 233-246.

Maddala, G.S., (1983), Limited Dependent Variables and Qualitative Variables in Econometrics, Cambridge University Press, New York.

McGuckin, R.H., and J. Redman, (1995), "Measuring and Monitoring Client Firm Competitiveness, mimeo, Center For Economic Studies, Washington, DC.

Moffitt, R., (1991), "Program Evaluation with Non-experimental," Evaluation Review, 15(3), pp. 291-314.

National Research Council, (1993), Learning to Change: Opportunities to Improve the Performance of Smaller Manufacturers, Commission of Engineering and Technical Systems, Manufacturing Studies Board, National Academy Press, Washington, DC.

Nexus Associates, (1996), "Evaluation of the New York Manufacturing Extension Partnership, mimeo, Nexus Associates, Belmont, MA.

Oldsman, E., (1996), "Does Manufacturing Extension Matter? An Evaluation of the Industrial Technology Extension Service in New York," Research Policy, 25, 215-232.

Shapira, P., Youtie, J., and J.D., Roessner, (1996), "Current Practices in the Evaluation of US Industrial Modernization Programs, Research Policy, 25, pp. 185-214.