Chapter 14

EVALUATING GOVERNMENT TECHNOLOGY PROGRAMMES: THE CASE OF MANUFACTURING EXTENSION

by

Ronald S. Jarmin
Center for Economic Studies, US Bureau of the Census, Washington, DC
and
J. Bradford Jensen
Heinz School, Carnegie Mellon University, Pittsburgh

Introduction

In 1994, federal and state governments in the United States spent approximately US$2 848 million[1] on programmes designed to encourage businesses to discover, develop and adopt improved technologies and business practices, and this sum is growing. It is important to taxpayers and policy makers to see if these technology programmes have the desired impact and that this money is well spent. Unfortunately, very few programmes are ever credibly evaluated. This paper discusses the issues involved in evaluating such programmes in the context of reviewing recent efforts to evaluate a particular programme: Manufacturing Extension. While differences exist between this and other technology programmes, we believe that many of the lessons learned in evaluating manufacturing extension can be applied elsewhere.

Contributions of this paper

By reviewing these efforts, we hope to provide some perspective that we believe has been absent in both discussions about manufacturing extension, in general, and in evaluation studies, in particular. Although there have been previous reviews of studies evaluating manufacturing extension (see Shapira, Youtie and Roessner, 1996; Feller, Glasmeier and Mark, 1996), we believe that this paper makes two important contributions.

First, considerable progress has been made since these earlier reviews were written. The quality of the data available for evaluating manufacturing extension programmes has improved greatly and many of the recent studies employ more rigorous econometric techniques than previous analyses. Thus, we review some of the important recent contributions, 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 programme stakeholders.

Second, much of the evaluation work to date ignores the competitive environment within which client plants and firms operate, and where the data available for programme evaluation are generated. We believe that it is important to have an understanding of this environment in order to optimally design, provide and evaluate programme services and to ensure that programme objectives are not at odds with those of the client plants and firms the programme is intended to serve (*e.g.* the policy maker's desire to promote job creation/retention may not always be consistent with the firm's desire to increase productivity and profits). Recent years have seen a large number of both theoretical and empirical studies in economics that suggest that this competitive environment is very dynamic, and that the plants and firms that operate within it are very heterogeneous (see Jensen and McGuckin, 1996). Policy makers, programme managers and researchers need to be cognisant of this when they, respectively, design, implement and evaluate technology programmes.

To address this shortcoming in the literature concerning manufacturing extension, we provide a stylised model of the environment in which client plants and firms operate. We then show how manufacturing extension fits into this stylised model. Finally, the model provides an excellent framework for our discussion of recent evaluation studies.

What is manufacturing extension?

Before characterising this framework, we briefly describe manufacturing extension. Manufacturing extension is the term used to describe the collection of organisations that provide industrial modernisation 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 US manufacturing industries. The MEP supports several manufacturing technology centres (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 centres can also channel more recent innovations generated in government and university laboratories to smaller US 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 market-place.

In order to maximise the effectiveness of the programme, it is crucial that programme stakeholders (*e.g.* client SMEs, MTCs, NIST, state and local governments and Congress) have detailed information about its current performance, and that a reliable evaluation framework be in place to analyse its future performance. However, different programme 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 single evaluation methodology is suitable to meeting the needs of all programme stakeholders. Our discussion below primarily examines the advantages and disadvantages of the various methodologies for determining the overall effectiveness of manufacturing extension.

Methodologically, the most reliable way to measure the effectiveness of programmes is to collect experimental data. Namely, plants and firms would be randomly assigned to treatment and control groups. Evaluation would then consist of a simple comparison of the performance of treatment and control firms. Unfortunately, this has not been done, nor is it likely to be done, for manufacturing extension. Instead, evaluation has proceeded using the non-experimental methods we discuss below. Of these, we believe that careful econometric studies incorporating control groups are the most reliable method available for evaluating the overall effectiveness of technology programmes such as manufacturing extension.

Stylised model of competition

We begin this section by providing a highly stylised model of the competitive environment US firms operate in, and in which these policies and programmes are intended to work. Next, we characterise the dynamics of firm decision making within this competitive environment. Finally, we discuss how one government programme, manufacturing extension, fits into this dynamic framework.

Basic framework for the model stylised model

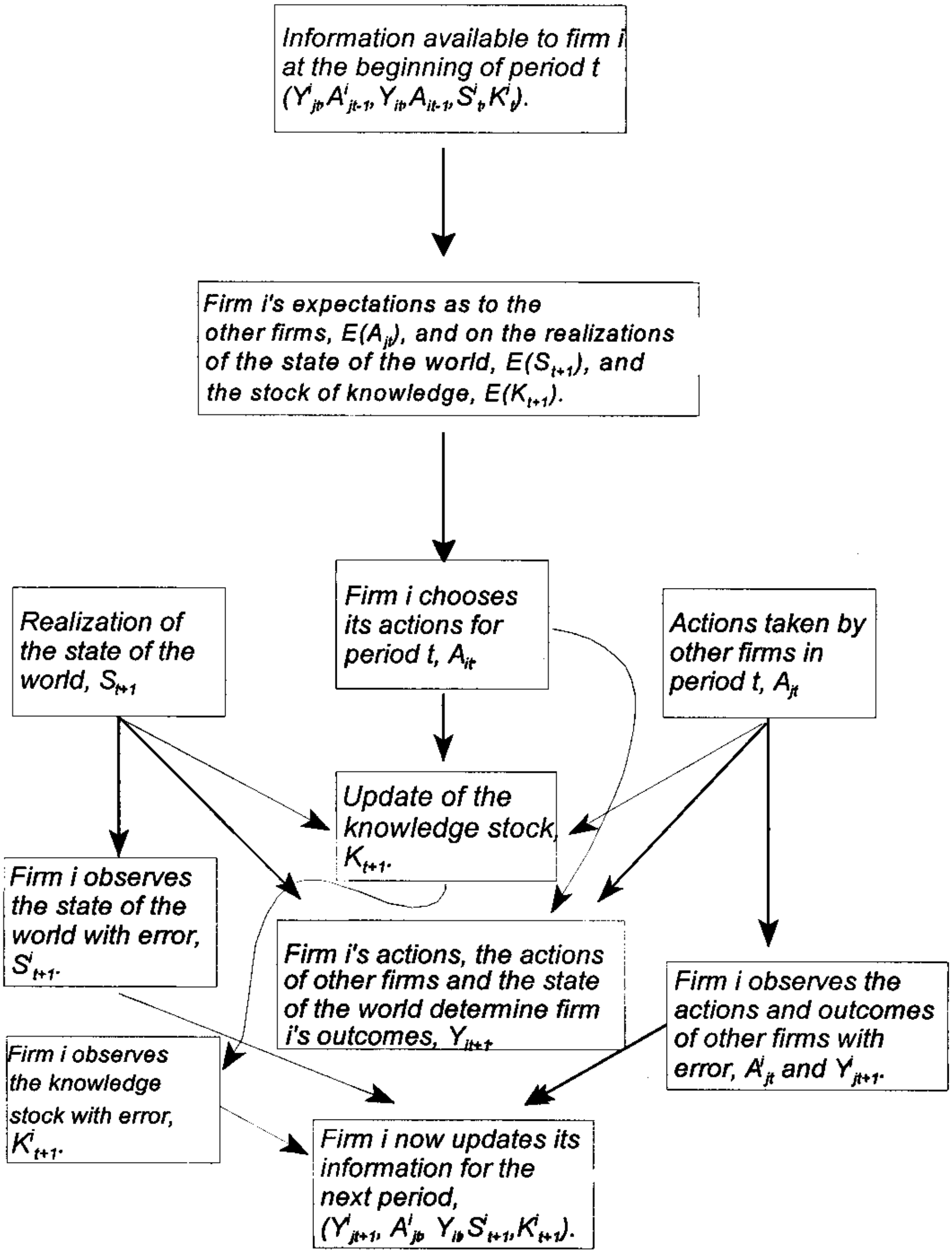
Our starting point is to note that firms and their constituent plants are very heterogeneous, even within industries. Important observable dimensions of this heterogeneity include, but are not limited to, size, age, capital intensity, productivity and wages. Jensen and McGuckin (1996) review an extensive 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[2]. This heterogeneity causes firms to respond differently to common shocks[3] and to government policies and programmes. In the case of programmes, such as manufacturing extension, plants and firms will respond differently to the availability of programme services. Some will choose to take advantage of the programme'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 programme. The different strategies pursued by heterogeneous firms cause their observable (and unobservable) characteristics to evolve differently over time. The implication for programme 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 stylised model should incorporate as many features of this environment as possible.

Figure 1 provides a flow chart of the basic components of our stylised model. It shows what takes place during a single decision-making period[4] for an arbitrary firm. At the beginning of the period (indexed by t in Figure 1), the firm (indexed by i in Figure 1) 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 realisation 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.

Figure 1. Stylised model



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 available to the firm. At the beginning of each decision-making period, firm i has an information set: $I_{it} = \{Y_{it}, Y^i_{jt}, A_{it-1}, A^i_{jt-1}, S^i_t, K^i_{it}\}$, 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 sub-set ($Y^g_{it} \subset Y_{it}$) of any firm's characteristics is observed by organisations, such as trade associations and government statistical agencies. Generally, econometricians and policy analysts exploit these sources of data to analyse firm behaviour and performance and how government programmes 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.* labour 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.* labour, 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.[5] 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).

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 maximise its discounted profit stream:

$$\begin{aligned} V_{it} &= \pi_{it}\left(A_{it}, A_{jt}, S_t, Y_{it}, I_{it}\right) + \delta_i V_{it+1} \\ s.t. = Y_{it} &= f\left(Y_{it-1}, A_{it-1}, A_{jt-1}, S_{t-1}, K_{it}\right) \\ I_{it} &= g\left(I_{it-1}, Y_{jt}^i, A_{jt-1}^i, Y_{it}, A_{jt-1}, S_{it}^i, K_{it}^i\right) \\ and\ K_t &= h\left(K_{t-1}, A_{it}, A_{jt-1}, S\ \right) \end{aligned} \quad (1)$$

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 time. The firm's single period profit function is given by π_{it} and it depends on the current actions of it and its rivals, the realised state of the world and its beginning of period characteristics. Discounted future profits are captured in $\delta_i V_{it+1}$ where δ_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 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.[6]

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 maximise 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.

The role of manufacturing extension in this framework

Having characterised the stylised model, we would now 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 counterparts.[7] In the example above, SMEs either do not observe the technology shock, do so with more error, or may be less able than large firms to process information about the shock and formulate the most profitable response.

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.[8] With this additional information, extension clients can then choose actions that lead to better outcomes.

We assume that policy makers have some sub-set, $Y^p_{it} \subset Y^g_{it}$ (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 centre provides i with additional information on the stock of technological knowledge. The 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.

Evaluation methodologies

The question we are interested in is this: How can policy makers know with reasonable confidence that improvements in the performance of client plants or firms are due to participation in the programme? In this section, we contrast the methodologies used in various studies that seek to evaluate manufacturing extension.[9]

Our discussion is from the point of view of trying to assess the overall effectiveness of manufacturing extension. However, it is important to recognise that various stakeholder groups associated with programmes such as manufacturing extension have different needs concerning programme evaluation. For example, the staff of the extension centres 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 knowing about short-term, intermediate outcomes, such as whether a client bought a new machine as a result of extension services. In contrast, the policy maker wants to know whether extension services led 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 with whether or not the actions ultimately led to the desired policy outcome.

An 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 policy makers (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 programmes 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 non-experimental data. The last four have all been exploited in studies attempting to assess the effectiveness of manufacturing programmes in the United States. Table 1 summarises the major evaluation methodologies and lists some of their characteristics.

Table 1. Review of evaluation methodologies

Methodology	Advantages	Disadvantages	Studies
Administrative data	1. Provides detailed information on the type and amounts of services provided, typically on a project-by-project basis. 2. Provides the ability to identify clients in third-party databases such as the LRD. 3. Provides 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 centres.	Nexus Associated (1996), Centers and NIST/MEP
Case studies	1. Provides very detailed information about the mechanisms through which programme 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 utilise 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)
Experimental methods	1. Control group. 2. Random assignment avoids the problem of selection bias.	1. Not a practical option for evaluating manufacturing extension.	None

Experimental methods

Recall that the goal of any effort to evaluate technology programmes is to determine whether observed changes in the behaviour and performance of client firms can be attributed to their association with the programme. The most reliable method available is to randomly assign plants and firms to treatment (those that receive services) and control groups. One then analyses the experimental data to see if the behaviour 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 programme's impact, he/she could simply compute the mean value added per worker for both the treatment and control groups and compare them.

Because of random assignment, he/she can be confident that any observed difference in value added per worker is attributable to programme 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.

Administrative data

Many programme 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 artefact 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 programme 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 databases also vary considerably across individual manufacturing extension centres, which limits cross centre 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 programmes.

Nevertheless, recording activity levels, the types of services provided and other administrative data is crucial to programme 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 minimising the reporting burden, is critical to subsequent evaluation efforts and essential for MTC and client co-operation.

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.[10] 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, implementation of statistical process control and on-the-job training programmes, 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 small share 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 non-client firms. Finally, case studies are often criticised 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 assess the overall effectiveness of a programme.

Client surveys

Another method used to collect information and evaluate programmes is the client follow-up survey. Client follow-up surveys typically survey all or some sub-set of the client population.[11] Follow-up surveys have the advantage that the questionnaires can be customised to collect detailed information on both intermediate outcomes (efforts at job training, investment in technology, etc.) and final outcomes (increased sales, increased employment, etc.). Client follow-up surveys can be performed by service providers or collected by independent third parties.[12]

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 centre 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 non-clients in the survey. However, this increases the cost of the survey. Another is to structure the questionnaire to provide time-series information on client behaviour. 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 it requires respondents to remember details of their operations prior to receiving services. Thus, if not carefully constructed, these questions can produce responses of dubious value.

Follow-up surveys have been commissioned and performed by several state programmes. 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 but 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-centre comparisons. This should aid both NIST-MEP and the individual MTC in identifying which services provided by manufacturing extension are effective and which are not.

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 non-experimental data sets. Of the methods discussed here and used to evaluate manufacturing extension, these come the closest to replicating an actual experiment in that they explicitly compare the performance of extension clients to non-client control groups. We believe that these studies offer researchers and policy makers the best opportunity to assess the overall performance of programmes such as manufacturing extension.

Jarmin (1995) discusses the modelling 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 some economic theory that suggests how these variables are related within the framework described above, the researcher then specifies an econometric model and estimates the impact of programme participation on the relevant characteristic or policy variable. For example, the researcher may estimate a regression model like the following:

$$Y_{it}^{p} = \beta^{\circ} X_{it}^{\circ} + \beta^{u} X_{it}^{u} + \gamma Z_{it} + \varepsilon_{it} \qquad (2)$$

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, $X^{o}_{it} \subseteq Y^{g}_{it}$, 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.* $X_{it}^{u} \notin Y_{it}^{g}$), Z_{it} is a programme participation variable, β^{o} and β^{u} are vectors of parameters, γ is the parameter that measures programme impact and ε_{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, γ measures the mean difference in Y^{p} between clients and non-clients controlling for the characteristics measured in X. If the researcher controlled for all the differences between clients and non-clients, other than client status, then γ will be an unbiased estimate of programme 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[13] of plants or firms and track them over time. One would collect data on several measures including information about programme participation from both clients and non-clients. Surveys such as this are very expensive, typically exceeding the evaluation budgets of most programmes.

Second, one can use pre-existing data sets. While these may not be designed and maintained specifically for programme evaluation, they are much less costly than special surveys. In the case of manufacturing extension, two such data sets have been utilised: *i)* the Performance Benchmarking Database maintained by the Industrial Technology Institute (ITI); and *ii)* 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 subsidises 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 non-clients. 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 non-clients.[14] However, they found no link between programme 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 non-client 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 to be subject to different shocks than plants and firms in New York.

These two studies are restricted to a small number of client and non-client observations. Further, as in many evaluation studies, the estimates of programme impact may be biased due to self-selection.[15] Self-selection arises when plants are not randomly assigned to client (treatment) and non-client (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 estimates of the impact of the programme on performance.[16] That is, Z_{it} and X^u_{it} in equation (2) are correlated and, because X^u_{it} is unobserved, Z_{it} and ε_{it} will be correlated and the estimate of programme impact, γ, 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 non-clients, across time and across physical space. An important advantage of the LRD is that it contains data for all manufacturing plants in the United States. When linked to the administrative data, it becomes a powerful tool for programme 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 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 per cent higher growth in value added per worker than did non-clients.

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 caused the improved performance.

Conclusions

As the budgets for government technology programmes have grown, so has the need to credibly demonstrate their effectiveness. In this paper, we have discussed some of the issues involved with evaluating such programmes in the context of reviewing recent efforts to measure the effectiveness of manufacturing extension. Although programmes differ, we believe that many of the lessons learned from these efforts are applicable to the evaluation of any programme intended to impact firm behaviour and performance.

In order to more effectively design, implement and evaluate technology programmes, we need to be aware of the dynamic competitive environment in which plants and firms operate and in which these programmes are intended to function. In the case of manufacturing extension, discussions about the programme, in general, and evaluation studies, in particular, have not demonstrated this awareness. To address this shortcoming, we offered a stylised 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 programmes seek to improve the ability of SMEs to make wiser choices, which hopefully leads to improved SME performance.

The stylised model has important implications for programme evaluation. First, heterogeneous 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 programme impact is by comparing 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. If better performing plants and firms self-select into the programme, then estimates of programme impact may be biased. Only one study (Jarmin, 1997) has dealt explicitly with the issue of selection bias.

A particularly important unobserved characteristic is whether clients also receive assistance from non-extension sources (*e.g.* consultants and suppliers). If this is the case, it is difficult to obtain an estimate of the impact of the programme that is independent of these other influences. None of the studies performed to date have controlled for non-extension services.

We conclude that the best method currently available to evaluate the overall effectiveness of technology programmes, such as manufacturing extension, is to match administrative data to existing plant-level panel data sets like the LRD. 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 analysis are more credible than case studies and client surveys where no control groups are employed.

Nevertheless, there is considerable room for improvement in these studies. First, both the administrative and third-party data can be improved. On the administrative data side, NIST/MEP should continue to push for the collection of accurate data on project characteristics (*e.g.* type of project, dosage of treatment) that are consistent across MTCs.

On the third-party side, both the PBS data and the LRD can be improved. The PBS data set, while it contains an impressive array of metrics useful for evaluation, has limited geographic, industry and panel coverage. Clearly, the utility of this data set would be improved immensely if it had more respondents, broader geographic and industry coverage and more observations per plant (*i.e.* better panel characteristics, which might be achieved by matching it to the LRD). The LRD has broader geographic, industry and more complete panel coverage, but for a smaller set of metrics. If the LRD is going to find more use in programme evaluation studies, we would urge the Census Bureau (and other statistical agencies) to take this into account in their data collection and processing operations. This would involve making the construction and maintenance of longitudinal microdata sets an explicit goal of these efforts.

Finally, the studies themselves should be extended and improved. Analysis should incorporate updated data and additional MTCs. Further, researchers should estimate the impact of extension services on additional variables, such as plant survival, and do more to control for the type and intensity of the services provided. Work on some of these issues is underway. However, work on controlling for non-extension services awaits appropriate data.

NOTES

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

2. 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 unforeseen events such as machinery break downs, strikes and the like that affect an individual plant or firm.

3. Common shocks include unforeseen changes in consumer tastes, the business cycle, wars, and so on.

4. Firms may differ in the length of their decision-making periods and an individual firm may use different planning horizons for different types of 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.

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

6. 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 shocks in different ways.

7. The National Research Council (1993) cites five barriers to improved performance of 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.

8. 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 organisations. However, proponents of manufacturing extension argue that the private sector provides an inefficiently small quantity of industrial modernisation services to SMEs (see National Research Council, 1993; McGuckin and Redman, 1995). Thus, they argue that there is a role for state and federal support for MTCs.

9. See McGuckin and Redman (1995) for a discussion of the characteristics of a "good" evaluation of manufacturing extension.

10. These are accessible on the NIST-MEP World-Wide Web page at http://vehicles.nist.gov/internet/success.nsf/$about.

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

12. An obvious advantage of having a third party conduct the survey is objectivity.

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

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

15. 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 programmes and discuss the issue of self-selection.

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

REFERENCES

BERGLUND, D. and C. COBURN (1995), *Partnerships: A Compendium of State and Federal Co-operative Technology Programmes*, Battelle Press, Columbus, OH.

FELLER, I., A. GLASMEIER and M. MARK (1996), "Issues and Perspectives on Evaluating Manufacturing Modernization Programmes", *Research Policy*, 25, pp. 309-319.

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 Economic Studies Working Paper CES 96-5.

JARMIN, R.S. (1997), "Evaluating the Impact of Manufacturing Extension on Productivity Growth", mimeo, Center for Economic Studies, Washington, DC.

JENSEN, J.B. and R.H. McGUCKIN (1996), "Firm Performance and Evolution: Empirical Regularities in the US Microdata", mimeo, Center For Economic Studies, Washington, DC.

LURIA, D. and E. WIARDA (1996), "Performance Benchmarking and Measuring Programme Impacts on Customers: Lessons from the Midwest Manufacturing Technology Center", *Research Policy*, 25, pp. 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), "Programme Evaluation with Non-experimental Data", *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, pp. 215-232.

SHAPIRA, P., J. YOUTIE and J.D. ROESSNER (1996), "Current Practices in the Evaluation of US Industrial Modernization Programmes", *Research Policy*, 25, pp. 185-214.

STROMSDORFER, E.W. (1987), “An Overview of Recent Findings and Advances in Evaluation Methods”, *Evaluation Review*, Vol. 11, No. 4, pp. 387-394.