

Measuring Efficiency in Government: Techniques and Experience*

Jack Diamond

BUDGETING FOR PERFORMANCE

Traditional views of the budget and budgeting procedures have undergone some recent changes. The previous view of the budget was as an exercise primarily in resource allocation and input control, which was usually highly centralized. This approach to the budget accepted a set of policy objectives-often poorly articulated and usually unquantified-and allocated inputs to reach those objectives. At the same time, central budgeting agencies focused almost exclusively on control and compliance as the primary modus operandi in financial management. There was often little follow-up in examining the subsequent performance of spending departments. The new trend, indicated here as “performance budgeting,” but with different titles in different countries,¹ aims to forge a more direct link between the allocation of resources through the budget and performance in reaching objectives. This reorientation has

* This article is a chapter of the book *Government Financial Management*, A. Premchand (Ed.), International Monetary Fund (IMF), Washington, 1990. It is reproduced here with non-profit purposes and it is also available in: <http://books.google.com.ar/books?hl=es&lr=&id=wrOgKfzv1GAC&oi=fnd&pg=PA142&dq=Measuring+Efficiency+in+Government:+Techniques+and+Experience+%2B+jack+diamond&ots=kCCN3X50NZ&sig=-ChIHLRIYE6h6C10OU1sTKY4cUM>

¹ In Australia, the Financial Management Improvement Program (FMIP); in Canada, the Increased Ministerial Authority and Accountability Reforms (IMAA); in the United Kingdom, the Financial Management Initiatives (FMI); and in the United States, the Productivity Improvement Program (PIP). For a fuller discussion of country experiences, including those of Denmark and Sweden, see Organization for Economic Cooperation and Development (1988).

necessitated a number of changes in traditional budgeting. Most noticeably, there has been a greater emphasis on measures of performance. Indeed, the basis of recent efforts to improve the management of government services has focused on developing measures of output and on assessing productivity in government.

A number of reasons can be advanced for this change in focus.² First, central budgetary agencies had a major incentive to concern themselves with productivity in the face of “fiscal stress” that they were experiencing. As operations and maintenance expenditures were squeezed, governments became concerned that this squeeze was resulting in reduced levels of service rather than gains in productivity. To check that the financial restraint was indeed resulting in greater efficiency necessitated renewed attempts to measure productivity. Second, in an environment of budgetary cuts there was heightened concern about the relative price effect (or Baumol’s disease),³ and the view was widespread that government services could not match the productivity gains of the private sector, with a consequent increase in the relative size of government. Third, institutionally, measuring productivity was found to be less controversial than trying to measure the effectiveness of programs in attaining their ultimate objectives. The former approach did not question whether present activities were actually effective in reaching the desired goals. Moreover, it focused entirely on those aspects that were most under the control of the resource managers, whereas performance and effectiveness could be more sensitive to exogenous factors.

The new emphasis on performance indicators had its genesis in previous attempts at improving budgeting. At first, the focus was on the management of inputs, where economizing was the prime concern. This approach was followed by attempts to manage outputs, with spending departments concentrating on work loads. Most recently, there have been more sophisticated attempts to relate inputs to outputs, with improved efficiency as the primary

² Elaborated in Premchand (1987).

³ See Baumol (1967), where it is demonstrated that if labor productivity is assumed constant in the government sector and that of the private sector rises exponentially, then if government goods and services in real terms are to remain as fixed proportions of total goods and services, government expenditures as a proportion of total expenditure must rise.

goal. The current renewed emphasis on performance signals an attempt to go one stage further and to assess the effectiveness of outputs in reaching objectives. While the latter stage is still in its infancy, over the last decade there have been numerous attempts aimed at improving efficiency in providing government services. It is hoped that in reviewing this practical experience much can be learned about the obstacles likely to be encountered in developing indicators of performance and analyzing the effectiveness of resource allocations.

MEANING OF EFFICIENCY

Economists generally distinguish two distinct concepts. First, there *is technical efficiency*, which is met when the maximum possible output is produced with a given supply of resources. It should not be possible to reduce any input without reducing the volume of the output. Technical inefficiency in terms of not being able to attain this level of resource use for a given output is symptomatic of more general shortcomings in internal organization of the producing unit (whether owing to political interference, bureaucratic procedures, wage and personnel rigidities, etc.). This type of inefficiency has been labeled X-inefficiency by Leibenstein, who argues that it is likely to be the greatest source of inefficiency, particularly in the public sector.⁴ Second, there is the concept of *allocative efficiency*, which moves from physical quantity measures to the costs of inputs. Allocative efficiency is met when the cost of any given output is minimized by combining inputs in such a way that one input cannot be substituted for another without raising costs.

Using this theoretical framework, and following the methodology of Farrell (1957), it is possible to devise measures of relative efficiency. If an agency is on its technical output frontier, such that with given inputs it could not be expected to produce more output, its efficiency score would be 100 percent. Shortfalls from this technically efficient output level would result in correspondingly lower scores. In comparisons of relative efficiency, it is essential to distinguish output from input efficiency. As described above, efficiency has two

⁴ See Leibenstein (1966).

dimensions: the management unit can minimize inputs to reach a given output, or maximize its output with given inputs. *Output efficiency* is therefore the ratio of the unit's actual output to expected output, given its inputs. *Input efficiency*, on the other hand, is defined in terms of the ratio of its expected inputs to its actual input, given its output. Only under very restrictive assumptions will the efficiency score of a management unit in terms of its input efficiency match its score in terms of output efficiency.⁵

Only under stringent perfect market conditions with multiple suppliers and demanders of goods and services is it possible to maximize efficiency. In practice it is difficult to ascertain whether allocative efficiency is being attained in the public sector. Rather, in most cases in the public sector, performance can only be measured in technical terms relative to other agencies offering similar services. However, with no "ideal" or "standard" performance in the production of government services, and where competitive conditions obviously do not exist, the most that can be achieved is to develop management and information systems that mirror as closely as possible conditions that will maximize efficiency. It has been found that to do so causes great demand for certain types of information:

- quantify intended outputs and associated costs;
- quantify change in output associated with costs;
- ascertain whether output maximization or input minimization is the governing objective;
- discover the scope for improving technical efficiency;
- assess with a given output whether inputs can be substituted for one another to reduce total costs;
- quantify the extent to which efficiency can be improved by bringing worst performers closer to the average.

In answering such questions, the measurement problems faced and the difficulties encountered in applying quantitative techniques have come to be recognized as a critical factor in adopting a performancebased approach to budgeting.

⁵ A demonstration is contained in Levitt and Joyce (1987), pp. 95ff.

MEASURING PERFORMANCE IN TERMS OF EFFICIENCY

Conceptually, the term “efficiency” appears clear cut and precise. An efficient allocation of resources is one where the greatest possible output is obtained from available resources (or where a given output cannot be produced with less input). In practice, there is real ambiguity about the output of government services and, as a consequence, the linkage between resource allocation and performance is far from direct and explicit.

Producing government services does not conform to the simple technical production function relationship linking inputs to outputs, described in the previous section. Rather, government outputs are often social consequences that are difficult to define precisely and tend to be multidimensional and interrelated. At the same time external social influences are also at work, masking the impact of government activity. As a result the output of any government good or service must often be measured along a number of dimensions by often imperfect, incomplete, multiple proxy indicators. It is perhaps with these considerations in mind that national accountants typically measure real government output by reference to input prices and volumes and avoid output measures. However, it is far from satisfactory. There is no reason in principle why input and output values should change in proportion, and conversely, there is no reason to assume zero-productivity gains and the existence of constant returns to scale in government.

One can characterize the production of a public service as resulting from a process that first begins with the purchase of inputs (labor, capital, etc.) to support certain activities (patients treated, benefits paid, children taught, etc.) that have certain social consequences that are the final output of the process (relief of poverty, better health, improved knowledge, etc.). While the measurement of inputs is the most straightforward and is usually reflected in costs, the final output is much more problematic. Often, without precisely definable impacts, analysts have settled for intermediate measures of output that describe the activities of government agencies (benefits paid per employee, patients treated per doctor, children taught per teacher, etc.). However, insofar as these intermediate outputs could be viewed as the inputs to reaching the government’s final objectives, the simple discussion of the meaning of efficiency above needs further elaboration.

To a large extent, translating efficiency into operational terms has floundered on this conceptual distinction between outputs and inputs. The confusion is encountered throughout the public service literature. For example, take the case of crime prevention, where certain inputs (men and equipment) are combined in certain activities (patrol units) to produce an output (crime prevention). In some cases, as in Bradford, Malt, and Oates (1969), police patrols are cited as “outputs.” Craig (1987), in analyzing police distribution, in his terminology called the number of police patrols an input toward a final output, crime prevention. Thus the “clearance rate,” the percentage of crimes that are considered solved, usually by arrests, is not the final product—which is public safety—but is merely a stage along the way. For him, the clearance rate is an “intermediate” output (p. 335). The amount of “output” (final output to consumers), that is, public safety, is measured by subtracting the crime rate per capita from a constant term, this term being “slightly larger than the maximum crime rate” (p. 335). To add to the confusion, what Craig terms “output” is referred to as “outcome” in Behrman and Craig (1987, pp. 4243). In other areas what the latter term “outcome” is called “impact.”

This distinction between outputs and outcomes (or impacts) has important implications for the concept of efficiency. Efficiency is usually defined as an optimal input-output relation that takes into account all factors of production used in outputs. Obviously, it is of critical importance how output is defined. If, as was implied above, output is defined as intermediate output, there is no problem. But if the focus of interest is in the ultimate attainment of policy objectives, perhaps another term should be used, such as “effectiveness,” which has been suggested (Burkhead and Ross, 1974), and which is now used extensively. Thus in public expenditure management it has been useful to distinguish three concepts: economy refers to the actual use of resources in relation to planned levels; efficiency refers to the use of resources in relation to outputs; and effectiveness refers to the impact on and fulfillment of program objectives.⁶

Measures of effectiveness or “performance indicators” are at an early stage of development. Part of the problem lies in developing practically applicable indicators of the degree to which policy objectives have been attained. In much of the public sector,

⁶ See discussion on the meaning of efficiency in International Monetary Fund, 1988, pp. 26ff.

objectives have been defined in generally vague and noneconomic terms, and it is partly for this reason that most empirical work has focused on inputs and processes and has little to say about outputs and objectives. Although some work has been done by evaluators in defining desirable properties of performance indicators,⁷ it has been overcome by a number of complications. First, there is the problem of the time span of the ultimate impact of policies. For many social programs the outcome can only be expected after a long time lag, whereas for the evaluator it is easier to deal with changes that manifest themselves quickly. It should be admitted that often the best an evaluator can do, given the usual constraints and lack of knowledge, is to find out how well any short-run goals have been met. Second, there is the problem of unintended outcomes, which arise for many reasons. For example, the program may be poorly conceived. A loan program to inefficient small businessmen may only lead them deeper into debt. Or, owing to social interaction, contagion effects appear. People who have never attended a health education program learn the new ideas of behavior through contact with participants. Third, the evaluator must take care to distinguish a program's impact from those of other forces working in a situation. Exogenous factors include both systematic and random influences outside the control of the executing agency. Examples of the former include the diagnosis of patients admitted to hospital (the hospital's case-mix), the social background of pupils in a school, and the unit costs of inputs used by the organization. We will return to the third problem and the ways of allowing for external factors when we review quantitative techniques used to measure efficiency' and particularly relative efficiency of different units providing the same service.

Efficiency measures or indicators have usually concentrated on estimating ratios of outputs to inputs, which allow managers comparisons with the ratio achieved in earlier years, with the planned ratio, with the ratio of others delivering the same service, and perhaps with the ratio that might be achieved by alternative policies. Ideally, these ratios can relate both to ultimate and to intermediate objectives. A common measure of efficiency is the unit cost of delivering a service: the cost of creating a new job, paying a pension, or building a mile of road (although, strictly speaking, unit cost is an

⁷ For a discussion of the progress made in developing social indicators, see Eilon (1984) and De Neufville (1975).

inverse ratio of efficiency-cost divided by output rather than the other way round). Unfortunately, such comparisons are complicated by the multiple objectives and outputs generally pursued by public sector providers. Nor is the use of several ratios particularly useful for the comparison of effectiveness,⁸ because some organizations are better than average according to certain indicators and poorer than average according to others.⁹ There is thus no clear means of deciding which organization is the more efficient, because no one unit is *dominant*, that is, superior in all measures. In addition, ratio analyses cannot capture the effect of factors that affect the performance of the organization but are not under the control of management (for example, demographic characteristics, weather, and general economic conditions). Further, a multiservice environment gives rise to the problem of trade-offs. That is, success in one ratio can only be achieved at the expense of performance in some other ratio. Without weights to indicate the relative importance of different ratios it is difficult to be definitive on relative efficiency.

These problems in employing simple ratios to compare the relative efficiency of different units in a multiproduct environment have led to the adoption of more sophisticated techniques. The next section examines critically two of the techniques most used. This section is followed by one that illustrates some attempts to employ these techniques to improve the management and performance in three broad categories of government services.

A REVIEW OF QUANTITATIVE TECHNIQUES

From the above discussion it is clear that efficiency is essentially an empirical problem and is thus amenable to quantitative analytical techniques. Although some overlap occurs, the techniques have generally fallen into two distinct categories: statistical and nonstatistical.

Regression Analysis

Of the statistical techniques, regression analysis has enjoyed the

⁸ See Eilon (1984), for a discussion on choosing ratios.

⁹ See, for example, Lewin, Morey, and Cook (1982).

widest application. It seems particularly suited to relating output to a set of inputs, as it explains the variation in one data series in terms of variation in one or more other series. In this way a production function of the following form can be specified

$$Q = a + bR + u$$

where Q is the variable to be explained, and the values of a and b (the regression coefficients) are chosen to minimize the variation in u , the unexplained variation in Q . This method generalizes to more than one resource, that is, more than one explanatory variable (multiple regression), and to nonlinear relationships. To use the method, however, involves tackling a number of problems at three distinct levels.

Specification of the Model

As indicated above, efficiency can be defined in terms of minimizing costs for a given output, or its dual, maximizing output with given costs. In terms of the regression model an important prior question must be answered: (a) are differences in output the result of differences in resource provision; or (b) are resources allocated as a result of a policy toward output. In the first case, the regression model should be specified

$$Q = a + bR + u,$$

and the implicit behavioral model implies the spending agency is trying to maximize its output with given resources; and in the second case the model is specified

$$R = a + bQ + v,$$

and the behavioral model implies the agency is trying to minimize the cost of its output. A third hybrid model is one of interaction: resources are provided to improve output, but output reflects resource provision in a simultaneous system, that is, Q and R feed back onto one another (see example in the next section).

Choice of specification is not academic, since fitting different models to the same data set can give quite different results in the most likely case of imperfect correlation between Q and R .

Estimation Problems

In specifying the production function, it is quite likely that output will differ between agencies or regions or between different time periods, not simply because of differences in R but owing to other socioeconomic variables. The production of government services takes place in an open system, and a large number of social influences might affect the extent to which resources and outputs are related.

This problem has two aspects: (a) identifying these important conditioning factors; and (b) once they have been identified, allowing for their influence.

A weakness of the regression technique in calculating coefficients a and b is that it must make the very strong assumption that the error term u is distributed in a purely random way. Exclusion of an important explanatory variable breaks this assumption and causes a and b to be misspecified. Unfortunately, in the absence of perfect knowledge, one cannot always be certain that all important explanatory variables have been included in the regression equation. Even if one is fairly sure that all conditioning factors have been included as explanatory variables, other problems must be overcome. It is possible to test for the influence of other determinants of output by including them in a multiple regression equation of the form:

$$Q = a + bR + cA + dB \dots xZ + u,$$

where variables $A \dots Z$ are other determinants of output. However, two problems are typically encountered in this approach. First, it is difficult to isolate the influence of each variable individually because they are often correlated with one another. As a result, those that matter in reality might not appear statistically significant. This is the so-called multicollinearity problem. Second, the number of potential explanatory variables may be so large that they begin to dwarf the size of the sample and lead to unreliable estimates of the coefficients. This is called the degrees-of-freedom problem.

A number of statistical techniques have been used to circumvent these difficulties, none of which is entirely satisfactory. Multicollinearity may be detected by a number of stepwise procedures. Alternatively, another approach is to use principal

components analysis, which transforms the initial large set of variables into weighted combinations of the original variables. These components, defined to be independent of each other, can be used as variables in the regression equation, thereby circumventing the multicollinearity problem. Of course, causal interpretation of the constructed components is often difficult.¹⁰

One approach to the degrees-of-freedom problem is to use the technique of stepwise regression. This will select a smaller set out of the total set of possible determinants on the basis of those determinants that provide the best statistical fit to the data. Unfortunately this criterion may not be legitimate in the presence of multicollinearity. One less statistically sophisticated approach is simply to take the total or average value of all seemingly relevant indicators and to use it as a composite summary indicator. Certain studies in the education and health fields have used this procedure. Some composite indicators assign equal weights to the individual social indicators chosen, others give them different weights according to some judgment of their relative importance.

Interpretation Problems

In comparisons of efficiency, a measure of an agency's efficiency with respect to the chosen output is then said to be the residual, $Q_i - \hat{Q}_i$, the difference between actual output (Q), and the output predicted by the regression model (\hat{Q}). In this way regression analysis only allows us to compare individual performance with the mean, not with the feasible level of efficiency. If all agencies are grossly inefficient, interpreting an agency with a positive deviation from the mean as having a relatively good performance may be misleading, interpreting it as being most efficient is certainly illegitimate.

Moreover, regression analysis typically identifies the relationship between inputs and a single output. It therefore explains how one objective is being met. However, it says nothing directly about the relationship between a wide range of objectives usually vested in the public services. It must be recognized that in a multiple-output world, inverse relationships will exist: success in one dimension can only be achieved at the expense of performance in some other area (that is,

¹⁰ Discussed in Mayston and Smith (1987).

as well as interdependency of inputs or multicollinearity, there is interdependency of outputs). Although, multiple regressions can be employed to estimate the effects of some inputs on different outputs, multiple outputs cannot be taken into account simultaneously.

Data Envelope Analysis

Given the need to accommodate multiple outputs and inputs, there have been attempts to derive practical measures of public sector efficiency that do not depend on common units of measurement, or on an a priori set of weights to reflect social objectives. One nonstatistical technique that has received much attention recently is data envelope analysis (DEA).

This technique, based on the early work of Farrell (1957),¹¹ examines output and input relationships not from the viewpoint of average relationships of regression analysis but from the perspective of best possible performance as represented by the outlying observations. Its primary use is in measuring the relative technical efficiency of different decision-making units by comparing each unit with a comparable hypothetical unit formed as a weighted average of a number of efficient units, a reference set. Based on linear programming, DEA weights inputs and outputs so as to produce a single summary measure of the relationship between inputs and output of each decisionmaking unit. Weights are calculated so as to maximize the efficiency score of each unit, and an efficiency score can be derived equal to the weighted sum of outputs or the weighted sum of inputs.

However, a number of problems have been encountered in applying the DEA technique. First, the technique is entirely mathematical: it simply calculates the ratio of specified outputs to the specified inputs. It does not test for the statistical significance of the implied relationship between outputs and inputs. By assuming that the direction of causation is Known, which is not always clear a priori, the results of the DEA could be completely irrelevant. Moreover, because of its nonstatistical nature, deviations from best practice efficiency arise solely from technical differences in efficiency with no

¹¹ For a useful survey of the technique and its development, see Barrow and Wagstaff (1989).

allowance for stochastic deviations (for example, resulting from misspecification or measurement error).

Second, the production frontier is constructed from a subset of the data (the most efficient management units), making it vulnerable to extreme observations and to measurement error. That is, by producing a relative rather than an absolute measure of technical efficiency, the results and rankings may well alter with a change in the sample. Of course, nonhomogeneity in sampling is also a problem of regression analysis but one that has led to the development of diagnostic tools. In DEA, however, if a unit concentrates on one particular output to the exclusion of others, and is the only unit to do so, it will automatically be deemed efficient. It will form part of its own unique "facet" of the efficiency frontier, at one extreme of one axis. In this way DEA will tend to overestimate efficiency among "unusual" authorities. One suggestion has been that cluster analysis be used to produce discrete groups of more homogeneous units on which to perform a DEA.

Third, it shows each management unit in the best light possible by calculating weights for outputs and inputs that produce the highest possible efficiency score. The latter occurs because DEA gives greater weight to those inputs/outputs where the management unit concerned does relatively well. However, two rather unsatisfactory results emerge. First, by evaluating anything as technically efficient if it has the best ratio of any output to any input, one can end up with an entirely inappropriate input/output ratio. As a result a careful a priori selection of the variables is essential. Second, one well-established property of the technique is that efficiency scores will tend to improve as the number of inputs or outputs included in the analysis increases, since there is more chance of finding a more favorable comparison with the rest of the sample. As a result care should be taken to avoid including more environmental factors than are relevant to the operation of the service being examined. It has been suggested therefore that it is important to carry out a sensitivity analysis to test how much the results are changed when one output or input is excluded from the analysis.¹²

Fourth, the technique restricts the functional form of the production frontier to linear subsets, which in turn implies the

¹² Demonstrated in Mayston and Smith (1987).

assumption of constant returns to scale. If the linearity assumption is known to be unrealistic, then the factors should be transformed so that the relationship between output and environmental factors becomes linear. Normally, one views the production function as convex to the origin, so that up to a point increasing inputs increases output more than proportionately. After a certain point further inputs lead to diminishing returns in terms of output. By assuming constant returns, DEA can produce extremely misleading results.

Recognizing the importance of these restrictions, a number of attempts have been made to circumvent some of the DEA's main deficiencies. Perhaps the most important is to loosen the very restrictive assumption of constant returns to scale. In a subsequent paper Farrell and Fieldhouse (1962) were able to extend this approach to nonconstant returns to scale technologies.¹³ The deterministic nature of bestpractice production frontier has been relaxed by the introduction of a disturbance term in its specification by Aigner and Chu (1968). However, without statistical assumptions on the behavior of this disturbance term it is not possible to conduct statistical tests. Thus the Aigner and Chu model has been extended by making some statistical assumptions about the disturbance term. It has been assumed that observations on the disturbance term are independently and identically distributed and independent of the explanatory variables used. This assumption allows the production frontier to be estimated by either corrected least squares or maximum likelihood techniques.¹⁴ With the introduction of a stochastic dimension to the determination of the production frontier, it is no longer possible to attribute all variations in performance from the frontier to inefficiency. In contrast, the stochastic frontier of Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977) evades this problem by decomposing the error term into two parts. One symmetric component, v , is intended to capture random effects from factors outside the control of the individual decision unit. The second, u , is intended to capture the effects of inefficiency relative to the stochastic frontier. Unfortunately, this approach does not enable us to obtain measures of technical efficiency by observation, but it can provide measures of average

¹³ See also Grosskopf (1986).

¹⁴ See Richmond (1974) and Schmidt (1976).

Table 1. Comparison of Regression and Data Envelope Analysis

Technique	Limitations	Solutions
Regression Analysis		
<p>A statistical technique to relate average expected output to given inputs. The expected output can then be compared with actual output to define efficiency scores.</p>	<ol style="list-style-type: none"> 1. Efficiency measured in terms of one output at a time in a multiproduct environment. 2. Interaction between inputs and outputs hides causal process. 3. Identification of all relevant determining variables is difficult. 4. Determining variables may be highly correlated (multicollinearity). 5. Determining variables may be too numerous compared with sample size, causing a cdegreesof-freedom problem. 6. Defines efficiency in terms of deviations from average performance not best-practice performance. 	<ol style="list-style-type: none"> 1. Employ more than one regression, but no real solution. 2. Simultaneous equation estimation techniques. 3. Stepwise regression. 4. Reduce number of variables by performing principal components analysis, and use the components as variables in regressions. 5. Use stepwise regression on principal components to identify a smaller set of determinants. 6. No solution.
Data Envelope Analysis		
<p>Mathematical programming used to identify best possible performance for comparisons with actual performance. Can accommodate multiple outputs as well as inputs.</p>	<ol style="list-style-type: none"> 1. Based on selected output to input ratios, does not test for statistical significance of the relationship. 2. Vulnerable to extreme observations and measurement error. 3. Shows each resource-using unit in best light, regardless of the appropriateness of the ratio employed. 4. Efficiency scores improve as number of inputs or outputs included increases. 5. Restricts production relationships to linear subsets, i.e., assumes constant returns to scale. 	<ol style="list-style-type: none"> 1. Deterministic nature of production frontier has been relaxed by introduction of a disturbanceterm. 2. Use a priori selection or cluster analysis to ensure more homogeneous groups on which to use the technique. 3. More careful a priori selection of relevant variables. 4. Perform sensitivity analysis by dropping one output or input at a time. 5. Techniques have been developed to allow for different assumptions, in particular, increasing returns to scale, but results increasingly difficult to interpret.

inefficiency over the sample. In this way the model moves closer to the regression analysis approach. While on policy grounds there are clear reasons for preferring frontier models to regression models that predict average expected performance, doubt still remains about their usefulness in practice, particularly regarding the plausibility and robustness of their results.

However, the DEA approach seems to have clear advantages. Unlike regression analysis it has the ability to handle multiple outputs. Also, an important by-product of the technique is its diagnostic ability whereby it can be used to pinpoint particular sources of poor performance in inefficient organizations-referred to in the jargon of DEA as an analysis of slack variables. Certainly, in the growing number of published applications of the technique, it clearly displays enormous potential in measuring public sector efficiency. This is particularly true in areas where there exist a large number of agencies to compare.

A comparison of these two techniques and their most important limitations is summarized in Table 1.

APPLICATIONS OF QUANTITATIVE TECHNIQUES

Health

Attempts to apply quantitative techniques in this area of public services have been constrained by the measurement problems discussed above. In the past, analysis has been undertaken on performance indicators that have largely reflected activities or processes, such as patients treated, hospital throughput, patient turnover, and detailed costs per treatment. These indicators have tended to describe inputs to the system rather than the outcome. Moreover, they have often proved controversial as indicators of relative performance. For example, the "length of stay" has often been used as an indicator of efficiency. However, variations in the average length of stay among hospitals may not reflect relative efficiency but reflect differences in patient mix, the severity of their conditions, and the availability of nonhospital care opportunities. Indeed, a hospital that killed its patients off faster than others would gain a higher rating by this indicator.

On the other hand, analysis using indicators of outcomes,

involving the impact on avoidable deaths from conditions amenable to treatment, is not so easy to find in the literature. Such performance indicators tend to be of three types: (1) population-based indicators, such as standardized mortality rates, where many influences other than health services have an impact; (2) patient-based indicators, such as death rates for those who undertook a particular treatment; and (3) controlled trials where the effectiveness or Output of a treatment is assessed by providing it for some patients but not for others in matched samples.¹⁵

The difficulty in correcting for other determining influences, or in generalizing the findings of clinical trials, has meant that other more imaginative avenues have been explored. One approach, attempted by Rosser (1983), is to construct a psychometric indicator by defining ill health according to two dimensions: an eight-point scale of physical incapacity/immobility and a four-point scale of pain or distress. Each point on the two-dimensional matrix has a score derived from sampling responses, where one year of healthy life scores 100 on the QALY (quality-adjusted life year) index. The impact of a treatment for a given condition is measured by the change in the QALY that it induces. Notwithstanding Williams's proposal (1985) that the QALY approach be used to compare the effectiveness of alternative health service resource allocations, the approach is not without its critics.¹⁶

Indeed, the measurement problems in quantifying even intermediate output in this sector has severely restricted the application of the quantitative techniques reviewed above. However, Feldstein (1967) used regression analysis to estimate technical and allocative inefficiency in 177 large nonteaching hospitals in the United Kingdom. Subsequent applications of the technique were carried out by Lavers and Whynes (1978) and McGuire (1987). Unfortunately, the many measurement problems mentioned above have often cast doubts on the interpretation and reliability of these results.

¹⁵ For a review of the different approaches to output measurement in the health sector, see Drummond (1989).

¹⁶ One complaint is that this approach only takes into account the immediate impact on the individual and not on other family members. This and other problems are discussed by Culyer (1989), pp. 53ff. For a more general attempt at developing survey-based indicators of performance, see Reid (1989).

Education

Although the main objective of education is more than the development of basic skills such as literacy and numeracy, the latter skills are more amenable to quantification through objective tests. Consequently, the main quantitative indicator of education output or performance has been education results.¹⁷ With greater ease in measurement it is not surprising that education has been an area that has exhibited many more applications of quantitative techniques than the health sector. However, even here the scope of quantification has been constrained. Most educational systems generate multiple examination results. In practice no indicator is likely to be ideal and no unambiguous widely acceptable way of weighting them to construct a summary index exists. The education system is also characterized by multiple inputs with, for example, many studies differentiating between teaching and nonteaching expenditures. However, it is likely that spending on teaching staff is an inadequate measure of the quantity and quality of teaching. Indeed, it is sometimes suggested that differences in teaching input have little impact on educational outputs. Nonteaching inputs present even more of an interpretative problem. Data, which measure changes over time, or differences between regions, usually tell nothing about differences in the stock of buildings, books, and equipment. Therefore high spending might be intended to compensate for the poor stock of capital or might be a symptom of inefficiency in their use.¹⁸

Even if these basic measurement problems can be overcome, a primary problem in gauging the relative effectiveness of different educational authorities and schools lies in isolating the influence of important socioeconomic factors, such as the educational qualifications of parents. For example, in one of the leading U.K.

¹⁷ In most countries such indicators tell us nothing about cognitive performance at lower levels of ability, for which most examinations are not designed. Nor do they tell us about noncognitive achievement. However, in the United States a number of tests, such as the scholastic aptitude test and the national assessment of educational progress, now exist.

¹⁸ There have also been disputes over the proper level of analysis and whether pupil-level outcomes should be aggregated to the school level for analysis. This and other methodological problems are discussed more fully in Aitkin and Longford (1986).

studies on the effectiveness of the educational system, the Department of Education and Science (1984) employed as many as 15 socioeconomic explanatory variables. Since the number of possible background variables influencing attainment scores is large (creating a degrees-of-freedom problem), with variables typically highly correlated (the multicollinearity problem), some researchers have used principal components analysis to reduce the explanatory variables to a smaller number of independent variables.¹⁹ However, although such procedures help us apply regression analysis to judge whether a particular education authority's pupils are achieving more or less than we expect, given social circumstances and spending, there are still likely to be difficulties in interpreting the residuals. Normally, regression residuals would be viewed as indicating relative efficiency and inefficiency, after making some allowance for factors outside the authorities' control. However, they may well reflect local practices and priorities that differ from the national average.

The relative ease of quantification in this area has also led to the application of DEA. Jesson, Mayston, and Smith (1987) and Mayston and Smith (1987) report the results of two applications of DEA to the provision of secondary education in the United Kingdom. The latter study was based on four outputs, two resource inputs (teaching and nonteaching expenditure), and three environmental variables. Their results suggested that examination performance is a poor guide to efficiency, and they also illustrate the sensitivity of the results to "outliers." Bessent and Bessent (1980) also have demonstrated the versatility of DEA in calculating inefficiency scores for 167 elementary schools in the Houston Independent School District, with two output indicators, seven resource input indicators, and five socioeconomic characteristics.

Police

Attempts to measure police impact on crime have also resorted to these quantitative techniques despite obvious measurement problems. The police face many types of crime, and it is difficult to weight them to reach an aggregate index. In the provision of this

¹⁹ Levitt and Joyce (1987), Chapter 10, report the use of principal components to reduce the number of highly correlated socioeconomic variables to two main independent dimensions, subsequently used in their regressions.

service one also encounters many different inputs and input mixes in terms of different types of manpower and resources. Moreover, the definition and measurement of the crime rate is problematical, with important interactions between inputs and outputs. For example, increases in police manpower might induce the public to report more crime, and increased numbers of police on the streets may notice and record more crime, so that there is no guarantee of a negative relationship between the recorded crime rate and the number of police. In this situation analysis has generally concentrated on the number of recorded crimes cleared up as a measure of the intermediate output of policing.

The use of regression analysis in devising indicators of relative efficiency has had to cope with this interdependence between the amount of police manpower, the recorded crime rate, and the recorded clear-up rate. To do so requires a Simultaneous equation approach, such as that suggested by Carr-Hill and Stern (1979). They used a three-equation model:

A Recorded Crime Rate Equation

$CR = f(\text{clear-up rate, police numbers, various social factors}).$

The recorded crime rate is partly a function of the clear-up rate (a proxy for the probability of detection), the number of police (insofar as they detect and record more crimes, this number is likely to be positive), and various social factors like demographics and economic conditions.

The Clear-Up Rate Equation

$CUR = f(\text{crime rate, police numbers, number of social factors}).$

This equation indicates that the clear-up rate depends partly on the crime rate (an indicator of work load), the number of police (as an input), and a number of factors influencing efficiency of policing (for example, urban/rural environment).

The Police Manpower Equation

$POL = f(CR, CUR, \text{various social factors}).$

The number of police deployed is dependent on the recorded crime rate, the success in clearing up crimes, and various environmental factors felt to determine policing needs.

Owing to the simultaneous nature of these relationships, the estimation technique of three-stage least squares was employed. CarrHill and Stern found that increased police resources do not in themselves lead to lower recorded crime rates (a finding also of Willis (1983)). Also, the crime rate is negatively related to the clear-up rate (confirmed by Pyle, 1983), and police inputs are positively related to the clear-up rate (also found by Burrows and Tarling (1982)). It was also possible to rank regional police forces in terms of the magnitude of their deviations from the regression equations, as indicators of their relative efficiency,

As demonstrated by Levitt and Joyce (1987), the problem of devising indicators of relative efficiency was also amenable to DEA. Indeed, the latter technique allowed more flexibility in the inclusion of additional inputs and multiple outputs. For example, it was possible to deal with the different clear-up rates of different types of crime, to include different policing inputs, and to distinguish between those factors within management control and those not within their control. Furthermore, two sensitivity exercises were performed. The relative efficiency rankings of police forces were compared under differing assumptions of constant and diminishing returns to scale, as well as the differences in their rankings with respect to input and output efficiency.

EFFICIENCY AND BUDGETARY PLANNING

As suggested at the beginning of this paper, the performance-based approach to budgeting is a management-oriented reform, requiring spending departments to focus on what they produce and the efficiency with which they produce it. At least on the surface this reorientation seems to offer even greater scope for the quantitative approaches surveyed above. However, the above review of these techniques and their principal applications highlights some important technical limitations to their more widespread use. The first concerns the scope of these analyses, which has usually been partial in approach. These quantitative methods have been employed primarily in intrasectoral comparisons to differentiate efficient from nonefficient

producers of similar services. In this way, efficiency has been defined in a sector-specific manner. This is an obvious limitation for integrating these techniques into the budgetary process, which must allocate resources between sectors. The second limitation concerns the feasibility of these methods. All are bedeviled by the problem of finding acceptable ways of measuring the output of government activities. All face the constraint of data availability and the problem of intangible social impacts that severely circumscribe quantitative analysis. The third limitation concerns the hidden costs of adopting any quantitative methodology in a bureaucracy. One should not overlook the considerable administrative costs and extra demands on skilled manpower in terms of the increased work load in applying these techniques. This point brings us to the more general considerations about the practicality of institutionalizing this methodology.

The more widespread adoption of quantitative analysis seems to require the resolution of a number of conflicts inherent in the performance-based approach to budgeting. The first is that faced by the central budget agency in balancing the needs of decision making with its more traditional objective of holding spending departments accountable for their actions. Implicit in the new approach of allowing spending units more freedom in managing resources to meet performance objectives lies a need for decentralization of controls and decision making. While it provides more flexibility to spending departments in resource management, the central agency must also try, as a price for this new freedom, to make resource managers accountable. Unfortunately, quantitative techniques designed to help central budget agencies to assess performance are unlikely to encompass the details necessary for day-to-day agency operations. As a result they are unlikely to elicit sufficient commitment by resource managers to make them effective. Thus one has to admit a possible conflict of interest between the central agency and the spending departments. By linking performance to resource allocation, what is foreseen is that the budget will become a contract for performance and lead to the formulation of precise targets. In such an environment any request for increased resources would have to be based on an indication of achievable improved performance. Defense of current resource levels would similarly be based on achieving performance targets, with any unjustifiable shortfall in performance resulting in a cut in available resources. But

how is one to achieve a behavioral change on the part of resource managers if the new system risks damaging their agency's budget? How can the center persuade the agency to provide data on performance if it is not perceived to be in its interest? The success of any of these procedures depends on the extent to which incentives within agencies are changed so that economic efficiency becomes a high priority. Agency staff may have their own interests and policy aims, with objectives defined in terms other than efficiency. If this is so, the danger exists that any imposed quantitative procedure will be treated by bureaucrats as just one more hurdle to be overcome by dressing up their budget proposals in pseudoeconomic analyses.

Other conflicts arise from the usual trade-offs faced by economic policy. For example, there is likely to be a conflict between the short-run budget exercise, usually taking place annually, and longer-run efficiency goals. On the whole, short-run demands of economic management often take precedence over longer-run considerations such as efficiency in the provision of public services.²⁰ For example, the U.S. Civil Service Commission (1972) noted that productivity improvements had been "forced" during the budget process by reducing manpower in face of increased work load. This reduction resulted in a short-run improvement in efficiency, but in the longer term it adversely affected the quality of work and alienated the spending agencies. Also, efficiency in government operations is only one element of a government objective function, and usually a second-order one. In practice the budget process has usually been determined by other macroeconomic considerations. In recent years austerity measures or expenditure containment have dominated the budgeting process with adverse effects on efficiency. For example, budget constraints may have adversely affected capital improvements with the result that outdated equipment has become a characteristic feature of public sector organizations (Grace, 1984).

Such considerations provide a perspective on the many problems likely to be faced in trying to integrate efficiency concerns with budgetary planning. Certainly, the pursuit of efficiency as an independent goal is likely to be counterproductive. The usually complex budgetary process may be overloaded if efficiency is to be added as an invariable feature in the evaluation of all programs.

²⁰ See Diamond (1977) for further discussion.

Rather, a strategy on two fronts could be suggested. The first is the continued encouragement of quantification and improved resource allocation at the sector level. At this level, in the absence of identifiable outputs, measuring efficiency in terms of intermediary outputs is likely to be more useful. Further refinement of the techniques outlined above, and perhaps use of them in combination, or gearing them to the needs of particular sectors, is also likely to prove productive. The second is that there should be concurrent efforts at the central level to develop indicators of the impact of different programs, to form the basis for intersectoral comparisons of efficiency. In this way, there will be some check to ensure that in pursuing efficiency the purpose and activities of spending departments are not in conflict with other objectives. Inevitably, these indicators may have to be partially in qualitative terms and may be used as indicative aids rather than prime determinants in budgetary planning.

REFERENCES

- Aigner, D.J., and S.F. Chu, "On Estimating the Industry Production Function," *American Economic Review*, Vol. 58 (September 1968), pp. 826-39.
- Aigner, Dennis., C.A. Knox Lovell, and Peter Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, Vol. 6 (July 1977), pp. 21-37.
- Aitkin, M., and N. Longford, "Statistical Modelling Issues in School Effectiveness Studies," *Journal of the Royal Statistical Society, Series A*, Vol. 149, Part 1 (1986), pp. 1-43.
- Banker, R.D., A. Charnes, and W.W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelope Analysis," *Management Science*, Vol. 30, No. 9 (September 1984), pp. 1078-92.
- Barrow, Michael, and Adam Wagstaff, "Efficiency Measurement in the Public Sector: An Appraisal," *Fiscal Studies*, Vol. 10, No. 1 (February 1989), pp. 72-97.
- Baumol, William J., "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *American Economic Review*, Vol. 57 (June 1967), pp. 415-26.
- Behrman, Jere R., and Steven G. Craig, "The Distribution of Public Services: An Exploration of Local Governmental Preferences," *American Economic Review*, Vol. 77, No. 1 (March 1987), pp. 37-49.
- Bessent, A.M., and E.W. Bessent, "Determining the Comparative Efficiency of Schools Through Data Envelope Analysis," *Education Administrative Quarterly*, Vol. 16, No. 2 (1980), pp. 57-75.
- Bradford, J.W., R.A. Malt, and W.E. Oates, "The Rising Cost of Local Public Services: Some Evidence and Reflections," *National Tax Journal*, No. 2 (June 1969).
- Burkhead, J., and J.P. Ross, *Productivity in the Local Government Sector* (Lexington, Massachusetts: D.C. Heath, 1974).
- Burrows, J., and R. Tarling, "Clearing Up Crimes," Home Office Research Study, No. 73 (London, 1982).
- Carr-Hill, R.A., and N.H. Stern, *Crime, Police and Criminal Statistics* (New York: Academic Press, 1979).
- Charnes, A., W.W. Cooper, and E. Rhodes, "Measuring the Efficiency of Decision-Making Units," *European Journal of Operational Research*, Vol. 2 (1978), pp. 429-44.
- Craig, Steven G., "The Impact of Congestion on Local Public Good

Production, "Journal of Public Economics, Vol. 32, No. 3 (April 1987), pp. 331-53.

Culyer, A.J., "The Normative Economics of Health Care Finance and Provision," *Oxford Review of Economic Policy*, Vol. 5, No. 1 (Spring 1989), pp. 34-58.

De Neufville, J.I., *Social Indicators and Public Policy* (Amsterdam: Elsevier Scientific Publishing Co., 1975).

Diamond, J., "Planning Resource Transfer Between Private and Public Uses," *Policy Analysis and Information Systems*, Vol. 1, No. 1 (1977), pp. 43-58.

Drummond, Michael F., "Output Measurement for Resource Allocation Decisions in Health Care," *Oxford Review Of Economic Policy*, Vol. 5, No. 1 (Spring 1989), pp. 59-74.

Eilon, S., *The Art of Reckoning: Analysis of Performance Criteria* (London: Academic Press, 1984).

Farrell, M.J., "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Series A, Vol. 120, Part 111 (1957), pp. 253-90.

—, and M. Fieldhouse, "Estimating Efficient Production Functions Under Increasing Returns to Scale," *Journal of the Royal Statistical Society*, Series A, Vol. 125, Part 11 (1962), pp. 252-67.

Feldstein, Martin S., *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service* (Amsterdam: North-Holland, 1967).

Forsund, F.R., C.A.K. Lovell, and P. Schmidt, "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement," *Journal of Econometrics*, Vol. 13 (1980), pp. 5-25.

Grace, J.P., *War on Waste* (New York: Macmillan, 1984).

Grosskopf, S., "The Role of the Reference Technology in Measuring Productive Efficiency," *Economic Journal* Vol. 96 (June 1986), pp. 499-513.

International Monetary Fund, Seminar on Public Expenditure Management, Fiscal Affairs Department, August 29-September 8, 1988.

Jesson, J., D. Mayston, and P. Smith, "Performance Assessment in the Education Sector," *Oxford Review of Education*, Vol. 13 (1987), pp. 249-66.

Lavers, R.J., and D.K. Whynes, "A Production Function Analysis of English Maternity Hospitals," *Socioeconomic Planning Sciences*, Vol. 12 (1978), pp. 85-93.

Leibenstein, Harvey, "Allocative Efficiency Vs. 'X-Efficiency'," *American Economic Review*, Vol. 56 (June 1966), pp. 392-415.

Levitt, M.S., and M.A.S. Joyce, *The Growth and Efficiency of Public Spending* (Cambridge, England: Cambridge University Press, 1987).

Lewin, A.Y., R.C. Morey, and T.J. Cook, "Evaluating the Administrative Efficiency of Courts," *Omega*, Vol. 10 (1982), pp. 401-411.

McGuire, A., "The Measurement of Hospital Efficiency," *Social Science and Medicine*, Vol. 24 (1987), pp. 719-24.

Mayston, D., and P. Smith, "Measuring Efficiency in the Public Sector," *Omega*, Vol. 15 (1987), pp. 181-89.

Meeusen, W., and J. van den Broeck, "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review*, Vol. 18, No. 2 (1977), pp. 435-44.

Organization for Economic Cooperation and Development, Conference Proceedings, "Measuring Performance and Allocating Resources" (May 1988), pp. 1-8.

Premchand, A., "Government Budgeting and Productivity," *Public Productivity Review*, No. 41 (Spring 1987), pp. 9-19.

Pyle, D.J., *The Economics of Crime and Law Enforcement* (London: Macmillan, 1983).

Reid, Gary J., "Measuring Government Performance: The Case of Government Waste," *Tax Notes*, Vol. 42 (March 1989), pp. 29-44.

Richmond, J., "Estimating the Efficiency of Production," *International Economic Review*, Vol. 15, No. 2 (June 1974), pp. 515-21.

Rosser, R., "Issues in the Design of Health Indicators," in *Health Indicators: An International Study for the European Science Foundation*, ed. by A.J. Culyer (New York: St. Martin's Press, 1983).