8. Measuring public sector efficiency

The government has staked its political credibility on delivering significant and noticeable improvements to public services. Over the period from April 1999 to March 2006, the resources allocated to public services are set to increase substantially: NHS spending is forecast to increase by 7.3% per year in real terms, spending on transport by 6.3% and spending on education by 5.9%.[1] The extent to which these increased resources are translated into improved outcomes for users will depend on the efficiency, or productivity, of public service providers. In addition, the extent to which public sector productivity improvements materialise will have implications for the government's aim of increasing the rate of productivity growth of the UK economy.

The 2002 Spending Review set out around 130 Public Service Agreement (PSA) performance targets covering areas such as health, education and crime. The PSA targets cover both outcomes, such as NHS waiting times, and the efficiency with which those outcomes are delivered. For example, Department of Health objective 12 states that 'value for money in the NHS and personal social services will improve by at least 2% per annum, with annual improvements of 1% in both cost efficiency and service effectiveness' and Home Office objective 10 aims to 'ensure annual efficiency gains by the police of at least 2%'.[2]

These targets raise issues about how the efficiency of individual service providers is measured, and also about whether the set of efficiency measures produced can be used to give service providers incentives for improved performance. Reliable measures of both outcomes and provider efficiency are particularly important if they are to be used to target resources to more efficient providers, to act as a motivational tool or to detect failing providers.

In April 2000, HM Treasury's Public Services Productivity Panel published a report detailing a new approach to measuring the efficiency of the police.[3] The wider aim was to use the efficiency measures as part of an incentive system to

---

[1] HM Treasury, *2002 Spending Review*, Cm. 5570, London, 2002 (www.hm-treasury.gov.uk/Spending_Review/spend_sr02/spend_sr02_index.cfm?); HM Treasury, *Public Expenditure Statistical Analyses 2002–03*, Cm. 5401, London, 2002 (www.hm-treasury.gov.uk/mediastore/otherfiles/pesa_2002to2003.pdf); authors' calculations. For comparable historical figures, see table 3.4 in A. Dilnot, C. Emmerson and H. Simpson (eds), *The IFS Green Budget: January 2002*, Commentary no. 87, IFS, London, 2002 (www.ifs.org.uk/gbfiles/gb2002.shtml).

[2] 'Public Spending and Services: Links to Departmental Performance Documents', www.hm-treasury.gov.uk/Documents/Public_Spending_and_Services/publicservice_performance/pss_perf_table.cfm.

[3] C. Spottiswoode, *Improving Police Performance: A New Approach to Measuring Police Efficiency*, HM Treasury Public Services Productivity Panel, London, 2000 (www.hm-treasury.gov.uk/mediastore/otherfiles/231.pdf).

improve police performance. Police authorities were to be placed in efficiency bands, which would be used to determine future performance targets and inspection and review arrangements. It was also suggested that the proposed efficiency measurement techniques – namely, 'stochastic frontier analysis' and 'data envelopment analysis' – might have wider applicability within the public sector.[4] The Office of the Deputy Prime Minister has also commissioned research into the use of these techniques in measuring the cost-effectiveness of local authorities.[5] This chapter outlines the techniques that were proposed and assesses their suitability for measuring the efficiency of public sector organisations.

8.1 Efficiency measurement in the public sector

Measuring an organisation's efficiency is about the relationship between the outputs it produces and the inputs it uses. An efficient organisation would be one that produces the maximum possible outputs given its inputs, or one that produces a certain level of output with the minimum amount of inputs. The process of trying to measure an organisation's efficiency can therefore be broken down into three steps. First, its inputs and outputs need to be defined and measured. Secondly, it is necessary to define what is feasible – in other words, what outputs could be achieved for any given set of inputs. Finally, the organisation's actual inputs and outputs are compared with the set of feasible inputs and outputs. At this stage, one of two questions can be asked: 'Is it feasible to achieve superior outputs, given the set of inputs being used?' or 'Is it feasible to use less inputs to achieve the same outputs?'. The way the first two steps are carried out will typically be highly influential on the outcome of the third.

Definition and measurement of outputs, inputs and environmental factors

Efficiency measurement is relatively straightforward for an organisation producing one type of output with one type of input. But most organisations – public and private – produce a wide range of outputs and use numerous inputs. In the case of a private firm selling its output in a competitive market, different outputs can be aggregated by using the observed prices.[6] But public sector organisations usually produce goods that are provided either free at the point of use or at a price that is not determined by market forces. This makes it

---

[4] HM Treasury, 'Improving Police Performance: A New Approach to Measuring Police Efficiency', Press Release PSP4, 17 April 2000.

[5] Office of the Deputy Prime Minister, *Local and Regional Government Research Programme Newsletter 2002–03* (www.local.dtlr.gov.uk/research/02.htm).

[6] Under certain assumptions, prices will reflect market or buyers' valuations of the outputs. They therefore act as a natural set of weights that can be used to aggregate outputs into a single measure. See, for example, W. E. Diewert, 'Fisher ideal output, input and productivity indexes revisited', *Journal of Productivity Analysis*, 1992, vol. 3, no. 3, pp. 211–48.

very difficult to define the aggregate output of a public service provider such as a school, hospital or police force. Inputs, such as hospital beds, are usually easier to deal with, as prices are more often observed.

Another difficulty arises because some inputs are not under the control of an organisation. These can include environmental variables such as the characteristics of individuals using the service – for example, their underlying health – that may have an impact on measured efficiency. It is therefore necessary to find a way to take account of their effects when comparing organisations.

Defining what is feasible

With an appropriate set of inputs and outputs at hand, the next task is to define the efficient set of inputs and outputs against which an organisation can be compared. This is, however, unknown. The procedure therefore is to compare an organisation with an ideal comparator constructed from information on other organisations operating in the same field (and with similar size and environmental factors etc.).[7] In practice, this is often difficult, as similar organisations may be few and far between, especially in the public sector. Environmental differences are particularly difficult to control for in this respect. For example, schools will differ in the average ability of children in their catchment area and ambulance service response times will vary according to population density. If these environmental factors are taken into account when selecting the comparison group, then the group will often be rather small or even empty. If they are not, then any efficiency measurements that emerge may simply reflect differences in these environmental factors.

Measuring efficiency

Setting the difficulties of finding comparable organisations aside, two particular technical methods were put forward in the Public Services Productivity Panel report as useful ways of deriving efficiency measures for a group of 'similar' organisations.

The two methods that have been proposed are stochastic frontier analysis (SFA) and data envelopment analysis (DEA). It is important to ascertain whether these techniques are really robust and general enough to achieve the purposes for which it has been suggested they are used.[8] To look at this, we begin by describing the two techniques. Consider Figure 8.1. The dots represent observed input and output combinations for six organisations. The two sets of lines (the solid curved line and the dotted line made up of lots of straight segments) are 'frontiers' and show the maximum output that could be

---

[7] Of course, given the information available, this comparator may in itself be inefficient compared with what is possible.

[8] For a discussion of these issues with particular reference to the Spottiswoode Report, see M. Stone, 'How not to measure the efficiency of public services (and how one might)', *Journal of the Royal Statistical Society A*, 2002, vol. 165, issue 3, pp. 405–34, and the discussion in the same volume.

produced for each level of input. The two frontiers represent the outcomes of the two methods of measuring efficiency, SFA and DEA.

Figure 8.1. SFA and DEA
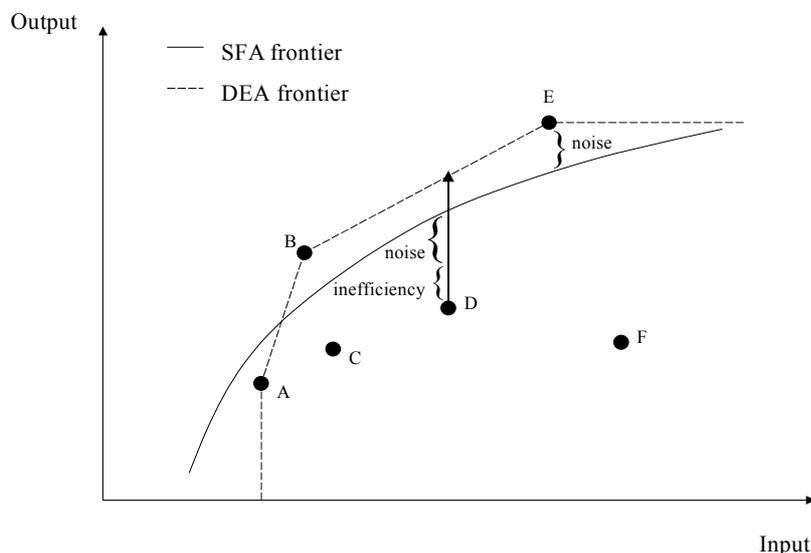
Stochastic frontier analysis

Stochastic frontier analysis uses statistical methods to fit a frontier like the solid curve in Figure 8.1. The idea is to identify the relationship between output and input(s) whilst allowing for two types of deviation from this relationship. One is statistical 'noise' – in other words, random variations in the data caused by inaccuracy in the measurement of output and by other errors. This first type of deviation is assumed to be zero on average, so that, on average, output is measured accurately.[9] The second type of deviation is a measure of inefficiency. It is one-sided: if a firm were fully efficient, it would be zero, and the more inefficient the organisation is, the more negative the deviation. These two types of deviation from the efficient frontier are shown in the figure by the curly brackets for organisations D and E. In this case, organisations B and E are classed as efficient as they lie above the frontier and organisations A, C, D and F are inefficient to some degree. The extent to which an organisation's total deviation from the frontier is designated to be noise versus inefficiency depends on the choices made about the joint distributions of the two components.

Data envelopment analysis

Data envelopment analysis is a non-statistical approach to the problem of efficiency measurement. Put simply, it takes data on organisations' outputs and inputs, and measures the efficiency of a particular organisation by its distance from the 'outer envelope' of the data. This outer envelope is shown in Figure 8.1 by the dashed line for the case where there are assumed to be

---

[9] As long as this noise is not correlated with either the inputs or the second type of deviation, it is of no particular interest.

variable returns to scale. With this technique, all deviations from the efficient frontier are classed entirely as inefficiency. In the figure, the solid arrow represents a measure of organisation D's inefficiency. Organisations A, B and E are measured as efficient and organisations C, D, and F as inefficient. It is worth noting that this procedure (and this variable-returns-to-scale version of it in particular) can designate an organisation as completely efficient simply because it produces more of a particular output than other organisations. In this single-output example, the organisation that produces the most will find itself on the efficient frontier simply because there is no larger organisation with which to compare it.

Measurement error and the form of the frontier

Stochastic frontier analysis requires a great deal of knowledge, both about the shape of the frontier and the distributions of the two types of deviation, if it is to yield a useful answer. The choice that is made about the shape of the frontier and the distributions of the deviation components can have significant effects on the efficiency rankings and absolute efficiency measures generated. These choices can be pretty much arbitrary: economic theory often provides little information about the shape of the frontier, and the data can be uninformative about the distributions of the two types of deviation. Failing to measure the inputs accurately can further complicate the task of correctly identifying organisations' relative efficiency.[10]

Data envelopment analysis does not require any assumptions about the shape of the frontier or about statistical distributions. But as the whole approach, by definition, focuses on extreme observations, it is very sensitive to mismeasurement. For example, in Figure 8.1, if the output of organisation E were inaccurately recorded and overstated, inclusion of E in the frontier would mean that the frontier was mismeasured and that the inefficiency of organisations such as F would be overstated. Furthermore, actual organisations are generally compared with hypothetical organisations (for example, the point on the DEA frontier with which organisation D is compared is a hypothetical organisation), which can mean that results are fragile if there are few real observations from which to determine the potential performance of these hypothetical organisations.

How many inputs and outputs to include

Both approaches will give results that depend upon the choice of inputs and outputs considered. An important issue is how many to include. As the number of inputs and outputs measured increases, the task of measuring efficiency becomes rapidly more difficult.

In the SFA approach, one problem is trying to choose appropriate forms for the frontier where the data are very sparse and are barely informative about the appropriate choice. Another is specifying how the many noise and inefficiency components are distributed.

---

[10] More generally, the problem is one of measurement error in the explanatory variables. See P. W. Bauer, 'Recent developments in the econometric estimation of frontiers', *Journal of Econometrics*, 1990, vol. 46, pp. 39–56.

In the DEA context, as the number of outputs increases, it turns out that the measured efficiency of the organisations cannot go down and, in the extreme, all organisations will end up being regarded as 100% efficient.[11] The intuition for this is quite straightforward: if you specify outputs very finely, you will eventually end up defining one output unique to each organisation, which – not surprisingly – it is uniquely efficient at producing. Choosing the level of aggregation of inputs and outputs is therefore an essential aspect of the exercise. There is no really good way of knowing how far or how little to disaggregate, yet the results will be heavily influenced by this choice. In order for these methods to yield useful results, the number of dimensions must be kept fairly moderate, and rules of thumb abound – for example, 'the number of organisations must be more than three times the combined number of inputs and outputs'. Decisions must therefore be made about which inputs and outputs to ignore, which to include and which to aggregate.

SFA and DEA in practice

Applications of these techniques have illustrated the sensitivity of results to the methods used to measure efficiency. A relatively recent application of statistical methods in measuring public sector efficiency can be seen in *The World Health Report 2000 – Health Systems: Improving Performance*,[12] which compares the efficiency of different countries' health systems. A scientific peer-group report on the methods employed in the paper was commissioned.[13] The report raised a number of the general issues discussed above, as well as a large number of reservations specific to the World Health Organisation's application of the methods. In particular, the group questioned the publication of league tables (page 123) based on the methods and data used. Another application is work comparing the efficiency of hospital trusts,[14] in which the authors highlight the sensitivity of the results to the efficiency measurement techniques used.

8.2 Conclusions

The obvious practical question arises of how concentrating on a subset of performance or input indicators might affect organisations' incentives. Clearly, formulaic application of these methods could have implications for providers' resource allocation, leading to the underprovision of outputs that are not included in the efficiency measurement exercise. Similarly, if certain organisations find it relatively expensive to produce certain outputs, they may

---

[11] T. Nunamaker, 'Using data envelopment analysis to measure the efficiency of non-profit organizations: a critical evaluation – reply', *Managerial and Decision Economics*, 1985, vol. 6, pp. 50–8.

[12] World Health Organisation, *The World Health Report 2000 – Health Systems: Improving Performance*, Geneva, 2000 (www.who.int/health-systems-performance/whr2000.htm).

[13] 'World Health Report 2000 Consultation and Peer Review', www.who.int/health-systems-performance/consultation.htm.

[14] D. Dawson, R. Jacobs and A. Street, 'Comparing the Efficiency of NHS Hospital Trusts', www.niesr.ac.uk/event/jacobs.pdf.

tend to redirect resources to other outputs. This has obvious implications for universal service obligations.[15] More generally, trade-offs exist between having large numbers of targets or a few, in terms of accurate performance measurement, the costs of gathering and analysing performance information, the incentives given to public service providers, and transparency and accountability in public service provision.

The outcomes of DEA and SFA exercises, in the form of rankings or league tables of organisations or other efficiency scores, are sometimes used in a second-stage analysis that tries to uncover the reasons for these measured differences in performance. Typically, the DEA or SFA efficiency scores are regressed on factors that may affect performance – for example, environmental factors that are thought to be under the organisations' control. Of course, this presupposes that the efficiency scores are correct and, for the reasons discussed above and others specific to any particular application, they may not be. Ultimately, very many auxiliary technical judgements are required when implementing either SFA or DEA techniques. While the methods are potentially useful, the results need to be treated with caution, as should their application as a motivational tool.

Ian Crawford, Alexander Klemm and Helen Simpson

---

[15] See A. Chesher, 'Discussion on the paper by Stone', *Journal of the Royal Statistical Society A*, 2002, vol. 165, issue 3, pp. 423–4.