New South Wales
TREASURY

GUIDE TO ECONOMIC PERFORMANCE MEASUREMENT FOR GENERAL GOVERNMENT SECTOR AGENCIES

Office of Financial Management

Policy &
Guidelines Paper

TPP 01-03

August 2001

PREFACE

Analysing the performance of general government agencies benefits from the support of rigorous quantitative techniques.

The *Guide to Economic Performance Measurement for General Government Sector Agencies* has been written to complement OFM's existing financial and operational performance measurement techniques for the sector.

Drawing on the major conceptual and practical contributions of production economics, the Guide introduces the major methods for measuring productivity and efficiency in a non-technical way, and sets out the basic steps for their use. Tracking the productivity of a single service delivery unit over time or benchmarking the relative efficiency of a group of units at a point in time can provide a useful source of management information about performance.

This paper is directed at both NSW Treasury analysts and agency personnel accountable for the delivery of public services in a cost-effective manner. It should stimulate thinking about how more detailed and rigorous analysis of efficiency and productivity can assist in improving the delivery of public services to the community.

John Pierce
Secretary
NSW Treasury
August 2001

Treasury Ref: **TPP 01-03**
ISBN: **0 7313 3123 0**

General inquiries concerning this document should be initially directed to:
Mr Richard Cox, Director, Fiscal Strategy (tel: 02 9228 3469, or e-mail: richard.cox@mail.treasury.nsw.gov.au) of NSW Treasury. This publication can be accessed from the Treasury's (Office of Financial Management) web site [http://www.treasury.nsw.gov.au/]. For printed copies contact the Publications Officer on 02 9228 4426.

Document prepared by Andrew Hughes.

TABLE OF CONTENTS

TABLES and FIGURES

1. INTRODUCTION

The *Guide to Economic Performance Measurement for General Government Sector Agencies* has been written to complement OFM's existing financial and operational performance measurement techniques for the NSW general government sector.

The Guide's objectives are to provide users with an introduction to the basics of productivity and efficiency measurement, and an appreciation of the major practical issues that need to be considered in their application to the general government sector.[1]

The underlying rationale for economic measures of performance stems from the major structural characteristics of most general government activity. In particular, the monopoly position that many agencies hold and the absence of any threat of takeover mean there are relatively weak organisational incentives to improve productivity and cost efficiency.

This situation contrasts with the typical position of private sector firms, where the discipline from product market competition and capital market scrutiny provide incentives for ongoing cost efficiency. Given the absence of market pressures for public sector entities, economic performance measures can be used to provide a set of surrogate incentives to spur performance improvement. The techniques canvassed in this paper are summarised in Table 1.

Section 2 of this paper outlines the characteristics of the general government sector and summarises the performance management framework for general government agencies in New South Wales. The focus then shifts from the sector as a whole to the service delivery level.

Section 3 illustrates the basic concepts and practical issues of economic performance measurement using a simplified example of a hospital that produces a single output using a single input. Section 4 provides an overview of techniques for measuring productivity and efficiency using a more sophisticated hospital example.

In Section 5, a health care case study prepared for the Steering Committee for the Review of Commonwealth/State Service Provision is presented to give the reader an overview of some issues that arise from applying these techniques. Section 6 provides the reader with a list of recommended publications that provide a more detailed and advanced treatment of efficiency and productivity measurement techniques, and details on selected software programs for performing the calculations.

[1] The techniques outlined in this paper can be applied to assessing the performance of private and public commercial entities; for example, see the Independent Pricing and Regulatory Tribunal of New South Wales' Research Paper, *Benchmarking the Efficiency of Australian Gas Distributors* (Gas 99-9) – Internet: http://www.ipart.nsw.gov.au.

Table 1: Summary of Techniques

Type of analysis	**Index number techniques** *Partial factor productivity (PFP) and total factor productivity (TFP) indexes*	**Statistical techniques** *Ordinary least squares (OLS) and stochastic frontier analysis (SFA)*	**Mathematical programming** *Data envelopment analysis (DEA)*
Measurement of *productivity change* through time ***Example*** Measuring TFP growth/decline for a single entity for a period of two or more years	***PFP index*** • Uses time series data - minimum of 4 data points. • Single inputs/outputs - does not require price data (as weights). ***TFP unilateral index (eg, Laspeyres)*** • Uses time series data. • Requires price (or cost/ revenue) data to weight changes of multiple inputs/ outputs. • Sources of TFP change cannot be identified. • Both index types assume no measurement error.	***OLS*** • Can be applied to measuring productivity change – not covered in Guide. ***SFA combined with Malmquist index*** • Typically uses panel data. • TFP change can be decomposed into changes in technical efficiency, scale and technology. • Allows for measurement error.	***DEA combined with Malmquist index*** • Typically, uses panel data. • TFP change can be decomposed into changes in technical efficiency, scale and technology. • Assumes no measurement error.
Measurement of relative *technical efficiency* levels at a point in time ***Example*** Benchmarking the technical efficiency of a group of service delivery units of an entity for a given year	***PFP index*** • Uses cross-sectional data. • Involves simple comparison of PFP ratios across entities. ***TFP multilateral index*** • Uses panel data. • Entities compared with a hypothetical, average entity in sample.	***OLS*** • Uses cross-sectional data. • Entities compared with average industry/sector performance. • Assumes no measurement error – residual is attributed to inefficiency. ***SFA*** • Uses cross-sectional data. • Comparison against best performing entity. • Residual decomposed into random error (meas. error) and inefficiency parts.	***DEA*** • Uses cross-sectional data. • Entities compared with best performers in sample.

2. MEASURING AND EVALUATING NEW SOUTH WALES' GENERAL GOVERNMENT SECTOR SERVICE DELIVERY

Summary

- The *Guide to Economic Performance Measurement for General Government Sector Agencies* provides an introduction to the application of productivity and efficiency techniques to general government agencies.

- A major difficulty faced by *any* type of performance measurement is defining robust measures of outputs and inputs. This problem is particularly acute for general government agencies due to the unpriced and intangible nature of their outputs. Unfortunately, there are no performance measurement techniques available that bypass this problem.

- Despite these limitations, performance measures can play an important role in improving accountability, encouraging ongoing performance improvement and encouraging more efficient service provision.

- Performance measures will be more useful where the sources of performance variation across organisations or operational units can be accounted for. In particular, when making judgements about performance and accountability for performance, it is important to distinguish the contribution of those factors that can be controlled by an organisation's management from those that cannot (eg, population density).

2.1 Background

The diverse range of public services provided by the NSW general government sector represent a significant share of the State's resources; net underlying general government expenses account for about 10 per cent of NSW gross state product.

The performance of the sector in converting resources of labour, capital (fixed assets) and materials into services has an important impact on both the overall State economy and community welfare. Low productivity in the delivery of publicly provided services implies higher levels of taxation and/or public debt. This, in turn, leads to distortions in the investment and consumption decisions of firms and households.

The NSW Government is pursuing initiatives to help ensure that the State's general government agencies produce services efficiently and contribute to meeting broader social policy objectives. The central financial policy initiative is the *Financial Management Framework for the General Government Sector*. The Framework is a consolidation of, and advancement upon, recent financial management reforms. It is intended to facilitate further improvements in program and service delivery.

The Framework will encompass a suite of interrelated policies and guidelines and agency toolkits. Each component of the Framework, however, will be capable of implementation in its own right. The intention is to roll out the components progressively over the next few years.

The measurement and evaluation of agency performance is a key component of these initiatives, in particular:

- *Budget Outcomes and Outputs – A Guide to Performance Management*, which is to be released in time for the 2002-03 Budget process; and
- *Performance Analysis Guidelines for General Government Sector Agencies*, which is intended to assist analysis of the financial and non-financial performance of general government agencies, including their performance against annual Service and Resource Allocation Agreement targets. The document aims to provide a robust methodology for evaluating the efficiency of agency programs and services, and their effectiveness in contributing to broader government policy objectives.

The *Guide to Economic Performance Measurement for General Government Sector Agencies* complements OFM's existing performance measurement initiatives by drawing on the major conceptual and practical contributions of production economics. The Guide's objectives are to provide users with an introduction to the basics of productivity and efficiency measurement, and an appreciation of the major practical issues that need to be considered in their application.

2.2 Characteristics of General Government Sector Service Delivery

Public sector entities vary in the nature of their operations and the source of their funding. The international *Government Finance Statistics* reporting system defines the general government sector as consisting of those public sector entities that provide goods and services outside the market mechanism, and facilitate the transfer of income for public policy purposes. At the State level, general government agencies are funded primarily from taxation and/or Commonwealth grants.

General government service delivery is characterised by a number of factors that present challenges to the objective of achieving value for money. These include:

- complexities of the political process;
- tension between short term imperatives and the design of long term program and service delivery strategies;
- uncertainty about the most appropriate delivery strategies;
- unequal information among stakeholders pertaining to community needs, as well as program and service possibilities;
- different incentives for different stakeholders;
- absence of price signals to guide decisions on service provision and consumption; and
- lack of competition in service provision.

2.3 Analysing NSW General Government Agency Performance

The analysis of NSW general government agencies is focused on three broad areas:

- budget compliance – ensuring that agencies keep within annual spending limits;
- financial performance – assessing the financial health of an agency over time; and
- value for money – evaluating the efficiency and effectiveness of government service delivery.

The Service and Resource Allocation Agreements (SRAAs) provide the framework for analysis of value for money by general government agencies. The purposes of a SRAA are:

- to improve Treasury's relationship with general government agencies by developing a concise understanding of an agency's operations;
- to provide basic information to help reallocate resources to areas of greater value;
- to facilitate a more strategic approach to the allocation of resources across policy areas; and
- to link strategic planning (at both the agency and whole of government levels) to the Budget process.

Performance monitoring, using the SRAA framework, is principally concerned with evaluating effectiveness, efficiency and economy. Effectiveness is the extent to which an agency's programs and services (outputs) achieve the Government's desired outcomes. Efficiency is defined as the extent to which an agency maximises the outputs produced from a given set of inputs or minimises the input cost of producing a given set of outputs. Economy refers to buying inputs in the most economic manner (ie, obtaining appropriate quality resources at least cost).

The content of this Guide draws upon the contribution of production economics to provide an introduction to the basics of productivity and efficiency measurement. Changes in the efficiency of a single entity can be considered over a period of time (time series analysis), and this is what some performance measurement techniques are designed to do. Alternatively, the efficiency of an entity can be compared at a point in time with other similar entities (cross-sectional analysis), for which other performance measurement techniques are required.

One of the major difficulties faced by any performance measurement exercise is defining robust measures of outputs and inputs. This problem is particularly acute for entities providing unpriced services outside the market mechanism, which by definition is precisely the circumstance where such performance measures are most needed. And it is, of course, the situation for many government service providers.

Unfortunately there are no performance measurement techniques available that bypass this problem. In general, good performance measurement requires a clear definition of outputs and inputs, and data pertaining to those inputs and outputs.

However, some performance measurement techniques do provide scope for "testing" alternative input and output definitions in the following sense. Performance measures for different combinations of inputs and outputs can be derived, allowing the robustness of the measures to changes in definitions to be assessed. Ideally, the performance measures will not be very sensitive to changes in the choice of outputs and input measures.

Performance measures will be more useful where the sources of performance variation across organisations or operational units can be accounted for. In particular, when making judgements about performance and accountability for performance, it is important to distinguish the contribution of those factors that can be controlled by an organisation's management from those that cannot. The latter are often characterised as an organisation's particular **operating environment**, and include factors such as demographics, climate, and population density. Some of the techniques discussed below are capable of making such allowances in a consistent and robust way.

2.4 Analysing Performance – Application to a Hospital

2.4.1 Introduction

As a device for explaining economic performance measurement techniques in this Guide we will use the example of a hospital. We will progressively develop this example to draw out the key considerations.

Four broad categories of hospital output can be identified:

- inpatient treatment – the curative care of patients who are admitted to stay in hospital, where treatment is for a specified condition of an acute, short-term nature;
- outpatient visits – care of patients without admission;
- teaching – the provision of doctor and/or nurse education; and
- research – systematic inquiry aimed at expanding the stock of medical knowledge.

The core activity of a hospital is the provision of inpatient treatment. The reason for this is that the other categories of output can each be produced in institutions other than hospitals. For example, outpatient services can be provided in an outpatient clinic or a doctor's surgery.

This is not to say, of course, that the production of two or more of these output classes in a hospital do not share common costs. However, for an institution to be classified as a hospital, it must provide inpatient treatment.

The remainder of Section 2 considers the definition and quantification of outputs and inputs in the case of a hospital, drawing out some more general observations.

2.4.2 Hospital outputs

Any approach to efficiency analysis of general government agencies needs to conceptualise and measure outputs. As many of the outputs of the general government sector are intangible, like many service industries, this is rarely straightforward.[2] Moreover, improvements in the quality of public services make it difficult to measure changes in output over time.

An important **conceptual** issue when considering the performance of hospitals in treating inpatients is whether the output of a hospital should be defined as the provision of the medical treatment, or the resulting improvement in the patient's health status (if any).

To some readers this may appear to be an example of the general issue of whether the focus should be on outputs - ie, *treated patients*, or outcomes - ie, *patients with an improved health status*. However, some may contend that hospitals should be viewed as producing health as an output, or more correctly, changes in health status. The counter argument is that the treatments hospitals produce are only one influence on a patient's health status, and may not change that status due to factors outside the hospital's control.

The health status concept is not practical for two reasons. First, the measurement of changes in health status is difficult. Second, even if changes in health status could be measured, the marginal contribution of hospital services from the contribution of other factors which influence the outcome (eg, exercise and diet) would be difficult to isolate. In terms of application then, the treatment conception has a distinct advantage. Appendix A, "Economic Measures of Performance – Should they Relate to *Outputs* or *Outcomes?*" provides a more detailed discussion of these issues.

For the purposes of **measurement**, a connection must be established between the abstract concept of a treatment and an empirically workable unit of output measurement.

An episode of hospitalisation (or a treated case) could, in principle, be taken as an empirically workable counterpart of the treatment conception. A possible alternative unit of output – the patient day – could also be considered but, in the context of acute inpatient hospital care, is more of an input related measure. In particular, it is related to the time dimension over which a unit of output (a treated case) is produced. Inter-hospital comparisons of average cost per patient day can be quite misleading because of the confounding influence of variations in the average length of stay.

The issue of **classifying** hospital output arises because of the multiproduct nature of inpatient treatment provided by a hospital. Not all treated cases are the same since the treatment produced will depend on, among other things, the patient's diagnosis.

---

[2] For public services, government statistics typically assume that output simply rises in line with the volume of inputs (eg, number of hours worked). Thus, by definition, productivity never rises.

The term "case mix" has evolved to describe the mix of cases treated by a hospital classified according to a set of criteria aimed at achieving "like" units of outputs within case mix categories. There are two broad approaches to the measurement of case mix. The first of these adopts a disaggregated approach that aims to minimise the within-group heterogeneity of treatments provided. The International Classification of Diseases (ICD) and Diagnosis-Related Groups (DRGs) are the two major case-mix classification schemes. The DRG approach, while employing a case mix classification scheme, attempts to summarise the information in an index measure.

The DRG classification system has been criticised on the grounds of variation within individual categories, a problem that worsens the higher the level of aggregation. But a higher level of aggregation is necessary if the number of output categories is to be kept at a manageable level.

The more general point raised by the foregoing discussion is that the definition of outputs for any general government activity needs to take account of both theoretical and practical considerations. Higher level outputs, or outcomes, are more likely to be influenced by external uncontrollable factors, and it may be stretching the point to characterise them as being "produced" by the entity/entities in question. Ultimately, if an output cannot be measured it cannot be included in an empirical investigation of performance.

2.4.3 Hospital inputs

A hospital's labour and capital[3] resources are its most important inputs. Labour inputs can be measured using a simple workforce headcount or by calculating hours worked. The former ignores variations in skill level and neglects differences in work intensity. The best measure of the labour input is probably hours worked. For permanent employees, however, available hours are often related to hours *paid* for rather than hours worked. Ideally, an hours worked measure for this labour category should exclude hours credited for recreation and other types of leave.

The measurement of capital inputs poses more difficulties. For a hospital, they comprise a diverse range of items such as hospital beds and electronic medical equipment. One approach is to aggregate the different items using some indicator of relative importance such as a rental price. The rental price "weights" must be held constant for two or more periods being compared so that changes in the aggregate capital measure reflect changes in physical volumes rather than in prices.

Calculating a rental price or unit cost for the flow of services from a capital stock poses many conceptual and measurement challenges. A unit cost needs to be imputed because fixed assets are typically owned by a hospital rather than rented or leased at observable market prices. In principle, an implicit rental price for a given fixed asset should reflect depreciation from regular "wear and tear" and obsolescence, financial "holding" costs (ie, the nominal opportunity cost of the funds invested in fixed assets), and general asset price inflation.

[3] The term "capital" is used broadly to refer to all durable assets including machinery and equipment, and buildings.

An alternative method for estimating the stock of capital inputs is to deflate the nominal value of the hospital's capital stock by an appropriate price index to eliminate the effects of price changes. [4]

In practice, deficient quantity and price data often necessitate the use of a simple proxy variable. For example, number of beds is commonly used as a proxy for a hospital's capital stock in efficiency benchmarking studies.[5] A major difficulty is adjusting a hospital's capital stock for the impact of technological change. Many types of medical equipment have experienced significant improvements in performance, resulting in rapid depreciation and obsolescence for existing assets.

Other categories of input can be separately identified or simply grouped together as an "other inputs" category. The calculation of an "other inputs" measure is illustrated in Section 4.2 below.

In general, there tends to be less ambiguity about what are inputs than outputs. The problems associated with the definition and measurement of capital inputs identified in the hospital example are commonplace, and similar approaches to those suggested for hospitals can often be adopted.

2.4.4 Sources of variation in hospital performance

There are three potential sources of performance variation for a hospital:

- differences in management performance;
- variations in service quality and funding levels: and
- differences in operating environment.

The efficiency of management and the level of service quality are both controllable factors. The particular demographic and socio-economic characteristics of its operating environment are not subject to management control. Hospitals in some locations may have greater difficulty attracting skilled staff than others, which might also be regarded as part of its "operating environment". Since environmental characteristics usually vary across service providers and therefore influence performance, they must be accounted for so that "like with like" comparisons can be made.

In sum, the issues discussed in this section are relevant for *any* type of empirical work that is performed on hospitals, or more widely to other areas of general government sector service delivery.

[4] The "deflation" approach to estimating quantities is illustrated in Section 4.2.
[5] It is easy to think of reasons why this measure may be inadequate in some applications given the plethora of machinery and equipment used in hospitals.

3. CONCEPTUAL FRAMEWORK FOR ECONOMIC PERFORMANCE MEASUREMENT

Summary

- Productivity is defined as the ratio of outputs to inputs. Productivity can be analysed at various levels – economy-wide, industry, firm/agency and operational unit. The focus of this Guide is at a more disaggregated level; how well the service delivery units of general government agencies convert inputs of labour, materials and capital into outputs of services.

- Improvements in an organisation's productivity ratio over time (or productivity growth) can be attributed to three major sources: increases in technical efficiency, exploitation of scale economies, and technological advances. These concepts are illustrated using a simplified example of a hospital that produces a single output using a single input.

- Productivity movements are correlated to changes in an organisation's cost structure. Improvements in productivity, assuming no change in input prices, will reduce costs.

3.1 Introduction

In this section, we explain the basic concepts of economic performance measurement drawing on the theoretical framework of production economics. A simplified example of a hospital is used to explain the sources of productivity change within a single entity. This is followed by a discussion of the linkages between physical production relationships and costs when price information is introduced.

3.2 Basic Concepts

The concept of **productivity** is widely accepted as a key performance benchmark for entities. Rising productivity is related to increased profitability, lower costs and sustained competitiveness.

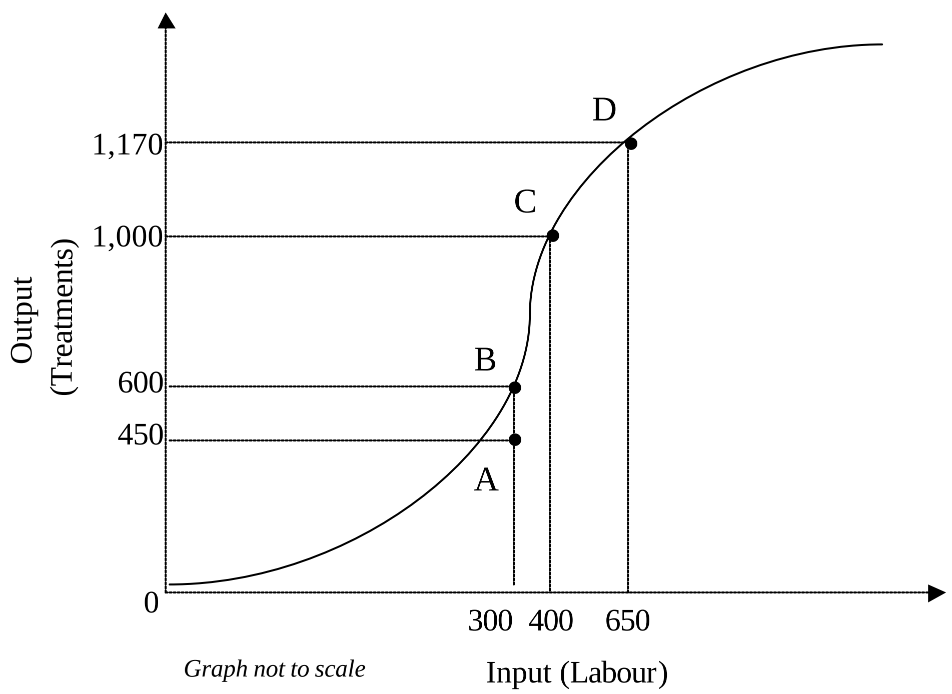
Productivity is defined as the ratio of outputs to inputs. Productivity can be analysed at various levels – economy-wide, industry, firm/agency and operational unit. The focus of this Guide is at a more disaggregated level; how well the service delivery units of general government agencies convert inputs of labour, materials and capital into outputs of services.

For a simplified example we assume that labour is the only resource required to treat patients in a hospital. Labour is measured as hours worked. The output is the number of treatments produced. If the hospital uses 500 hours of labour to produce 1,000 treatments, its productivity is 2 treatments per labour hour. If there is another hospital that uses 400 hours of labour to produce 1,000 treatments its productivity is 2.5 treatments per labour hour. Since the second hospital can produce more treatments per labour hour than the first hospital, the second one is more productive.

We can use the concept of the **production function** to understand more about productivity. [6] A production function describes the relationship between the output (number of treatments) and the input (number of labour hours). The nature of the output-input relationship depends on the particular **production technology** that is used to convert the inputs into the outputs. In this context, the term "technology" is intended to capture the skills of the labour input.

Figure 1 shows the production function for patient treatments. Points B, C and D represent the maximum number of treatments that can be produced at different input levels (or the minimum amount of labour hours required to produce a given number of treatments). The *shape* of the curve will depend on the particular production technology of the hospital.

Figure 1: Simple Production Function (single output/input)



[6] Using mathematical notation, the general form for a production function is *outputs = f(inputs)*; ie, *number of treatments = f(labour hours)*. The functional form is defined by the particular production technology.

The area from the curve to the horizontal axis comprises different possible combinations of the output and input (input-output combinations above the curve are not feasible unless a new technology is introduced). For example, at Point A it is possible to produce 450 treatments using 300 hours of labour. The productivity ratio, however, of this combination is not at the maximum possible since it is technically possible to produce more treatments using the same amount of labour hours. At Point B, the hospital can produce 600 treatments using the 300 labour hours; the given production technology does not allow more treatments to be produced using this amount of labour hours. Point B is a point of **technical (or productive) efficiency**.[7] Put differently, Point A is technically *in*efficient compared to Point B.

Points on the upper bound or "curve" of the production function (ie, B, C and D) are all technically efficient. These points represent the maximum output that can be produced for a given level of input. If the hospital were able to move from Point A to Point B through improved management practices, its productivity would increase from 1.5 treatments per labour hour to 2.0, representing a 33 per cent increase in productivity. Hence, one way for a hospital to increase its productivity is to improve its technical efficiency.

Next, we turn to the relationship between productivity and the hospital's **scale of operations**. Suppose that the hospital can readily adjust its scale or size of operations. When the hospital increases its scale by moving from B to C, its labour hours increase from 300 to 400 hours or 33 per cent, but the number of treatments produced increases from 600 to 1,000 units or 67 per cent. Since the proportionate increase in labour hours is less than the proportionate increase in treatments, the productivity ratio increases from 2.0 to 2.5. This illustrates the concept of **scale economies**. Hence, higher productivity can also result from exploiting economies of scale.

In general, increasing the employment of labour and capital resources allows managers to subdivide tasks so that inputs can be specialised and productivity improved. In a health care context, an increased scale of operations may improve inventory management. A hospital's inventory of medical supplies may not need to increase at the same rate as output growth due to probabilistic considerations.

From C to D, labour hours increase by 250 or 63 per cent but the number of treatments produced increases by only 170 units or 17 per cent. In this case, the input increases by a greater proportion than the increase in output. This results in a reduction in the productivity ratio from 2.5 to 1.8. At this size, the hospital faces **scale *dis*economies**.

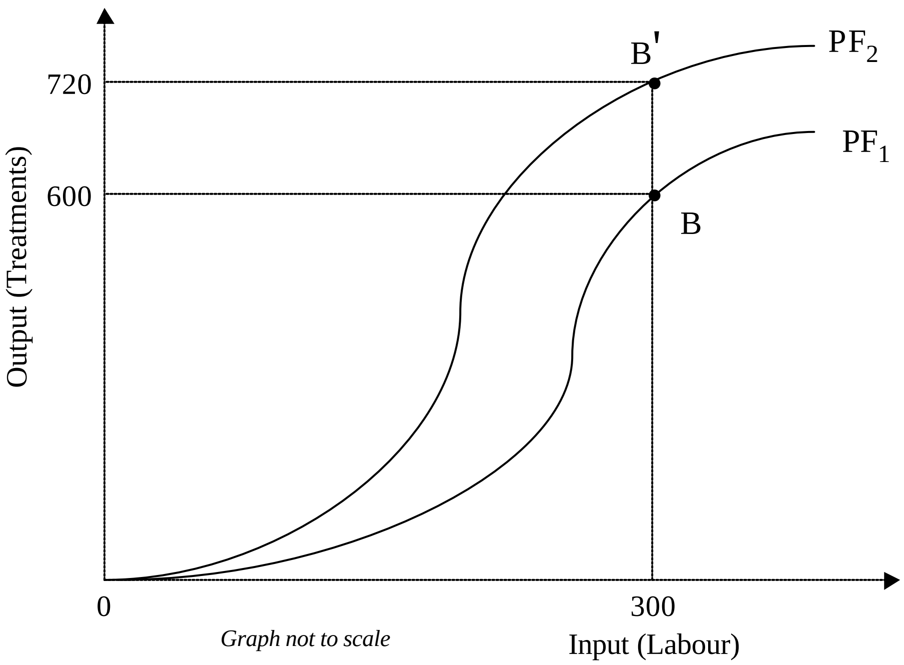
Scale diseconomies typically arise from coordination problems beyond a certain point. Restructuring a service delivery unit into two or more smaller entities is a possible solution to scale diseconomies.

[7] It is worth noting at this point that the terms "productivity" and "efficiency" are typically used interchangeably in business and government circles, which may be a source of confusion for a reader not familiar with economic techniques.

Based on our illustration in Figure 1, Point C represents the *optimal* scale for the hospital; that is, the size that gives the highest possible productivity ratio. This illustration shows that the hospital could improve its productivity if it were in a position to adjust its scale of operations until it reaches this optimum point.

Lastly, we turn to the relationship between productivity and **technological change**. A positive technological change is illustrated by an outward shift in the hospital's production function between two periods. In Figure 2, the production function from the previous figure is reproduced to represent Period 1.

Figure 2: Productivity and Technological Improvement (single output/input)



Providing staff with additional skills allows the hospital to produce more treatments using the same amount of labour at every level of production. The production function in Period 2 lies above the production function in Period 1. Point B' is on the production function in Period 2. At Point B', the hospital uses 300 hours of labour to produce 720 treatments, giving a productivity ratio of 2.4. Hence, the hospital's productivity increases from 2.0 at Point B to 2.4 at B' due to the use of advanced skills such as the implementation of "keyhole" surgery.

To summarise discussion to this point, improvements in an organisation's productivity ratio over time (or productivity growth) can be attributed to three major sources: increases in technical efficiency, exploitation of scale economies, and technological advances.

3.3 Productivity and Cost Minimisation

So far, all discussion has involved the relationship between the physical units of inputs and ouputs. We have not discussed financial concepts such as costs. If input price information is available, such as the wage rate, we can analyse the relationship between productivity and average costs.

Return to Figure 1 and assume that the price of labour is constant at $350 per hour. Consider Point B. The hospital uses 300 hours of labour. Therefore, the total cost is $105,000. Since the hospital at Point B produces 600 treatments the cost per unit (or average cost) is $175.

Similarly, at Point C the hospital uses 400 hours at $350 per hour. The total cost is $140,000 while the total output is 1,000 units. Therefore, the average cost is $140. Table 2 summarises the computation of average cost for the four illustrated points.

Table 2: Calculation of Average Cost

Point	Number of labour hours x	Labour price ($) w	Cost ($) (x*w)	Number of treatments y	Average cost ($) (x*w)/y
A	300	350	105,000	450	233
B	300	350	105,000	600	175
C	400	350	140,000	1,000	140
D	650	350	227,500	1,170	194

As the hospital increases its scale of operations from B to C, *total* cost increases but *average* cost decreases due to the impact of scale economies. As the hospital moves from C to D, however, it encounters diseconomies of scale. Total costs increase at a faster rate than the increase in output.

As a result, the average cost increases from $140 to $194. Point C represents the optimal scale for the hospital, where the productivity ratio is at a maximum and average cost is at a minimum. Hence, productivity movements are correlated to changes in an organisation’s cost structure. Improvements in productivity, assuming no change in input prices, will reduce costs.

3.4 Multiple Inputs and Outputs

When only a single input and a single output are involved in an organisation's production process the calculation of the productivity ratio is easy. For a production process that has more than one input or output, a method for combining these multiple measures into a single, aggregated quantity is needed to calculate the ratio. How this can be done will be discussed in Section 4.

When an organisation's production process uses multiple inputs the concept of technical efficiency becomes more sophisticated. It raises issues about the relationships among inputs in the production process; for example, whether they can be substituted for each other and if so at what rate.

Where inputs are substitutable, the organisation may be able to reduce its cost of production if it chooses an input mix that minimises its total costs at prevailing input prices.

When an organisation chooses an appropriate mix of inputs and/or outputs that takes into account given market prices, it attains **allocative efficiency**. The combination of allocative and technical efficiency gives a measure of **cost efficiency**. For a detailed discussion of allocative and cost efficiency concepts and applications the reader is referred to Coelli, Rao and Battese (1998).

4. OVERVIEW OF ECONOMIC PERFORMANCE MEASUREMENT TECHNIQUES

Summary

- There are three broad approaches to measuring productivity and efficiency – index numbers, statistical and mathematical programming. No single technique can provide a complete view of performance because they all require assumptions to be made that can affect the interpretation of results. Therefore, a sensible and prudent approach requires the use of several techniques.

- The measurement of productivity over time using a more sophisticated illustrative example (a hospital that produces two outputs from four inputs) is demonstrated using index numbers. The introduction of an entity with multiple outputs/inputs raises aggregation issues; in particular, how can multiple outputs/inputs be aggregated or combined into a single measure to allow for the calculation of total factor productivity?

- Aggregate output/input indexes are calculated using the Laspeyres index, a common index number formula. Price data is used to weight the contribution of individual input and output quantities. Where market prices are not available for outputs, which is typically the case for taxpayer-financed public services, it may be possible to use a proxy measure that reflects the resource intensity of the service.

- The empirical measurement of technical efficiency is framed in terms of benchmarking performance. The relative performance of a small group of hospitals at a point in time is conducted by comparing their partial factor productivity ratios. This is followed by a more general discussion of statistical techniques (ordinary least squares and stochastic frontier analysis) and a mathematical programming method (data envelopment analysis) for measuring efficiency.

4.1 Introduction

In this section we provide an overview of available economic performance techniques for measuring productivity and technical efficiency using a series of simple numerical examples to illustrate the main concepts. Three broad approaches to economic performance measurement will be discussed - index numbers, statistical, and mathematical programming. For a comprehensive treatment of these approaches the book by Coelli, Rao and Battese (1998) is recommended.

In Section 4.2, we use the index numbers approach to measure productivity change for a single hospital that produces two outputs using four inputs over a period of time. Where there are multiple inputs and outputs, the different inputs and outputs can be aggregated to allow for the calculation of measures of *total factor productivity.* We use this illustrative example to distinguish partial factor and total factor productivity measures. The ratio of individual outputs to individual inputs, or aggregated outputs to aggregated inputs provide indexes that are in common units, but may have little meaning in terms of their absolute levels.[8] Such indexes can provide for meaningful *relative* comparisons. For example, relative changes in the index for a given entity tell us by how much its productivity has changed. Alternatively, the index value of an entity can be compared with that of another entity as a guide to its relative productivity performance.

In Section 4.3, the empirical measurement of technical efficiency is framed in terms of benchmarking performance. The relative performance of a small group of hospitals at a point in time is conducted by comparing their partial factor productivity ratios. This is followed by a more general discussion of statistical techniques (ordinary least squares and stochastic frontier analysis) and a mathematical programming method (data envelopment analysis) to measuring efficiency.

4.2 Measuring Productivity

In this section, we consider a simplified model of a hospital, called XYZ. Measuring productivity performance first involves defining and then collecting data on the outputs and inputs used in the production process. Outputs are associated with a receipt (for commercial activities) while inputs are linked with a payment.

For our example, we will assume that the hospital is part of a tax-financed, public health system, and provides treatments that are free of charge to users (patients).

The hospital produces two outputs:

- *inpatient treatments* measured as number of cases; and
- *outpatient visits* measured as number of consultations.

These outputs are produced using four inputs:

- *labour* measured as number of full time equivalent medical (eg, doctors) and administration staff;
- *contractors* measured as hours worked by visiting medical officers and other contractors (eg, cleaners);
- *capital* measured by a proxy, number of beds; and
- *other inputs* measured on an imputed quantity basis.

[8] A well-known index is the consumer price index (CPI) which measures the value of a basket of commodities relative to its value at a base period. If the current value of the CPI is, say, 132, this level by itself does not tell us anything very useful. However, if we know that the CPI is based in 1990, ie. its index was set to a value of 100 for 1990, we can now interpret the current value of the CPI as being 32% above its level in 1990. That is, relative changes in the value of the index give us measures for the rate of change in consumer prices.

The measurement of the other inputs series requires some explanation. This series comprises a mixture of residual inputs that share no common physical measure; for example, pharmaceuticals and patient meals. A physical estimate can be "imputed" or measured by deflating the expenditure on these items by an appropriate price index.[9]

This can be simply explained from the simple accounting identity where:

$$\text{Cost} = \text{Price} \times \text{Quantity}$$

Rearranging the terms of the identity gives:

$$\text{Quantity} = \text{Cost/Price}$$

If a price index is used, and given the cost measure, a notional quantity measure can be defined.

The hypothetical data for XYZ Hospital is presented in Table 3.

Table 3: Output and Input Data of XYZ Hospital, Years 1 to 4

Year	Inpatient treatments (no. of cases) y_1	Outpatient visits (no. of consultn.) y_2	Labour (no. of FTE staff) x_1	Contractors (no. of hours worked) x_2	Capital (no. of beds) x_3	Other inputs (imputed quantity) x_4
1	800	45,000	1,300	250,000	300	200,000
2	820	47,500	1,250	275,000	305	210,000
3	850	50,000	1,200	300,000	310	215,000
4	875	52,000	1,175	330,000	314	230,000

A distinction is usually made between the performance of individual factor inputs (partial factor productivity) and overall productivity (total factor productivity). Both measures can be applied at any operational level where sufficient data is available.

Partial factor productivity (PFP) ratios are the most common form of productivity measurement. The PFP measure for XYZ can be calculated as the ratio of each of the outputs to each of the four inputs. For example, the labour partial factor productivity ratio using the inpatient treatment output is:

$$\text{PFP} = \frac{\text{Inpatient treatments (no. of cases)}}{\text{Labour (FTE workforce)}}$$

[9] A price index should be sector-specific where possible to improve the degree of precision for changes in the quantity estimate. Where one is not available, a broad-based price index such as the GDP deflator or consumer price index can be used.

Although the hypothetical data comprises four periods, a single period is sufficient for the calculation of a partial ratio and a minimum of two periods is required for the computation of a productivity change. XYZ Hospital's four series of PFP ratios using the inpatient treatment measure of output are presented in Table 4a:

Table 4a: Partial Factor Productivity Ratios (Inpatient Treatment Output Measure), XYZ Hospital

Year	Labour productivity ratio y_1/x_1	Contractor productivity ratio y_1/x_2	Capital productivity ratio y_1/x_3	Other inputs productivity ratio y_1/x_4
1	0.615	0.0032	2.667	0.0040
2	0.656	0.0030	2.689	0.0039
3	0.708	0.0028	2.742	0.0040
4	0.745	0.0027	2.787	0.0038

Each series of PFP ratios can be converted into an index by following three steps:

- selecting a base value for *each* PFP series, say their values in Year 1;
- dividing the base value by 100; and
- dividing every ratio in the particular series by the resulting amount.

The index series for each PFP ratio are presented in Table 4b.

Table 4b: Partial Factor Productivity Index Series (Inpatient Treatment Output Measure), XYZ Hospital, Year 1 =100.0

Year	Labour productivity ratio y_1/x_1	Contractor productivity ratio y_1/x_2	Capital productivity ratio y_1/x_3	Other inputs productivity ratio y_1/x_4
1	100.0	100.0	100.0	100.0
2	106.7	93.8	100.8	97.5
3	115.1	87.5	102.8	100.0
4	121.1	84.4	104.5	95.0

PFP percentage changes can be easily calculated from the either the ratios (Table 4a) or index numbers (Table 4b). For example, labour productivity increases by 21.1 per cent between Year 1 and Year 4: ratio calculation [0.745/0.615 – 1] or index calculation [121.1/100.0 – 1]. Annual percentage changes for the four PFP measures are presented in Table 4c.

Table 4c: Partial Factor Productivity Annual Percentage Changes (Inpatient Treatment Output Measure), XYZ Hospital

Year	Labour productivity change y_1/x_1	Contractor productivity change y_1/x_2	Capital productivity change y_1/x_3	Other inputs productivity change y_1/x_4
1	-	-	-	-
2	6.7%	-6.3%	0.8%	-2.5%
3	7.9%	-6.7%	2.0%	2.6%
4	5.2%	-3.6%	1.6%	-6.0%

From Year 1 to Year 2, the labour productivity ratio increases by 6.7 per cent. Contractor productivity declines by 6.3 per cent over the same period. Meanwhile, capital productivity increases by 0.8 per cent. The other inputs' productivity measure falls by 2.5 per cent.

Partial factor productivity measures are conceptually simple and easy to calculate. These measures, however, need to be interpreted with care as they do not provide a complete picture of performance. Notably, they can move in different directions as is the case in our example. XYZ may have improved its labour productivity figures by merely substituting contractors for permanent staff.

Total factor productivity measurement addresses this shortcoming by providing an overall indicator of performance. The general definition for a total factor productivity (TFP) measure is:

$$TFP = \frac{\text{Combined output quantity index}}{\text{Combined input quantity index}}$$

Table 5 shows TFP indices for the hypothetical data set presented in Table 3. Year 1 is set as the base period for the output and input index series. The output index is used to aggregate or combine inpatient treatments and outpatient visits into a single measure. In the same way, an input index is used to combine the four input quantities into one series.

While creating a combined index series may seem like adding apples and oranges, a standard approach is to use price data as "weights" for combining the different quantities. In this case, since there are no market prices for the outputs (as they are provided free to patients), a proxy price needs to be used to capture the different resource demands of the two outputs.

Table 5: Total Factor Productivity Analysis, XYZ Hospital

Year	Combined output index Yr 1 = 100	Combined input index Yr 1 = 100	Total factor productivity ratio Yr 1 = 100	TFP change on previous year (%)
1	100.0	100.0	100.0	-
2	102.9	99.8	103.1	3.1
3	106.8	99.2	107.7	4.4
4	110.1	101.1	108.9	1.1

In this case, based on actual practice we can justifiably assume that the resources required to treat one inpatient case equals, on average, those required to perform 40 outpatient consultations for each of the four years. Hence, if the inpatient treatment "price" is set at $2,400 per case then the price for outpatient visits will be $60 per consultation. The calculation of the input quantity uses the same approach, though actual market input price data are used as "weights" for aggregating or combining the different components.[10]

The calculation of the output and input indexes is based on a commonly used index number formula known as the Laspeyres index. The computation of this index requires data on prices of all outputs and inputs for Years 1, 2 and 3. The calculation of the combined output and input indexes using this index formula is presented in Appendix B.

The Laspeyres index number formula uses the "chain" form. This means that it uses a flexible or "moving" set of base-period prices to weight changes in outputs and inputs instead of a fixed set of base-period prices (say for Year 1). For example, using the chain form, the change in the combined input index between Year 3 and Year 4 (the current period) is calculated using Year 3 prices to weight the quantity changes.

Note that the partial factor productivity measures in Table 4c show positive changes for the labour and capital factors, and negative changes for contractors and other inputs (except for one year). In contrast, TFP performance improves consistently during the period. The TFP measure weights the individual contributions of the four inputs. Note also that total factor productivity techniques need more data than partial factor productivity measures since they require price information for construction of the index weights.

[10] In principle, a combined output index could be calculated in a similar way if the output was sold; the output prices (or relative revenue shares) would provide weights.

Where price data is either incomplete or distorted, alternative techniques can be used to produce a measure of total factor productivity change.[11] Typically, this is the case with the provision of public services by general government agencies, which do not assign market prices to their multiple outputs.

There is one major limitation of using the index numbers to measure total factor productivity. Total factor productivity change reflects the impact of all three sources of change identified in Section 3; that is, changes in technical efficiency, scale and technology. Index number techniques, however, do not provide any means for dissecting TFP change into these three components in practice.

4.3 Measuring Efficiency

4.3.1 Index Number Techniques

In principle, the scope for measuring technical efficiency can be assessed by simply comparing a set of productivity ratios across a group of organisations at a point in time. We present an approach to using index numbers to measuring efficiency by extending our previous example.

The focus shifts from analysing the productivity performance of a single hospital over a period of time (time series analysis) to comparing performance across a group of hospitals at a single point in time (cross-sectional analysis). The empirical measurement of technical efficiency is framed in terms of a benchmarking exercise.

Technical efficiency for each hospital can be viewed in terms of performance *relative* to its peers. In Section 3.2 we defined technical efficiency in relation to an individual firm's production technology. In practice, an organisation's production technology can not be easily observed.

Typically, we only have *observable* output and input data (eg, number of FTE staff) to work with. Our approach to measuring efficiency implicitly assumes that the best performers in a group are using their (common) production technology in an optimal manner; that is, they are operating at best practice "on the frontier".

Table 6a shows the inpatient output measure and full set of input data for the four hospitals, including XYZ in Year 1. Table 6b shows the set of partial factor productivity ratios and their rankings across the four hospitals.

[11] Where output price data is not available, statistical (stochastic frontier analysis) and mathematical (data envelopment analysis) techniques can be combined with the Malmquist index number formula to measure total factor productivity change at the organisational/service delivery level. These techniques are discussed in the context of their application to measuring efficiency in section 4.3.2.

Table 6a: Output and Input Data for Group of Hospitals

Hospital	Inpatient treatments (no. of cases) y_1	Outpatient visits (no. of consultn.) y_2	Labour (no. of FTE staff) x_1	Contractors (no. of hours worked) x_2	Capital (no. of beds) x_3	Other inputs (imputed quantity) x_4
XYZ	800	45,000	1,300	250,000	300	200,000
ABC	900	50,000	1,400	300,000	320	220,000
LMN	700	48,000	1,200	250,000	280	160,000
HIJ	1,000	60,000	1,500	340,000	420	240,000

Table 6b: Comparative Partial Factor Productivity Ratios and Rankings (Inpatient Treatment Output Measure)

Hospital	Ratio 1 y_1/x_1	Ratio 2 y_1/x_2	Ratio 3 y_1/x_3	Ratio 4 y_1/x_4
XYZ	0.615 (3)	0.0032 (1)	2.667 (2)	0.0400 (4)
ABC	0.643 (2)	0.0030 (2)	2.813 (1)	0.0041 (3)
LMN	0.583 (4)	0.0028 (4)	2.500 (3)	0.0044 (1)
HIJ	0.667 (1)	0.0029 (3)	2.381 (4)	0.0042 (2)

Using simple PFP ratios, there is no clear way of determining technical efficiency. For example, HIJ is the top-ranking performer for Ratio 1, the third ranking performer for Ratio 2, and the worst performer for Ratio 3. The picture would become even less clear if we considered the partial ratios incorporating the other output measure, outpatient visits.

This example illustrates that there is no clear way, even for a small group of only four organisations, to assess efficiency using partial productivity ratios because different ratios produce different performance rankings. Moreover, it is not possible to identify which organisation is inefficient and the magnitude for potential improvement.[12]

[12] To calculate an unambiguous measure of relative performance using an index number approach, a comparison of total factor productivity ratios is necessary. A "multilateral" index can be used to facilitate a comparison of TFP performance across a group of organisations over time. A multilateral index compares each organisation in an industry to a hypothetical representative entity. The representative organisation is "constructed" from average output and input data derived from all data in a given panel. Input price (or cost) data and output price (or revenue) data is required for weighting the quantity changes of the individual entities and the representative organisation (see Caves, Christensen and Diewert, 1982).

4.3.2 Statistical and Mathematical Techniques

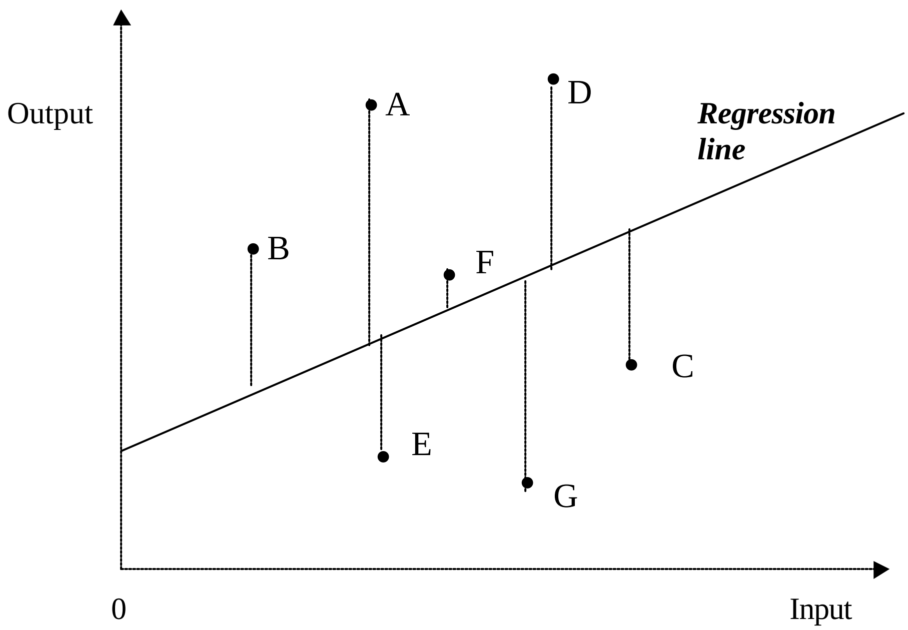
We now turn to the statistical techniques of ordinary least squares regression and stochastic frontier estimation, and the mathematical programming technique of data envelopment analysis. Unlike index number techniques, statistical and mathematical programming techniques do not require price information to calculate technical efficiency where organisations have multiple inputs and/or multiple outputs. These techniques, however, typically require data for a larger number of entities to be compared than the index numbers approach.

The statistical approach requires explicit specification of a production function (ie, the mathematical relationship between inputs and outputs) but assumes that the relationship between inputs and outputs is inexact due to measurement error and other factors. This property is captured by the inclusion of an error term that has well-defined probabilistic properties.

By comparison, the mathematical programming approach does not presuppose a particular functional form but "allows" the output and input data to determine the shape of the efficiency frontier. Moreover, the programming approach assumes an exact or deterministic relationship between inputs and outputs, which makes it sensitive to measurement error.[13]

The application of **ordinary least squares (OLS) regression** (see Figure 3) to estimate an industry production function produces a measure of efficiency that is influenced by average practice rather than best practice.

Figure 3: OLS Regression Approach to Measuring Efficiency

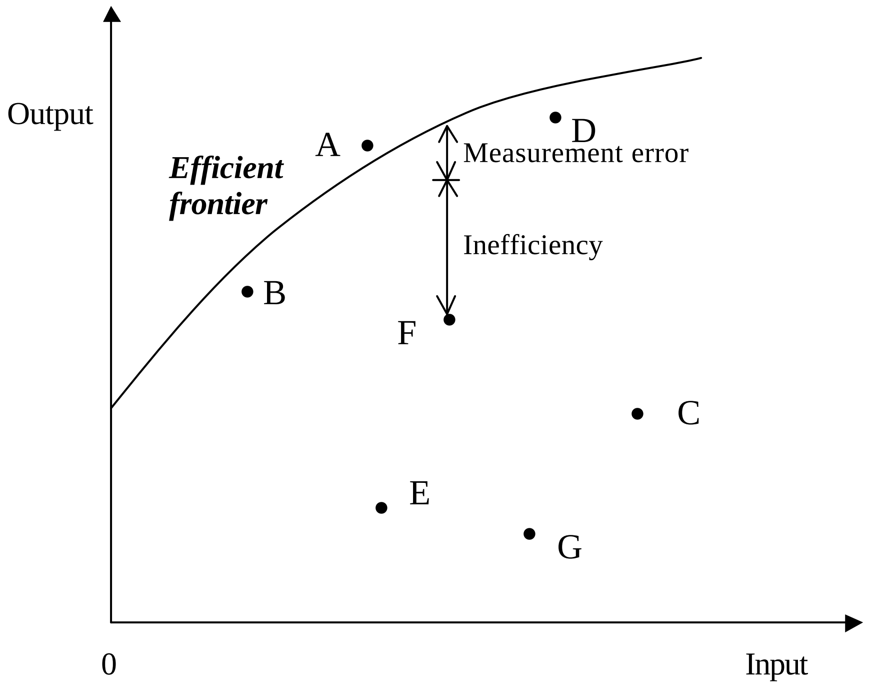


[13] In most practical applications measurement is inexact and therefore is subject to error.

The OLS technique identifies a "line of best fit" through a data set of output/input ratios for a group of organisations.[14] There will, in general, be a discrepancy between the output implied by the regression line for a given level of input (which represents average practice) and observed output at that level. This difference will, by assumption, be attributed entirely to systemic efficiency differences (rather than to random factors). The efficiency of organisations is be ranked according to these differences. The most efficient organisation will, by definition, be that with the largest positive difference (see Point A in Figure 3).

Stochastic[15] frontier analysis (SFA) is a more advanced statistical technique as it assumes that the gap between predicted and observed performance can be dissected into components for inefficiency and random noise (which is mainly measurement error).[16] A simple stochastic frontier is illustrated in Figure 4 for a group of organisations that each produces a single output using a single input.

Figure 4: Stochastic Frontier Analysis

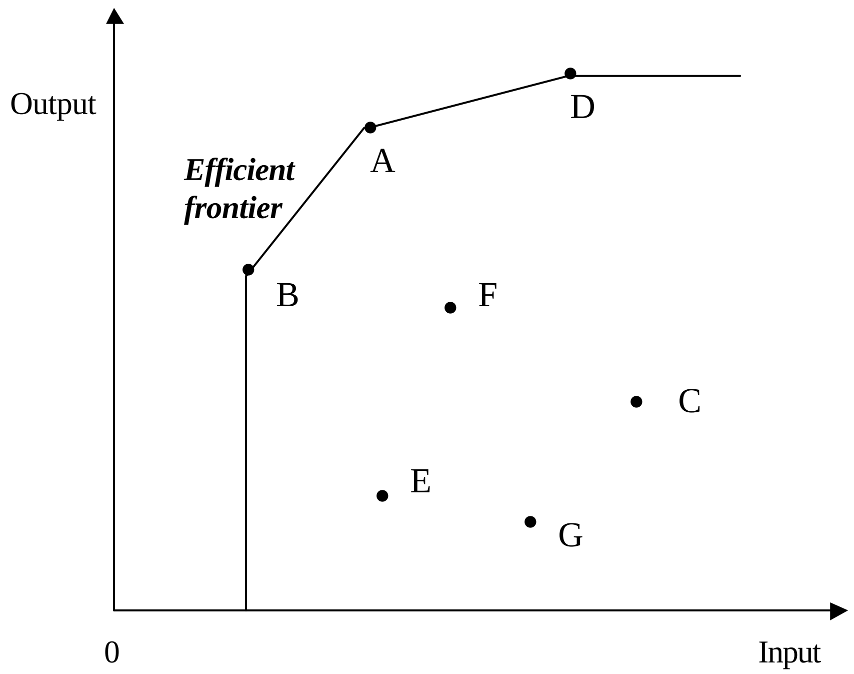


The stochastic frontier identifies the predicted performance for the best organisation (A), allowing for measurement error. The other organisations (eg, B) are below this frontier and are therefore relatively inefficient compared to the best. For these organisations SFA assumes that some of the gap between actual and predicted best performance will be measurement error. Empirical work using SFA requires the use of specialist computer software, such as FRONTIER (Coelli, 1996a).

[14] It should be noted that a regression function can be non-linear.
[15] Stochastic means "a problem involving probabilities, as opposed to a deterministic problem based on certainties" (The Economist, 1991).
[16] More formally, the gap is split into a symmetric "noise error term" (with a normal distribution) and an half-truncated "inefficiency error term" (with a one-sided distribution).

Data envelopment analysis (DEA) uses mathematical programming[17] to construct a production frontier comprising a set of linear segments. The frontier relates to best performance at a point in time. The points separating the segments are from the best practice organisations within a sample. A simple example is illustrated in Figure 5 for a group of organisations that produce a single output using a single input. The frontier "envelopes" the entities with the best output/input ratios. In comparison, a stochastic frontier is estimated using a regression approach from the most efficient organisation within a sample.

Figure 5: Data Envelopment Analysis



The distance of an inefficient organisation from the frontier is the measure of its inefficiency. A number of approaches can be applied to SFA and DEA efficiency scores to adjust for the influence of the operating environment (ie, factors beyond management control) such as climate and population density.[18] Like stochastic frontier analysis, DEA empirical applications require the use of specialist computer software, such as DEAP (Coelli, 1996b).

[17] Mathematical programming is a numerical way of finding the optimal solution to a problem subject to a set of constraints. In an efficiency measurement context, we may consider the extent to which inputs can be reduced or minimised while holding outputs constant.

[18] According to one approach outlined by Coelli et.al. (1998: 170) a DEA "two-staged method solves a [mathematical programming] problem in the first stage involving only traditional inputs and outputs. In the second stage, the efficiency scores from the first stage are regressed upon the environmental variables. The sign of the coefficients of the environmental variables indicate the direction of the influence, and standard hypothesis tests can be used to assess the strength of the relationship. The second-stage regression can be used to 'correct' the efficiency scores for environmental factors by using the estimated regression coefficients to adjust all efficiency scores to correspond to a common level of environment (eg, sample means)."

In general, frontier techniques need more data than total and partial productivity index measures as they are benchmarking techniques requiring a data set of organisations/ service delivery units for comparison. DEA results can be sensitive to the number of variables included (ie, both inputs and outputs) and the sample size. Reducing the sample size will tend to inflate the average efficiency score as it creates fewer comparable organisations and improves the likelihood of any entity being placed on the frontier "by default".[19]

As a non-statistical technique, DEA is sensitive to outliers in the sample, which are often due to measurement errors and/or random events such as climate. By contrast, SFA is less susceptible to outliers as it allows for random noise in measuring inefficiency. In addition, a mathematical functional form (eg, the shape of the curve for two dimensions or the shape of the plane for three dimensions) representing the underlying production technology has to be assumed *before* the stochastic frontier is estimated. By contrast, DEA does not presuppose a particular functional form for the frontier but allows the data to determine the shape of the frontier (eg, in effect as a set of linear segments in two dimensions, or as flat triangles in three dimensions).

For each organisation inside the frontier that is found by DEA to be inefficient, the technique identifies at least one organisation on the production frontier that is a "peer" or role model to the inefficient organisation. The technique assigns a weight to each peer, reflecting the relevance of that peer to the inefficient organisation. DEA can determine whether an organisation's technical inefficiency is primarily related to waste (ie, the use of too many inputs to produce a given level of outputs), or to the particular scale of operations.

If input and output data for a set of similar organisations is available over a period of time (panel data) then total factor productivity can be measured in a more sophisticated way than the "pure" index number approaches discussed earlier in the Guide. This is achieved by combining a frontier technique (DEA or SFA) with the Malmquist index number formula. This approach allows a change in productivity to be dissected into the three sources: changes in technical efficiency, scale and technology.

[19] In undertaking comparisons of efficiency, DEA recognises organisations as being comparable with one another where they have similar mixes of inputs and outputs.

5. APPLYING ECONOMIC PERFORMANCE MEASUREMENT TECHNIQUES

Summary

- The delivery of public services benefits from the support of robust measurement techniques. Techniques such as data envelopment analysis and stochastic frontier analysis can assist in the identification of best practice in the use of resources among a group of service delivery units.

- A summary of a DEA benchmarking exercise on Victorian acute health care services, conducted by the Steering Committee for the Review of Commonwealth/State Service Provision, is presented in this section.

5.1 Introduction

In recent years, Federal and State/Territory governments have pursued initiatives aimed at improving the efficiency of public services. These include the development of performance measures that may ultimately facilitate the linking of government service providers' budget allocations to agreed output/outcome targets, service quality standards and efficiency indicators.

In this context, the Steering Committee for the Review of Commonwealth/State Service Provision commissioned research on the application of data envelopment analysis to benchmarking the efficiency of government service delivery in the mid-1990s. The Steering Committee published *Data Envelopment Analysis: A technique for measuring the efficiency of government service provision* (1997) to promote a better understanding of DEA by explaining the technique's conceptual foundations, how to interpret the output from DEA models and its strengths and weaknesses.

The Report contains a set of case studies on acute health care services in Victorian hospitals, Queensland oral health services for school students, NSW correctional centres, NSW police patrols and NSW motor registry offices. In this section we will summarise the work undertaken on Victorian hospitals.[20] Appendix C provides the reader with a simple illustration of the DEA methodology for a hospital, which is adapted from Hughes and Yaisawarng (2000b).

[20] There have been few benchmarking studies on hospitals conducted in Australia. Preliminary work on NSW acute care public hospitals by NSW Treasury and NSW Health focused on adjusting efficiency scores for the impact of environmental factors (see Hughes and Yaisawarng, 2000a). By contrast, DEA has been widely used to benchmark hospital efficiency overseas; Hollingsworth et. al. (1999) identifies nearly 90 DEA applications to hospitals in the United States and Europe.

5.2 DEA Model

The objective of the study was to evaluate the potential for using DEA as a benchmarking tool for measuring the efficiency of acute health care services in Victorian public hospitals. A sample comprising 109 hospitals from the 1994-95 financial year was assembled for the exercise.

The sample was split into two sub-groups due to differences in input data availability and probable differences in the structure of production technology:

- metropolitan/large country hospitals (including teaching, research and base hospitals); and
- small rural hospitals (excluding base hospitals).

An "output-oriented"[21] model was selected to reflect the management objective of maximising service provision subject to resource and technology constraints. The rationale for adopting an output orientation is the fact that there is always unmet demand for public hospital services. In this situation, hospital managers seek to satisfy service demand to the maximum extent possible given limited resources and the technological constraints embodied in a hospital's production function.

The DEA model comprised the following variables:

Inputs

- Medical officers (doctors, specialists) – measured on a full-time equivalent basis
- Non-medical staff (nurses, administration staff) – measured on a full-time equivalent basis
- Non-labour operating expenditure – eg, pharmaceuticals

Outputs

- Number of patients treated by each hospital expressed in terms of "weighted inlier equivalent separations" (WIES)[22], which adjusts the number of separations (when a person leaves a hospital) by the expected resource intensity of each case. Treatments were classified as minor, intermediate or complex cases.
- Inverse of the unplanned re-admission rate (a proxy for the quality of care)

Inputs were restricted to variable resources; there was no measure of capital in the model specification. In other studies, hospital bed numbers is a common proxy measure for the capital input.

[21] An output-oriented approach to efficiency measures the extent to which an organisation's output quantities can be proportionately expanded without altering the input quantities used. This contrasts with the input-oriented approach which measures by how much input quantities can be trimmed without changing the output quantities produced.

[22] A WIES is similar, but not an equivalent, output measure to a Diagnostic-Related Group separation.

The output measure was focused on the dominant activity (with an associated quality of care variable). Adequate measures for other output categories, such as outpatient services, were not available.

In sum, the DEA model specification was focused on short-term controllable inputs and the core hospital output category.

5.3 Results

For the first sub-group, the range in efficiency scores was relatively small, with 24 of the 37 hospitals being on the efficient frontier. The mean efficiency score for metropolitan/ large country hospitals off the frontier was 1.11. This means that the average inefficient hospital could potentially increase its measured services by 11 per cent without using any additional resources, if it is able to operate at observed best practice.[23]

The efficiency scores for small rural hospitals were more dispersed, with an average score for hospitals off the frontier of 1.33. In this sub-group, only 14 of the 69 hospitals made up the efficient frontier.

The sensitivity of the two models was tested by changing the measure of labour inputs from a volume measure (full-time equivalent staff numbers) to a value measure (salary costs). This resulted in little change to both efficiency scores and the listing of efficient hospitals, which supported the proposition that that the labour measure was reasonably robust.

The DEA model specification lacked a measure of capital inputs. This absence may have biased the efficiency scores if there were substantial variations in capital resources across hospitals in the two sub-groups. In particular, hospitals with above average capital usage would produce relatively more services given their measured inputs. The model would be improved by adding a measure of capital usage.

[23] Each inefficient hospital has one or more "efficient peers" with which its performance is compared. See Appendix C for a simple illustration of the concept of efficient peers.

6. WHERE DO YOU GO FOR HELP?

6.1 Recommended Publications

- Blank, J. L. T. editor (2000), *Public Provision and Performance: Contributions from Efficiency and Productivity Measurement*, Elsevier Science B.V., Amsterdam.

- Coelli, T., D. S. Prasada Rao and G. E. Battese (1998), *An Introduction to Efficiency and Productivity Analysis,* Kluwer Academic Publishers, Boston.

- Fried, H. O., C.A Knox Lovell and S. Schmidt editors (1993), *The Measurement of Productivity Efficiency: Techniques and Applications*, Oxford University Press.

- Steering Committee for the Review of Commonwealth/State Service Provision (1997), *Data Envelopment Analysis: A technique for measuring the efficiency of government service delivery*, AGPS, Canberra.
  Internet: http://www.pc.gov.au/service/gsp/dea/index.html

6.2 Computer Software

- The Centre for Efficiency and Productivity Analysis at the University of New England (Internet: http://www.une.edu.au/febl/EconStud/emet/cepa.htm) has three "freeware" products available:

  - **FRONTIER Version 4.1 –** software program for estimating stochastic frontier models to measure technical and cost efficiencies.

  - **DEAP Version 2.1** – program for constructing data envelopment analysis frontiers to measure technical and cost efficiencies as well as Malmquist total factor productivity indices.

  - **TFPIP Version 1.0** – software program for measuring total factor productivity using Tornqvist and Fisher index number formulae.

- **Frontier Analyst** – Windows-based DEA software supplied by Banxia Software.
  Internet: http://www.banxia.com

- **SHAZAM** – widely used general econometrics package that can be applied to efficiency and productivity analysis.
  Internet: http://shazam.econ.ubc.ca

6.3 New South Wales Treasury

- Contact the Director, Fiscal Strategy, New South Wales Treasury
  (Tel: 02 9228 3469; Fax: 02 9228 4041).

APPENDIX A: ECONOMIC MEASURES OF PERFORMANCE – SHOULD THEY RELATE TO *OUTPUTS* OR *OUTCOMES*?

Economic performance measurement can be applied at different levels of aggregation. At the highest level, we can derive productivity measures for the whole economy. At the lowest level, measures for the performance of an operating unit within an entity could be derived.

Normal practice is for high level measures to be related to aggregate measures of *output* defined in the national accounts framework. However, there is no aggregate *outcome* construct, so even if we wanted to derive measures of efficiency related to *outcomes* we could not readily do so.

Lower level performance measures also tend to be cast in terms of the efficiency with which *outputs* are produced. However, there is no overwhelming in-principle reason why they could not be related to *outcomes*. Nevertheless, it is pertinent to consider how this could be accomplished, and what practical issues are raised. Before doing this, we will consider briefly what the purpose of performance measurement is. This has a bearing on the interpretation of *outcome* related performance measures.

Performance measures can serve two purposes:

- to gauge progress; and
- to diagnose performance in order to find possible ways of improving performance

If we are only interested in gauging performance - determining whether in some sense we are making progress, then any well constructed measure can help us to do this. In this regard, an *outcome* related performance measure could be quite useful.

If we wish to use performance measurement as a diagnostic tool, then to be useful it really needs to be able to tell us something about the relationship between the things we can control and the *outputs* or *outcomes* we are trying to achieve. In this respect, *outcome* related performance measures are likely to be less useful than *output* related measures. This can best be illustrated by applying the production theory that underlies economic performance measures.

For a production unit (firm, agency, or whatever) the *outputs* that are produced are determined by the *inputs* that are used. To express this as a simple non-specific mathematical relationship:

$$\textit{outputs} = f(\textit{inputs}) \qquad (1)$$

where f() is a general functional form.

The conversion of inputs into *outputs* depends on the technology employed, the skill applied by management in organising the *inputs*, and external "environmental" factors. In other words, if we were to express the production function as an explicit mathematical relationship, the mathematical formula would be influenced by both factors over which the entity has a degree of control, and those external to it that are a given and cannot be altered by the entity. Nevertheless, the function is predicated on *outputs* being dependent on (or "caused" by) the conversion of *inputs* within the entity.

When comparing the performance of different entities, it becomes important to try and understand the impact of the environmental factors, and the extent to which differences in performance can be attributed to those factors. But the premise still remains that for the entities being considered, their *output* is "caused" by their activity in converting *inputs*.

When considering the performance of government agencies, we recognise that their production of *outputs* is only as a means to an end, where the end is a desired *outcome*. This qualitatively distinguishes them from typical private sector firms (and indeed, PTEs) whose ultimate objective is to sell *outputs* at a profit. The distinction between *outputs* and *outcomes* is not relevant when considering profit making entities.

As such, the distinction between *outputs* and *outcomes* is not typically drawn out in production theory in economics – implicitly they are assumed to be the same thing. Nevertheless, we can still apply the production theory framework to a consideration of *outcomes*. Again we can state a simple non-specific mathematical relationship:

$$\textit{outcomes} = g(\textit{outputs}) \qquad (2)$$

where g() is a general functional form. Equations (1) and (2) combine to give:

$$\textit{outcomes} = g(f(\textit{inputs})) \qquad (3a)$$

or

$$\textit{outcomes} = h(\textit{inputs}) \qquad (3b)$$

where h() is a general functional form combining f() and g().

The "conversion" of *outputs* into *outcomes* depends on the appropriateness, quantity and quality of the *outputs*, and external "environmental" and other factors. To express the production function g() as an explicit mathematical relationship, the mathematical formula would be influenced by *output* related factors over which the agency has a degree of control, and external factors that are a given and cannot be altered by the agency. The premise that *outcomes* are dependent on (or "caused" by) the *outputs* produced within the agency would in general appear to be a little shaky.

By extension considering equation (3b), the premise that *outcomes* are dependent on (or “caused” by) the conversion of *inputs* within the agency is in general even more shaky. Nevertheless, as long as to some degree (however small) changes in *inputs* influence *outcomes*, equation (3b) will hold as a general expression. Given the potentially wide range of external factors that could influence *outcomes*, determining the precise functional form for h() will at best be very difficult and more likely impossible. So while as a general expression equation (3b) holds, in practice it may not be very useful.

Therefore we conclude that on theoretical grounds, the “efficiency” with which *inputs* are converted into *outcomes* is a legitimate line of inquiry. It is legitimate to posit a causal relationship between an agency’s “processing” of *inputs*, and ultimate *outcomes*. But given a plethora of influences that are outside the control of the agency, there is no theoretical basis for attributing responsibility for “performance” (in the sense of the level of efficiency, or changes in efficiency).

Furthermore, in general it is likely to be more difficult to measure *outcomes* than *outputs* because *outcomes* tend to be multi-dimensional. Therefore while such performance measures may be useful for the first of the purposes identified above for performance measures - gauging performance, they cannot be expected to be of much use for the second purpose – diagnosing performance.

APPENDIX B: CALCULATION OF TOTAL FACTOR PRODUCTIVITY – CHAIN LASPEYRES INDEX (XYZ Hospital Example)

A. OUTPUTS	**Year 1**	**Year 2**	**Year 3**	**Year 4**
A1 Output quantity data				
Inpatient treatment (number of cases)	800	820	850	875
Outpatient visits (number of consultations)	4,500	4,750	5,000	5,200
A2 Output non-market "price" data (index weights)				
Inpatient treatment (price per case)	2,400	2,400	2,400	-
Outpatient visits (price per consultation)	60	60	60	-
A3 Combined (or aggregate) output quantity index				
Combined output index number	-	1.029	1.039	1.031
Combined output index series (Year 1 = 100.0)	100.0	102.9	106.8	110.1
Annual percentage change	-	2.9%	3.9%	3.1%

Calculation of Year 4 combined output index number

[($2,400 * 875)+ ($60 * 5,200)]/ [($2,400 * 850) + ($60 * 5,000)] = 1.031

B. INPUTS	**Year 1**	**Year 2**	**Year 3**	**Year 4**
B1 Input quantity data				
Labour (number of FTE staff)	1,300	1,250	1,200	1,175
Contractors (number of hours worked)	250,000	275,000	300,000	330,000
Capital (number of beds)	300	305	310	314
Other (imputed quantity measure)	200,000	210,000	215,000	230,000
B2 Input market unit price data ($) (index weights)				
Labour	60,000	62,000	64,000	-
Contractors	70	75	78	-
Capital	3,500	3,600	3,650	-
Other	100	105	108	-
B3 Combined (or aggregate) input quantity index				
Combined input index number	-	0.998	0.994	1.019
Combined input index series (Year 1 = 100.0)	100.0	99.8	99.2	101.1
Annual percentage change	-	-0.2%	-0.6%	1.9%

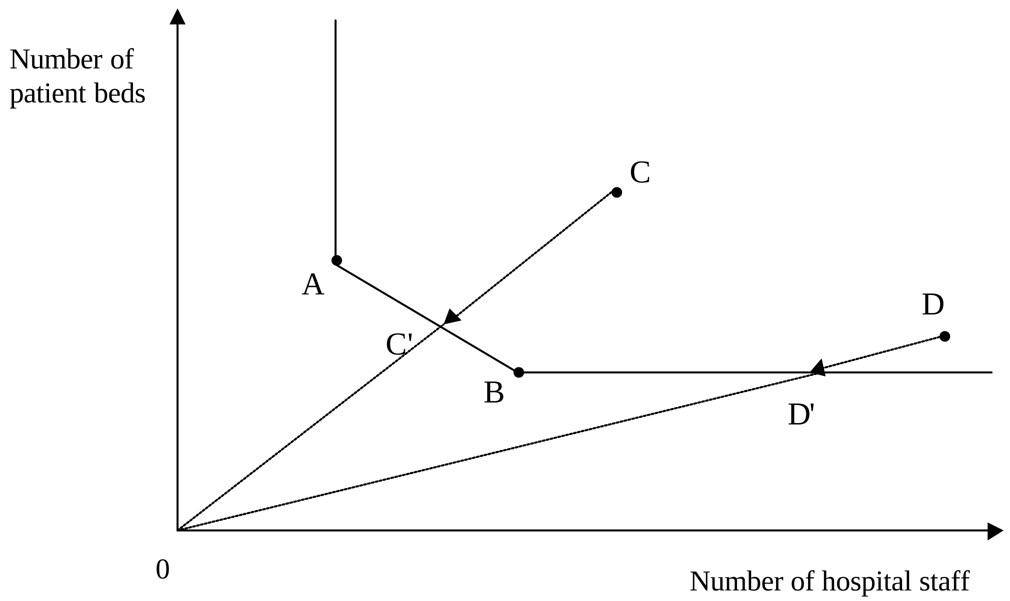
C. TOTAL FACTOR PRODUCTIVITY	**Year 1**	**Year 2**	**Year 3**	**Year 4**
Combined output index series	100.0	102.9	106.8	110.1
Combined input index series	100.0	99.8	99.2	101.1
Total factor productivity (TFP) index series	100.0	103.1	107.7	108.9
Annual percentage TFP change	-	3.1%	4.4%	1.1%

APPENDIX C: AN ILLUSTRATION OF DEA METHODOLOGY

This appendix aims to provide an intuitive understanding of the DEA efficiency measure and its interpretation, using a simple graphical illustration. The focus is on input saving and managerial efficiency; that is, to what extent can a manager trim resources without reducing the level of services (at the same scale of operations). The illustration uses a single output or service and two types of inputs.

Consider a simplified case where there are four hospitals in the sample. These hospitals deliver identical levels of services using two inputs: number of hospital staff (the labour input) and number of patient beds (proxy for the capital input). The four hospitals use different combinations of inputs to deliver services. Figure A depicts the hospitals, where each point represents the combination of inputs for a hospital.

Figure A: Input Oriented DEA Model



The best practice observations define the frontier, AB and its extensions. Hospitals A and B use the lowest combination of inputs and are accordingly judged as technically efficient. Hospitals C and D use more staff and patient beds than A and B and are therefore off the best practice observed frontier. These two hospitals are inefficient. DEA assigns efficiency scores of one to both A and B. Hospitals C and D receive efficiency scores of less than one.

To illustrate how DEA efficiency scores for inefficient units are calculated, first consider Hospital C. The target point for C is C' which lies on the best practice observed frontier and represents the same ratio of staff to patient beds as C. An efficiency score for C is computed as OC'/OC which is less than 1, say 0.8. This result indicates that C is 80 percent efficient, relative to the best hospitals in the sample.

To maintain the same level of service delivery, C could use 20 percent less staff and patient beds if the level of resource usage had been equal to that implied by the best practice observed frontier. To do so, C should imitate A and B since C' is a combination of these two hospitals.

Hospitals A and B are **efficient peers** for C. Since C' is closer to B than A, B has more impact than A. The effect that an efficient peer has on the inefficient hospital is called the **impact (or weight)**. For example, B may have a 60 percent **impact** and A may have a 40 percent impact. These impacts are used to compute the target Point C'.

Next, consider the other inefficient hospital, D. The efficiency score of D is OD'/OD which is less than one, say 0.92. If hospital D was on the frontier it would use 92 percent of currently used inputs to deliver services. This brings D to D'. However, D' is *not* a target point for D. As shown in Figure A, the number of staff at D' exceeds that at B while the number of patient beds is the same.

This suggests that, after uniformly reducing both inputs from D to D', it is possible for D to further reduce staff from D' to B. The difference in number of staff between B and D' is an **additional reduction** (**known as non-radial slack**) that is possible if D uses inputs efficiently and achieves the target Point B. In this case, B is an efficient peer for D, with a weight of one.

References

Butler, J. R. G. (1995) *Hospital Cost Analysis* (Developments in Health Economics and Public Policy, Volume 3), Kluwer Academic Publishers, Boston, Chapter 3.

Carrington, R., N. Puthucheary, D. Rose and S. Yaisawarng (1997), "Performance Measurement in Government Service Provision: The Case of Police Services in New South Wales", *Journal of Productivity Analysis*, 8: 415-430.

Caves, D.W, L. R. Christensen, and W. E. Diewert (1982), "The Economic Theory of Index Numbers", *Econometrica*, Vol. 50, No. 6, November.

Coelli, T (1996a), "A Guide to FRONTIER Version 4.1: A Computer Program for Frontier Production Function Estimation", *CEPA Working Paper 96/07*, Department of Econometrics, University of New England, Armidale.

Coelli, T (1996b), "A Guide to DEAP Version 2.1: A Data Envelopment (Computer) Program", *CEPA Working Paper 96/08*, Department of Econometrics, University of New England, Armidale.

Coelli, T., D. S. Prasada Rao and G. E. Battese (1998), *An Introduction to Efficiency and Productivity Analysis,* Kluwer Academic Publishers, Boston.

Deiwert, E. (2000), "Alternative Approaches to Measuring Productivity and Efficiency", Keynote address to the 2000 North American Productivity Workshop, Union College, Schenectady, New York, June.
Internet: http://idol.union.edu/~napw/programp.html

Fried, H. O., C. A. Knox Lovell and S. Schmidt editors (1993), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York.

Hollingsworth, B., P. Dawson and N. Maniadakis (1999), "Efficiency Measurement of Health Care: a Review of Nonparametric Methods and Applications", *Health Care Management Science*, vol. 2, no. 3, pp 161-172.

Hughes, A. and S. Yaisawarng (2000a), "Operating Environment Adjusted Efficiency of Acute Care Public Hospitals in New South Wales", in *Proceedings of the 2000 International DEA Symposium*, Ipswich, 3-5 July, University of Queensland.

Hughes, A. and S. Yaisawarng (2000b), "Efficiency of Local Police Districts: A New South Wales Experience", in J. L. T. Blank (editor), *Public Provision and Performance: Contributions from Efficiency and Productivity Measurement*, Elsevier Science B.V., Amsterdam.

Kang, K and D. Feeney (1997), "Reading the Results of Performance Measures – Has the Desired Change Occurred?", Invited paper for Conference of Performance Measures for Government, Sebel of Sydney, July.

Kendrick, J.W. and M. Frankel (2001), "Productivity", Encyclopaedia Britannica Article, Internet: http://www.britannica.com

Lawrence, D., E. Diewert and K. Fox (2001), "Who Benefits from Economic Reform: The Contribution of Productivity, Price Changes and Firm Size to Profitability", Paper presented to the Economic Measurement Group Workshop, University of New South Wales, Sydney, March.

London Economics, (1994) "Comparative Efficiency of NSW Metropolitan Distributors Benchmarking Methodology" (Report 2). Prepared for the NSW Government Pricing Tribunal (now the Independent Pricing and Regulatory Tribunal of New South Wales), May.

NSW Treasury (2000), *Financial Management Framework for the General Government Sector*, Policy and Guidelines Paper, TPP 00-4. Internet: http://www.treasury.nsw.gov.au/pubs/tpp00_4/tpp00-4.pdf

Peacock, S., C. Chan, M. Mangolini and D. Johansen (2001), *Techniques for Measuring Efficiency in Health Services*, Productivity Commission Staff Working Paper, July.

Spottiswoode, C., (2000), *Improving Police Performance: A new approach to measuring police efficiency*, Public Services Productivity Panel – HM Treasury. Internet: http://www.hm-treasury.gov.uk/pspp/studies.html

Steering Committee for the Review of Commonwealth/State Service Provision (1997), *Data Envelopment Analysis: A technique for measuring the efficiency of government service delivery*, AGPS, Canberra. Internet: http://www.pc.gov.au/service/gsp/dea/index.html

Steering Committee for the Review of Commonwealth/State Service Provision (2001), *Report on Government Services 2001*, AGPS, Canberra. Internet: http://www.pc.gov.au/gsp/2001/index.html

Steering Committee on National Performance Monitoring of Government Trading Enterprises (1992), *Measuring the Total Factor Productivity of Government Trading Enterprises*, Industry Commission, Canberra.

The Economist (1991), *Numbers Guide: The Essentials of Business Numeracy*, The Economist Books Ltd.

Tirole, J. (1994), "The Internal Organisation of Government", *Oxford Economic Papers* 46, 1-29.

Glossary/Index

Allocative efficiency The degree to which an entity uses the "right mix" of inputs and/or produces outputs in optimal proportions, given their respective prices. Page 15

Average cost Defined as total cost divided by output. It is used in the Guide to illustrate the relationship between productivity and cost minimisation. Page 14

Cross-sectional data Data for a number of observations pertaining to one point in time (eg, output data for a group of service delivery units for a particular year). Pages 5, 22

Data envelopment analysis A mathematical programming technique which identifies best practice within a sample of firms/service delivery units. Efficiency scores are calculated relative to best practice.
Pages 26-27, 28-30, 31, 36-37

Benchmarking The activity of defining a desirable standard or level of performance, and making comparisons against the standard for a group of organisations/ service delivery units. Pages 8, 16, 17, 22, 27, 28, 29

Cost efficiency The combination of allocative and technical efficiency gives a measure of cost efficiency.
Page 15

Index number techniques A mathematical approach to combining disparate quantities or prices into a single, combined measure (eg, the consumer price index). In the Guide we consider the application of these techniques to the construction of partial factor and total factor productivity indexes. We use the Laspeyres index to illustrate the main concepts and practical issues associated with the index number approach. Pages 17-23, 31, 35.

Inputs Physical resources (eg, labour, capital, materials) used in the production process to produce outputs. Inputs are associated with costs. Pages 8-9

Laspeyres index formula A widely used index formula where base period prices (quantities) provide weights for the calculation of changes in an aggregate quantity (price) index.
Pages 21, 35

Malmquist index formula A specialist index formula used with data envelopment analysis or stochastic frontier analysis to decompose productivity change into efficiency, scale and technical change components. Pages 27, 31

Mathematical programming techniques These techniques provide a numerical way of finding a solution to a problem which optimises some function subject to a set of constraints. In an efficiency measurement context, we may consider the extent to which inputs can be reduced or minimised while holding outputs constant. In the Guide, we focus on the technique of data envelopment analysis.
Pages 26-27, 28-30, 31, 36-37

Measurement error These arise from practical imperfections in identifying, recording and computing data.
Pages 24, 25, 27

Operating environment Factors that impact on an entity's performance (eg, in the use of inputs) that are not in the direct control of managers (eg, climate, population density of a service catchment area). Pages 6, 9, 26

Ordinary least squares (OLS) One of the most popular methods of regression. OLS regression identifies the equation of the line of best fit which links multiple variables and reveals the strength of their relationship, using the magnitude of estimated measurement error terms.
Pages 24-25

Outputs Physical products or intangible services that are created from a production process. For commercial activities (not covered in the Guide) outputs are associated with revenues. Pages 7-8, 32-34

Panel data A cross-sectional collection of data with a time series dimension (eg, output data for a group of service delivery units over several years). Page 27

Partial factor productivity Ratio of a single or combined measure of outputs relative to one particular input (eg, labour productivity is the ratio of output to labour input).
Pages 18-20, 23

Production function/frontier A production function describes the relationship between outputs and inputs, capturing the set of all possible (both efficient and inefficient) input-output combinations. A production frontier is the upper boundary of a production function, representing the sub-set of efficient input-output combinations. Page 11

Production technology Refers to the embedded engineering/technical knowledge in a production process. It determines the particular nature of the input-output relationship within a production function. Page 11

Productivity Defined as the ratio of outputs to inputs. Productivity growth through time can be attributed to three sources: improvements in technical efficiency, exploitation of scale economies and technological advances.
Page 10

Scale economies The extent to which an entity can improve its productivity by altering its scale of operations. The presence of scale economies can be represented graphically by a movement along a production function (or a cost function). Page 12

Statistical techniques These techniques require explicit specification of a production function but assume that the relationship between inputs and outputs is inexact due to measurement error and other factors. This aspect is captured by the inclusion of an error term that has well defined probabilistic properties. In the Guide, we consider the statistical techniques of ordinary least squares and stochastic frontier analysis. Pages 24-25, 31

Stochastic frontier analysis A method for estimating a production (or cost) frontier function using econometric methods. Page 25

Technical (or productive) efficiency Determined by the gap between an entity's actual productivity ratio and the optimal ratio, set by the particular production technology. It is represented graphically by a movement towards the upper (lower) bound of a production (cost) function. Pages 12, 15, 22-27, 28-30, 36-37

Technological change Advances or declines in the state of technical knowledge. A technological advance is a source of productivity growth. It can be represented graphically by an upward shift in a production function. Page 13

Time series data A collection of observations over time (eg, output data for a singe service delivery unit over several years). Pages 5, 22

Total factor productivity Ratio of the quantity of all outputs to the quantity of all inputs. Pages 20-22