

**What we know about the efficiency of higher  
education institutions: The best evidence**

Carlo Salerno

Senior Research Associate  
The Center for Higher Education Policy Studies  
University of Twente



## TABLE OF CONTENTS

<b>1. Introduction.....</b>	<b>5</b>
<b>2. Defining efficiency.....</b>	<b>7</b>
<b>3. Approaches to measuring efficiency.....</b>	<b>13</b>
3.1 Stochastic frontier estimation.....	13
3.2 Data envelopment analysis (DEA).....	17
3.3 Advantages and disadvantages to each technique.....	21
3.4 Output and input measures in higher education efficiency studies.....	24
<b>4. A survey of efficiency studies.....</b>	<b>28</b>
4.1 Australia.....	29
4.2 United States.....	33
4.3 United Kingdom.....	36
4.4 Canada.....	41
4.5 The Netherlands.....	43
<b>5. Discussion.....</b>	<b>51</b>
5.1 Across country comparison.....	51
5.2 Policy implications.....	57
5.3 A modest proposal.....	58
<b>References.....</b>	<b>61</b>
<b>Appendix A.....</b>	<b>65</b>

## LIST OF FIGURES AND TABLES

Figure 1: Technical Efficiency	9
Figure 2: Allocative and Overall Efficiency	10
Figure 3: Stochastic, Deterministic & Traditional Cost Functions	16
Figure 4: Graphical Depiction of DEA Efficiency	19
Figure 5: Effects of Determining Relative and not Absolute Efficiency in DEA	22
Table 1: Studies of institution-level efficiency in higher education	29
Table 2: Input and output measures used in Coelli (1996) study	30
Table 3: Output measures used in Avkiran (2001) study	32
Table 4: Input measures used in Athanassopoulus & Shale's (1997) study	37
Table 5: Variables in Stevens (2001) used to identify sources of IHE inefficiency	40
Table 6: Input/Output measures & models in McMillan & Datta (1998) study	42
Table 7: Explanatory variable used in Goudriaan, <i>et al.</i> (1998) study	48

## 1. Introduction

Two higher education topics receiving considerable attention in policy circles and within academe today are productivity and efficiency. As enrollments in higher education continue to expand, public funding is becoming increasingly scarce, particularly as competition increases from other recipients of public funds such as healthcare and corrections. In light of this many policymakers have found themselves asking if higher education institutions are using their resources productively. Over the past decade, questions of this kind have given rise to a number of studies seeking to assess productive and cost efficiency.

Yet the increase in studies of this type can also be attributed to the development of parametric and non-parametric techniques for estimating efficiency that have only recently moved beyond theoretical construction and gained popularity in more applied settings. These increasingly sophisticated approaches have finally provided researchers both the ability and flexibility necessary for modeling the complex production processes and cost structures within higher education institutions. As a result, one can look across education systems in several countries and find a growing repository of empirical studies that shed new light on our understanding of higher education efficiency.

Remarkably though, within any given country it is not possible to identify more than a handful of empirical studies. As we will see later in this paper, an added layer of complexity is evident in the diverse methodologies different studies have employed. In short, the best evidence researchers have about higher education efficiency is scattered among a diverse set of educational systems that are more apt to be different than similar in many aspects.

Nevertheless, it is worth asking whether there is anything to gain by coalescing what is known about higher education efficiency and reflecting on the state of the art. What common threads exist to tie these different studies together? What can be learned by examining how one studies of efficiency in one system that have yet to be applied in another? How do the efficiency findings from one country's higher education system compare to another? Is it even possible to draw such comparisons? It is precisely these types of questions that motivated the Netherlands' Ministry of Education, Culture and Science (*Ministerie van Onderwijs, Cultuur, en Wetenschappelijk*) to commission the study at hand.

This paper takes the questions above to task and surveys what is currently known about efficiency in higher education. Briefly, the format of the rest of the paper is as follows. In the next (second) section I formally define what economists mean when they speak of efficiency and explain how it fits into production theory. This is followed by a description and critique of the different empirical tools researchers have at their disposal to derive estimates of efficiency. The fourth section details the various studies that have been conducted and discusses their findings. The fifth and final section is a “comparison and contrast” of sorts; it addresses the extent to which the findings from different studies can be more broadly applied, whether cross-country comparisons can be made, briefly outlines some guidelines for a more in-depth comparative analysis of efficiency.

## 2. Defining efficiency

Before turning our attention to the various studies that have already been conducted, it is first worthwhile to ask, “What do we actually mean when we are talking about efficiency?” This may seem superfluous yet oftentimes many people use terms like productivity and efficiency interchangeably, thinking they are equivalent when in fact they are distinctly different. As we will soon see, economists are very precise in what they mean by efficiency and, importantly, there are many different types of efficiency, each interconnected to the others.

What is productivity? In the most straightforward sense, productivity may be defined for any given firm, organization, or operating unit as simply the ratio of output produced to physical inputs used. In higher education, examples of single-input and single-output productivity measure might include the number of students (educated) per faculty member or the number of journal articles published per researcher.

While measures like these are interesting and useful in certain contexts, they are quite limited for describing the aggregate productivity of an institution like a university that uses several ( $m$ ) inputs to produce several ( $n$ ) outputs. Here, in order to arrive at a single-valued productivity measure. This is done by developing composite, or virtual, inputs and outputs by attaching some relative importance, or weight, to each input and output. From this it is possible to derive a Total Factor Productivity measure like that shown in equation (1):

$$\text{Total productivity} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{j=1}^m d_j x_j} \quad (1)$$

Where:

$y_i$  = the  $i^{\text{th}}$  output produced

$x_j$  = the  $j^{\text{th}}$  input used

$w_i$  = the relative importance of output  $y_i$  to the institution

$d_j$  = the relative importance of input  $x_j$  to the production process

In sum, productivity measures are nothing more than rank-free indicators of the rate at which inputs are translated into outputs.

How then does efficiency relate to productivity? In the most straightforward sense, the former can be seen as an index of the latter. If one were to calculate productivity estimates for a set of institutions (in keeping with the higher education focus) and seek to identify the most (least) productive unit, efficiency can be defined as the index used to rank the different productivity values. Productivity then, is a *value* assigned to the rate at which inputs are converted into outputs and efficiency is a *ranking* of different values.

To this point we have not distinguished between how inputs are characterized, though it is generally assumed that inputs are expressed in physical quantities. Yet it is also possible to show how these ideas work when inputs are measured not by physical input usage but by costs. If, for example, the relevant input measure was “expenditures on instructional activities” and output was proxied by the number of educated students, then the productivity measure becomes “number of students educated per euro spent on faculty.” By taking the inverse of this measure one gets the more familiar looking “faculty expenditures per student.” Again, if one identifies and ranks this cost-based productivity measure, that ranking assesses the cost efficiency of each institution relative to the others being evaluated.

There are many different forms of efficiency that can be estimated; here I focus on the four that are most often evaluated in the context of higher education institutions.<sup>1</sup> The first is referred to as *technical efficiency*. Intuitively, technical efficiency is a measure of the extent to which an institution efficiently allocates the physical inputs at its disposal for a given level of output.<sup>2</sup> This is made clearer by the graphical example shown in Figure 1. For the sake of simplicity here we consider the case of one output (education) and two physical inputs, numbers of staff and numbers of computers. Note that

---

<sup>1</sup> This paper only examines efficiency from the perspective of the individual institution or one of its subunits. One can however also speak of industry efficiency. Because no higher education studies address this issue, it is not discussed here. In general though, one can say that industry efficiency measures the extent to which resources are allocated efficiently between firms and whether these firms are producing the right goods given their resources. See Nicholson (1995) for a more detailed discussion.

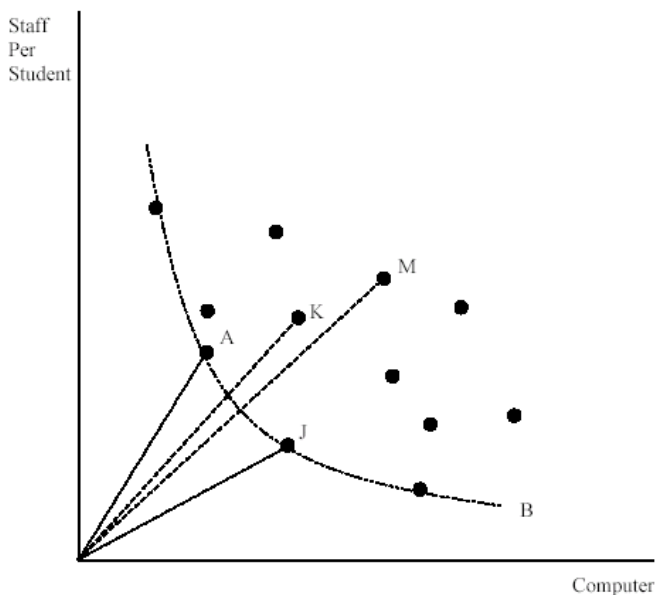
<sup>2</sup> A more formal, economic definition is as follows: A firm has allocated a fixed set of resources efficiently if it has them fully employed and the rate of technical substitution between pairs of inputs is the same for every output the firm produces (Nicholson, 1995, p. 549).



the two axis measure the inputs used per student. This is done so that we can graph institutions of different size.

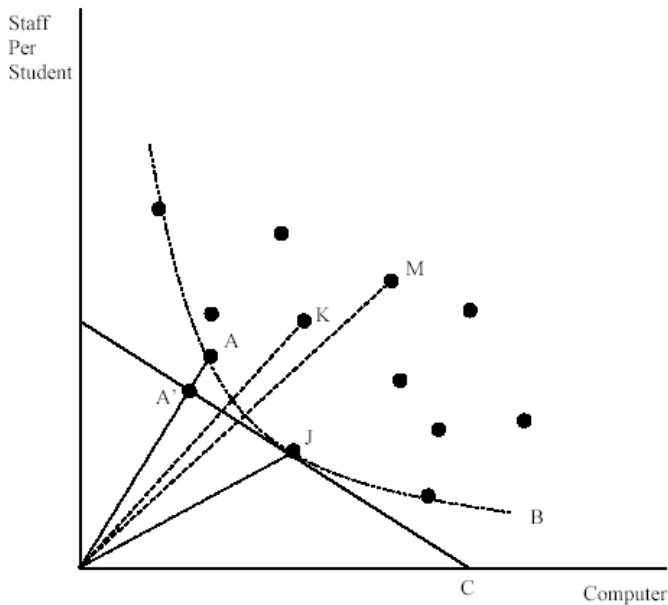
The first step in determining efficiency is to identify some standard, or benchmark from which estimates can be derived. This is done by identifying those institutions using the least amount of inputs per output; in other words those institutions that can be regarded as efficient. By fitting a line through these observations one creates an envelope, or frontier from which the inefficiency in other institutions can be evaluated. This is shown by line B in Figure 1 and any point on the line can be regarded as an equally acceptable combination of input-minimizing bundles. It is clear that any institution not lying on line B, which is in economics referred to as the isoquant, must be inefficient. For example, one can see that institution K in Figure 1 uses more computers per student and staff per student than both institutions A and J. If one wants to measure the extent to which institution K is inefficient, it can be done in several ways but simplest approach is to consider how much of a reduction it must make to move back to the isoquant (the efficient combinations); that is, moving along the dotted line in Figure 1 towards the origin. By measuring inefficiency in this way, it is possible to see how measures can be calculated for individual institutions. If we turn our attention now to institution M, it is clear that M is even further away from the isoquant than K and, given the discussion above, we can say that M is *less efficient* or *more inefficient* than institution K.

Figure 1: Technical Efficiency



Technical efficiency captures the extent to which physical inputs are efficiently allocated. Two additional measures of efficiency alternatively consider how efficient institutions are based on costs. These can be seen in Figure 2. Here, we have reproduced Figure 1 and simply included a line (C), what economists call an isocost, that describes the rate at which the two inputs can be traded off in the market (i.e. their relative costs). In short it represents the different input combinations that can be purchased from a fixed budget.

*Figure 2: Allocative and Overall Efficiency*



The second efficiency measure we are interested in, allocative or price efficiency, measures the extent to which inefficiency occurs because an institution is using the “wrong” combination of inputs given what they cost to purchase. It is measured, for each institution being evaluated, by the distance between the isoquant and isocost lines. Again it is helpful to utilize a simple example.

Above it was shown that an institution operating at point A in Figure 2 would be regarded as technically efficient. However, when costs are also considered we see that A is not totally efficient because it is operating above the isocost line (in the same way that institutions above the isoquant were shown to be inefficient). In other words, A could reduce the number of staff and increase the number of computers (equivalent to moving to the right on the isoquant) and instead operate at point J.

Note that at this point, the institution is still producing the same amount of output (because they are still situated on line B) but now is doing it for less cost than before. Thus, A is allocatively inefficient and we can measure the amount of that inefficiency by the distance between A and A'

A third type of efficiency, called economic or overall efficiency, jointly considers technical and allocative efficiencies; in essence it captures the extent to which each institution lies off of the isocost line (C). Thus, since point J in Figure 2 is on both the isoquant and isocost lines, the firm is technically and allocatively efficient; hence it is overall efficient. In contrast K is both technically and allocatively inefficient. Again the amount of each inefficiency can be measured by following the dotted line between K and the origin.

Finally, the last type of efficiency that is frequently estimated in studies on higher education institutions is scale efficiency. Many empirical studies of higher education costs frequently seek to measure the extent to which institutions are operating at increasing (decreasing) returns to scale, which in turn helps to determine the optimal size of an institution. However, since deviations from the optimal size are clearly suboptimal, they can be regarded as inefficiencies.

What condition must an institution meet if it is to be scale efficient? Economic theory suggests that, in the long run, competitive firms will continue adjusting their scale size to the point that they operate at constant *returns to scale* (CRS); thus scale inefficiency arises when institutions *are not* operating at CRS. Formally, we can say that an institution is operating at CRS if doubling all inputs results in a doubling of the output. From this we can now characterize deviations from CRS and subsequently, different forms of scale inefficiency. So if, on the one hand, doubling the inputs results in a less than equal increase in output then the institution is said to be operating at *decreasing returns to scale* (DRS). On the other hand, if scaling up inputs entails a greater than equal increase in output then it is said to be operating at *increasing returns to scale* (IRS).

Economists often equate DRS with increasing levels of institutional bureaucracy. For example, as paperwork increases and lines of authority grow longer (by doubling inputs) such factors are more likely to reduce production rather than increase it. On the other hand, IRS is often equated with small start up companies. To take a higher education example, one professor may be able to teach up to 100 students in a class but if there are only 70 students enrolled, the cost per student is relatively high.

As enrollments expand, per-student costs decline because the institution does not have to purchase more inputs (professors) but can increase output.

Thus it is not only interesting to know whether or not an institution is scale efficient at all but also whether it is too large or too small. Unfortunately, the mathematics behind computing scale efficiencies is beyond the scope of this paper. In general though, scale efficiency can be computed in several ways depending on which estimation technique the researcher chooses to employ.

### 3. Approaches to measuring efficiency

Now that we have a better understanding of what efficiency is, the logical follow-up question is, “What tools do researchers have at their disposal for empirically estimating these different types of efficiency?” Broadly speaking there are two general types of techniques: 1) parametric, or regression based estimators, and 2) non-parametric, or mathematically programming estimators. While both seek to characterize and quantify notions of efficiency, they are fundamentally different in their construction and underlying assumptions. Because each possesses its own strengths and limitations, neither is generally regarded to be superior to the other. The following three sub-sections look in greater detail at two of the most popular techniques for estimating efficiency in higher education: stochastic frontier estimation (parametric) and data envelopment analysis (non-parametric).

#### 3.1 Stochastic frontier estimation

Perhaps the most widely used tool for studying the cost structure of higher education is *cost function* estimation. In essence, this function traces out what economists refer to as a firm’s (or industry’s) *expansion path* of cost-minimizing production possibilities and is usually written in the following way:<sup>3</sup>

$$C = f(q_1, q_2, \dots, q_n, p_1, p_2, \dots, p_m) \quad (2)$$

Where:

C = total costs

$q_i$  = quantity of the  $i_{th}$  output produced

$p_j$  = price paid for the  $j_{th}$  input

An appealing feature of the cost function is its unique relationship to another fundamental concept in economics: the production function. In fact, one of the cornerstones of production theory in economics, attributable to Shepherd (1953), is that the production and cost functions are actually two different ways of examining the *same* production phenomenon. Thus, in cases where the production

---

<sup>3</sup> The arguments in a cost function are the  $n$  outputs produced by the firm and  $m$  prices it pays for inputs. Thus the function maps the minimum production costs for increasing levels of outputs and/or input prices. See Varian (1992) for a detailed discussion of the cost function’s mathematical properties.

function is not known or difficult to model (as is frequently the case with higher education), this *duality* between the two functions often makes it easier to empirically estimate cost functions because the data is more readily available.<sup>4</sup>

There are many ways to characterize the *functional form*, or the mathematical relationship between outputs and input prices, in (2) in order to derive empirical estimates. In industries like higher education, where researchers do not know, *a priori*, which functional form to use, the most common approach has been to estimate what is called a *flexible functional form* model. Unlike exact specifications (e.g. a Cobb-Douglas or Leontief cost technology), these models relax the limitations over the form in which the feasible technology could take. An example of such a flexible form can be seen in equation (3):

$$\begin{aligned} \log C(\mathbf{q}, \mathbf{p}) = & a_0 + \sum_{i=1}^n a_i \log(q_i) + \sum_{i=1}^m b_i \log(p_i) + 1/2 \sum_{i=1}^n \sum_{j=1}^n d_{ij} \log(q_i) \log(q_j) \\ & + 1/2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \log(p_i) \log(p_j) + 1/2 \sum_{i=1}^m \sum_{j=1}^n g_{ij} \log(p_i) \log(q_k) + u \end{aligned} \quad (3)$$

Where:

C = Total (or variable) costs

q<sub>i</sub> = output i

p<sub>i</sub> = input price i

a, b, d, f, g = coefficients to be estimated

m, n = numbers of input prices and outputs

u = randomly distributed error term

Equation (3) is a specific example of how (2) may be specified in an empirical analysis. It is a flexible functional form<sup>5</sup> model and was used by Glass, McKillop and Hyndman (1995) in their study on the cost efficiency of UK universities.

---

<sup>4</sup> Cost function estimations only require data on outputs and input prices, both of which are often routinely collected. In contrast, production function specifications may require data for numerous input variables that are not available. In such cases, it may be too costly to obtain the necessary information or require the use of proxy variables.

Cost functions are a useful tool in the eyes of researchers because they allow estimates to be made of a variety of interesting concepts including the marginal costs of producing different outputs, economies of scale, and economies of scope. Yet as early as the late 1960s, economists began questioning the consequences of assuming the firms under study minimize (or at least seek to minimize) costs. For one, economic theories of nonprofit behavior persuasively argue that organizations like higher education institutions have little incentive to engage in efficient production practices (James, 1990; James and Rose-Ackermann, 1986). Second, and more important from a modeling point of view, it is obvious that traditional parametric estimations like ordinary least squares fit a regression line *through* the data; they characterize the behavior of the average firm. By definition then some observations *must* lay below the regression line, which, contrary to the cost-minimization assumption, means some institutions' costs are less than the minimum cost!

Clearly such a rigorous application of textbook cost minimization is not likely to occur even in commodity industries like corn production much less in higher education. Nonetheless, for higher education where inefficiency is the rule and *not* the exception, traditional regression analyses will invariably distort estimates of important concepts like marginal costs and economies of scale. For these reasons, econometricians developed new parametric techniques that specifically sought to disentangle inefficiency from estimates. Today these are known as *stochastic frontier estimators* (SFE). The development of these models is largely credited to the work of Aigner, *et al.* (1977) and Jondrow, *et al.* (1982).

Stochastic frontier estimators are very similar to traditional parametric regressions (e.g. Equation 3). The difference between the two is that in an SFE regression, the traditional random error term is divided into two components: a normally distributed random error term ( $u$ ) and a second, strictly positive error term that captures inefficiency ( $v$ ):<sup>6</sup>

---

<sup>5</sup> There are multiple kinds of flexible forms that can be estimated. Equation (3) is referred to as the transcendental logarithmic model.

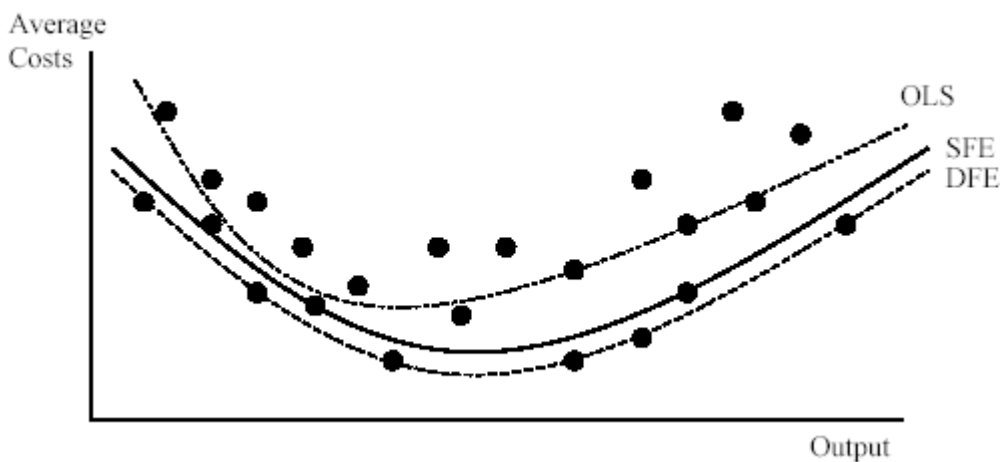
<sup>6</sup> In the empirical literature,  $v$  is generally assumed to be half-normally distributed. This simply means that if one were to take a picture of the standard bell curve from basic statistics, cut it in half and look only at the right hand side (the positive side), the distribution of  $v$  would be shaped like the right hand side.

$$\begin{aligned} \log C(\mathbf{q}, \mathbf{p}) = & a_0 + \sum_{i=1}^n a_i \log(q_i) + \sum_{i=1}^m b_i \log(p_i) + 1/2 \sum_{i=1}^n \sum_{j=1}^n d_{ij} \log(q_i) \log(q_j) \\ & + 1/2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \log(p_i) \log(p_j) + 1/2 \sum_{i=1}^m \sum_{j=1}^n g_{ij} \log(p_i) \log(q_k) + v + u \end{aligned} \quad (4)$$

Once a particular distributional form is selected for the efficiency residual term ( $v$ ), it is a relatively straightforward exercise to derive all of the estimates found in a traditional cost function<sup>7</sup> and also determines for each observation (through the  $v$  term) deviations from cost efficient behavior (Coelli, *et al.*, 1998).

Equations (3) and (4) are difficult to conceptualize; in order to get a more intuitive feel for how SFE works, it is better to illustrate the concept using the simple graphical example presented in Figure 3. The figure depicts the simple case where a single output (say education) is regressed against average costs (say educational expenditures). Because we are evaluating average costs on the Y-axis one gets the traditional U-shaped curve often seen in introductory microeconomics courses.

Figure 3: Stochastic, Deterministic & Traditional Cost Functions



As the figure shows in typical regression analysis like OLS, the objective is to fit a line through the data that minimizes the sum of the squared deviations from the line. Again though, because the cost function implicitly assumes cost-efficiency, technically speaking no point should lie below the

<sup>7</sup> Including estimates of scale and scope economies.



regression line. The first solution posed by econometricians for this problem was to fit the regression line through the *least cost* observations, in essence enveloping the data by a frontier. This is depicted by the line labeled DFE in Figure 3. In these early models all deviations from the frontier were attributed to inefficiency. Because they took no stock of random error in the data, they were referred to as Deterministic Frontier Estimators (DFE).

While they were seen as an improvement over traditional methods because of their ability to assess efficiency, economists were troubled by DFE's inability to account for random noise, which placed severe restrictions on the use of hypothesis testing. As a result, econometricians developed the stochastic, or error-inclusive, frontier estimator. This is depicted by the solid line in Figure 3. Note that the SFE line lies below the OLS regression, yet there still are observations lying below the SFE line. These low points reflect the random noise in the data. After accounting for positive random errors, it is then a straightforward task to measure inefficiency as the remaining error in the model.

### **3.2 Data envelopment analysis (DEA)**

Stochastic frontier estimators provide parametric estimates of efficiency. In other words, the parameters of a model are first specified and then estimated using real or simulated data. The second general approach to estimating efficiency is non-parametric. That is, rather than estimate values for selected parameters, the non-parametric approach relies on linear programming or some other form of mathematical programming to characterize the set of efficient producers and then derive estimates of efficiency for inefficient observations based on how far they deviate from the most efficient ones.

At the heart of non-parametric efficiency analysis is deriving measures of productivity and then determining efficiency by developing some unobtrusive way to rank them. Unfortunately because higher education institutions use multiple inputs to produce multiple outputs, developing a suitable measure of Total Factor Productivity (TFP) is very difficult. Recall from the beginning of the previous section that TFP measures require finding some way to fairly assign weights, or importance, to the various inputs and outputs. One approach might be for the researcher to choose such weights in advance. Unfortunately, this not only requires a high degree of expert judgment on the part of the researcher but also, since each institution is likely to be unique in how it values its resources, it would

require intricate knowledge of each institution. The likelihood of any researcher possessing such information is in all probability very small.

Finding a way to fairly assign such weights was first addressed by Farrell (1957) but was not fully resolved until Charnes, Cooper and Rhodes (1978) published their seminal paper in which they coined the term “data envelopment analysis” (DEA). Today DEA has almost become synonymous with non-parametric efficiency estimation.

What Charnes *et al.* did was construct an optimization algorithm, based on linear programming that identified the weights that would maximize, for each individual institution in a given analysis, equation (1). In doing so they showed how the efficiency of a given decision-making unit,<sup>8</sup>  $R$ , using  $m$  inputs to produce  $k$  outputs relative to  $n$  other institutions could be calculated:

$$\begin{aligned}
 TE_R(Y, X) = \underset{\lambda_j, \theta}{\text{minimize}} \theta & \qquad (5) \\
 \text{s.t.} \quad Y \cdot \lambda & \geq Y_R \\
 X \cdot \lambda & \leq X_r \cdot \theta \\
 \lambda_j, \theta & \geq 0
 \end{aligned}$$

Where:

$X = m \times n$  matrix of inputs

$Y = k \times n$  matrix of outputs

$\lambda = n \times 1$  vector of weights

$\theta =$  efficiency score of institution  $R$

$X_r = m \times 1$  vector of institution  $R$ 's inputs

$Y_R = k \times 1$  vector of institution  $R$ 's outputs

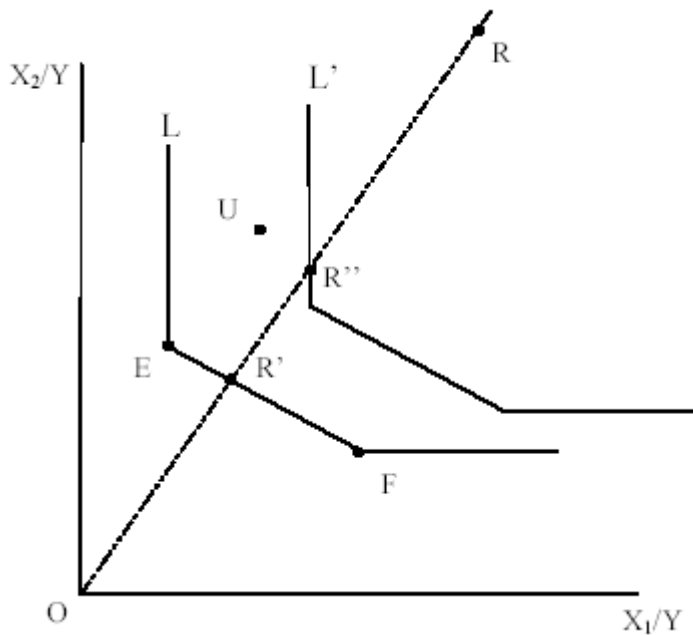
Equation (5) is the most basic of all DEA models and was the result of Charnes *et al.*'s seminal 1978 paper. Because (6) is difficult to conceptualize, it is again helpful to use a simple graphical example. Consider a set of four institutions (E,F,U and R) each using two inputs (say faculty salaries  $X_1$  and

---

<sup>8</sup> In order to make the model as general as possible, Charnes *et al.* used the term “decision making unit” (DMU) to describe the unit of analysis.

capital expenditures  $X_2$ ) to produce one output ( $Y$ , the number of educated students). These four institutions are plotted graphically in Figure 4 (In fact along the axes is the ratio of inputs to the output; in other words: our cost per student measure).

Figure 4: Graphical Depiction of DEA Efficiency



Intuitively, what equation (5) does is to construct a convex envelope<sup>9</sup> around the data defined by those institutions using the least amount of input to produce a given output level, here E and F. It is clear from the figure that both E and F use fewer inputs per student (in terms of cost) than U or R. In addition, notice that institution E uses less of  $X_1$  (salaries) than F, and F uses less of  $X_2$  (capital expenditures) than E. Hence both are considered “best practice” (cost efficient) institutions here since they are minimizing the use of at least one input relative to all other institutions being evaluated. That there is a line connecting E and F is a key assumption in the model. Specifically, for any two feasible production plans (i.e. points in the graph), any linear combination of those plans is also assumed to be feasible.

<sup>9</sup> The mathematical properties of the DEA model ensure that the space bounded by the frontier (representing technologically feasible production opportunities) is convex. As such, the frontier is also referred to as the convex hull of the technology.

As can be seen in Figure 4, R is using more of both inputs than all other institutions being evaluated; it is clearly cost inefficient. Other institutions can produce the same number of students at a lower per-unit cost (E and F) and this is directly observed behavior. In determining how inefficient institution R is relative to the other institutions, equation (5) intuitively looks at institution R and considers how far along the dotted line R must move (towards the origin) in order to be considered efficient. As the figure shows, this would occur at R'. To calculate the efficiency of R ( $\theta$ ), the model computes the ratio  $OR'/OR$ . To do so, it searches for the best possible set of weights to minimize this ratio.<sup>10</sup> As is evident in the figure, as R moves closer to the best practice line (L),  $\theta$  approaches a value of one: hence efficiency scores are bounded by 0 and 1 and can be represented as a percentage value (e.g. 45% or 90% efficient).

The efficiency model in (5) is a constant returns to scale (CRS) model; in other words, it calculates efficiency where the researcher assumes that it is feasible for all institutions to double their output if all inputs are doubled. This is a rather rigorous assumption and in some cases, particularly where institutions can notably vary in size, it may be more instructive to relax this condition. One flexible alternative is to estimate efficiency under the assumption of variable returns to scale (VRS) (Banker, Charnes, & Cooper, 1984). In Figure 4, this is shown by the second line (L'). It is readily apparent that when the CRS restriction is relaxed, institution R performs much better: now it is only compared to institutions the same size or larger. As such, efficiency is now measured by the ratio  $OR''/OR$ . These models together highlight two important observations about DEA efficiency. One, efficiency scores in a VRS analysis will always be higher than those in a CRS analysis. Two, if both VRS and CRS efficiency are estimated, then it is possible to determine the amount of scale inefficiency present by taking the ratio  $CRS/VRS$ . Thus, for an institution that is 90% CRS efficient and 95% VRS efficient, by taking the ratio of the scores we arrive at a scale efficiency score of 94.7%. Based on this value, we can say that 5.3% of the inefficiency found for this institution is due to not operating at a scale consistent with long-run equilibrium.

---

<sup>10</sup> Note that the weights from equation (4) do not appear in equation (5). As every linear programming model has a dual model, it is possible to reformulate (5) into what the literature refers to as “multiplier form,” which specifically seeks to identify the optimal weights from equation (4). Conversely, equation (5) can be regarded as the dual, or envelope formulation of the multiplier model. Most DEA analyses tend to depict (5) because it is much simpler to draw parallels with graphical measures of efficiency like that in Figure 4.

### 3.3 Advantages and disadvantages to each technique

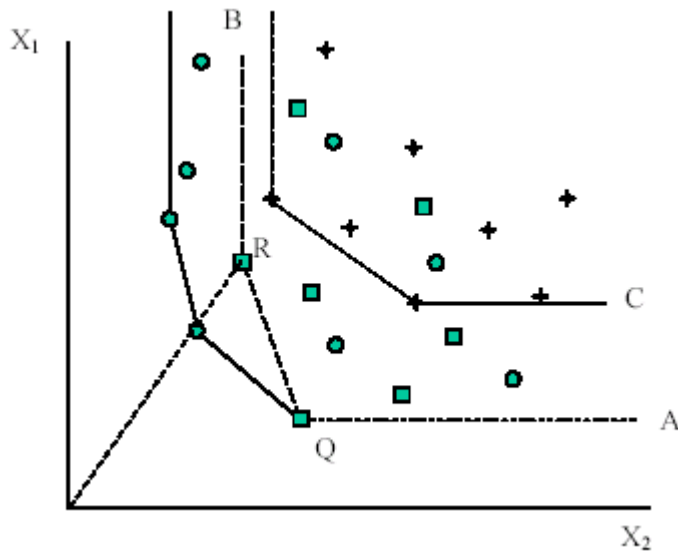
At the beginning of this section it was pointed out that both ways of assessing efficiency have their strengths and limitations. Here I briefly outline the main points to each approach, side by side, to give the reader a more balanced picture of the implications associate with using the different approaches.

Researchers preferring to estimate efficiency using SFE repeatedly cite two major limitations to the DEA approach: 1) sensitivity to data errors, and 2) the fact that it assesses relative, and not absolute, efficiency. DEA is not only non-parametric but is also a deterministic approach to assessing efficiency; as such it makes no allowance for the possibility of random errors in the data. Because efficiency is estimated *relative to other institutions being evaluated*, outliers in the data may alter the shape of the best practice frontier and distort the efficiency scores of institutions using similar input/output proportions. Mettas, Vargas and Whybark (2001), for example, have shown that DEA efficiency scores can be highly sensitive to data errors.

A second, closely related issue arises from the fact that DEA constructs a frontier from the data itself. Hence, the efficiency measures derived in any given analysis are only valid in as much as they reflect how efficient DMUs are, relative to others *in that particular sample*. Consider the scenarios outlined in Figure 5, which is an input-space map for a population of firms (the whole of which is represented by stars, circles, and triangles) producing a single output using two inputs.

In the first case, the researcher performs a DEA analysis on the entire population with the constructed frontier graphically represented by the line marked “A.” Now assume the researcher instead draws a random sample of firms (or selects a particular sample for another reason), resulting in the construction of the frontier represented in Figure 5 by line “B.” It is clear from the figure that when the entire population was considered, firm R was not found to be efficient but is efficient in the new sample. At the same time, firm Q is shown to be efficient in both instances. Had the researcher chosen, randomly or otherwise, another sample of firms, the results would again likely differ, as shown by the frontier marked “C” that evaluates only those firms depicted as stars.

Figure 5: Effects of Determining Relative and not Absolute Efficiency in DEA



What this example demonstrates is two important ways in which DEA results change with the selection of decision-making units to be evaluated. First the shape of the constructed isoquant may change depending on characteristics of firms in the sample. Second, as was discussed above, firms deemed efficient relative to one group may in fact be inefficient when compared to another. The main point, however, to be taken away from this discussion is that computing efficiency in this way is that it is not possible to develop measures of *absolute* efficiency. Even if the entire population of some group of firms was analyzed, one cannot say that the constructed frontier represents the absolute minimum input usage possible in the production of the outputs specified.

In contrast, neither of these concerns poses much of a problem in SFE analyses. Because it explicitly allows for the presence of noisy data, SFE can effectively deal with random error through statistical inference on the estimated parameters. In terms of relative versus absolute efficiency, since SFE characterizes the behavior of the “average firm” (after taking into account the distribution of efficiency scores) estimates are much less sensitive to changes in a single data point. As the frontier reflects the average firm *after* efficiency is taken into account, what is left is a hypothetically absolutely efficient frontier.<sup>11</sup>

<sup>11</sup> Researchers have increasingly been working toward the development of stochastic DEA models. These are, however, have yet to receive much attention in practical applications.

In the same vein, researchers preferring to use DEA point out that SFE is not without its fair share of limitations as well. The main concern here lies with the fact that it is necessary in SFE to make assumptions over how efficiency is distributed. It seems a questionable practice to make assumptions about efficiency when that is what is specifically being assessed! Nonetheless, incorrect assumptions about the distribution of inefficiencies will invariably lead to the possibility of biased and/or inconsistent efficiency estimates. In two of the efficiency studies looked at in the next section (Salerno, 2002; Ahn, *et al.*, 1988), researchers found the distribution of efficiency scores did not follow a half-normal or exponential distribution as is commonly assumed in SFE analyses.

The other concerns have less to do with SFE and more to do with parametric estimation in general. First, in industries like higher education where the production process is largely unknown, parametric estimation requires imposing additional assumptions about the technology, which again if they are incorrectly specified may lead to biased and inconsistent estimators. In his review of the higher education cost function literature, Brinkman (1990) found both marginal cost estimates and estimates of scale and scope economies to be highly sensitive to the choice of functional form. Second, the dependent variable in cost function and production function regressions can only be a single value. As a result, it is not generally possible to jointly estimate the influence explanatory variables have on multiple expenditures (in the case of cost functions) or multiple outputs (in the case of production functions).

In direct contrast, these limitations are effectively dealt with in the DEA approach. Efficiency estimates in DEA are based on the behavior of other institutions; as a result there is no need to draw assumptions about efficiency *a priori*. The efficiency identified is directly observable since other institutions have already demonstrated higher levels of efficiency can be achieved. Second, the danger of imposing incorrect assumptions on the model is mitigated because the non-parametric nature of DEA means only very few assumptions are imposed on the underlying technology.<sup>12</sup> Finally, DEA is widely lauded for its ability to estimate efficiency where firms use multiple inputs to produce multiple outputs and the underlying production relationship is not well understood (Cooper, Seiford and Tone, 2000). Whereas an SFE cost function can only consider one expenditure category at a time (i.e. a

---

<sup>12</sup> The “typical” DEA model imposes three assumptions: 1) free-disposability of inputs and/or outputs, 2) that if any two production plans (observations) are feasible then a linear combination of those plans is also feasible (the convexity assumption), and 3) what is called the “trivial” assumption that the inputs specified can produce the outputs specified.

single-valued dependent variable), in DEA it is possible to examine how multiple expenditure categories are likely to influence cost efficiency.

What one sees then is that the advantages to using either approach tend to rectify the disadvantages in the other. Shortcomings of the DEA approach, like only being able to assess relative efficiency or being highly sensitive to data errors, are the primary reasons why some researchers prefer using SFE. At the same time, limitations imposed by having to use a single-valued dependent variable or the need for assumptions about the nature of efficiency in SFE are easily handled using DEA and widely considered one of the strongest attributes of the approach. In both cases more sophisticated estimation techniques are continuously being put forth and their eventual adoption will be the only true test of their success. In the meantime, it is just as necessary to consider what these estimation methods cannot tell us as what they can.

### **3.4 Output and input measures in higher education efficiency studies**

While higher education institutions produce a variety of outputs, nearly all efficiency studies of higher education institutions focus on what Estelle James (1978) refers to as “academic products” (p. 78) of universities, the advancement and transference of knowledge. Translating these tasks into two identifiable outputs presents the standard notions of research and education. However, while straightforward to conceptualize, both resist detailed characterization.

Outputs possess both tangible and intangible aspects (Hopkins and Massy, 1981 cited in Hopkins, 1990) that are often difficult to capture empirically. Cost and production studies of higher education institutions readily acknowledge that the coefficients estimated are distorted by the difficulty in effectively accounting for input and output quality (see King, 1997; Dunder and Lewis, 1995; Nelson and Hevert, 1992; de Groot, et al., 1991; and Cohn, et al., 1989). Unfortunately, lack of consensus on the part of researchers over how to adequately account for quality and the substantial costs, in both time and resources, of obtaining meaningful data has left this issue largely unresolved. This has led many research efforts to follow Nelson and Hevert’s lead of “bowing to tradition” (p. 474) and using traditional measures while simply recognizing that the limitation exists.



There is considerable disagreement across economic studies of higher education institutions as to what is the “best” way to quantify the output “education.” For example, consider two institutions both educating the same number of students where one provides what might be called an “excellent” education while the other provides only a “standard” education. To simply compare the two based on the physical number of “educations” (e.g. FTE enrollments) seriously masks the *effort* put in to educating students and in an efficiency study, the institution educating more students per faculty member will invariably be regarded as more efficient. In other words, one institution may provide an education to many students but not do a good job at it while another institution educates significantly less but puts considerable effort into teaching but the former would be deemed more efficient.

Though researchers suggest that an ideal measure of an institution’s education output attach some institutional “quality weight” to the physical number of students it educates (Nelson and Hevert, 1992), the opportunity costs of obtaining the necessary data are simply too great for it to be feasible. This is evident in nearly all empirical studies of higher education production and costs: education output is almost exclusively proxied by physical headcounts of full time equivalent (FTE) enrollments or number of degrees while recognizing the quality limitation exists.

Of these two commonly used measures, degrees granted is rarely used in studies of higher education efficiency because of several drawbacks that render it highly suspect as a measure of education output. Specifically, it strongly neglects the fact that students who do not complete degrees may still receive one, two, or even three year’s worth of education. To draw on yet another example, the number of degrees granted by an institution in say 2000 would not account for students enrolled but who have not yet finished their studies though they most certainly received a year’s worth of education during the same time period. In this respect, degrees granted does not satisfactorily capture the production of education; it puts a downward bias on the number of students receiving a year’s worth of education at any given time.

The same quantity/quality conundrum is also evident in researchers’ attempts to develop measures of research output. Empirical studies almost exclusively use either publication counts (usually number of journal articles) or research expenditures (usually based on the amount of sponsored research funding). Those advocating the former usually point to the fact that expenditures are actually an input and not an

output of the production process. At the same time, they also suggest that publication counts are preferable because expenditure measures tend to neglect quality aspects and “define away any variation in productivity (de Groot, et. al., 1991, p. 425). On the other hand, those in favor of the research expenditures approach argue that not all research output is in the form of journal articles. Book reviews, plays, musical scores, and patents issued are all viable outputs for certain disciplines and simply choosing one or several can have serious efficiency implications for institutions who do not specialize in those areas. While not all research is based on sponsored funds, most studies employing research income as an output measure defer to Cohn, *et al.* 's (1989) persuasive argument that “the ability of an IHE [institution] to generate such funds is closely correlated with its research output, at least insofar as it is perceived by sponsors” (p. 285).

The ideal output measure, as suggested by Cohn, *et al.*, would involve a weighted measure of all the different research outputs an institution produces. Unfortunately, specifying the weights *a priori* requires value judgments as to the respective worth of any given output and there is no substantive basis in the literature for making such judgments. Johnes and Johnes (1995), for example, showed in their study of economics departments in the United Kingdom that efficiency scores are highly sensitive to the weight assignments given to different publications like journal articles, books, and book reviews.

Specifications of higher education inputs essentially face the same problems with effectively accounting for quality and effort that are evident on the output side though there is less controversy over how they should be quantified. Different input measures exist, but instead of being substitutes for each other (i.e. research income versus publication counts) as in the case for outputs, which input measure is used is a function of what type of efficiency is being assessed: technical efficiency estimates routinely employ physical input units while cost efficiency uses expenditure-based units. In the case of the former, studies tend to use a mixed bag of measures: physical units of academic labor which, in almost all circumstances are measured by FTE faculty numbers, and some cost-based measure of capital, such as library expenditures or physical investment expenditures. Under practically no circumstances is quality controlled for, especially in terms of capital inputs, though in a smattering of studies researchers have made the case that faculty salaries can be used as a rough index of faculty quality (e.g. Dunder & Lewis, 1995).

Where cost efficiency is being assessed, inputs are naturally either measured by total institutional expenditures or by several components of it (e.g. personnel and capital expenditures or academic and research expenditures). In these cases two particular measurement problems arise. First, because accounting practices vary across institutions, what may be regarded as “academic expenditures” by one institution may not be for another. As a result, the more (less) that is included in a particular expenditure category relative to other institutions being analyzed, the more likely it becomes that cost efficiency estimates could be distorted. Second, there is no practical way to index input quality. In other words, two institutions may each spend €70,000 on a faculty member yet a “labor expenditures” input measure will not discriminate between institutions purchasing high, versus average, quality people.

#### **4. A survey of efficiency studies**

To this point we have defined the different types of efficiency, shown how they can be estimated using different techniques, and briefly discussed the different ways in which input and output measures are specified. With this information in mind, we now turn our attention to surveying the various studies of higher education efficiency that have been undertaken over the past decade. In order to develop some uniform criteria from which a broader discussion can be developed in the next section, each study presented addresses, to the extent possible, the following six points:

1. Groups of institutions (or other units) being analyzed
2. Data sources
3. Types of input and output measures used
4. What type of efficiency is being measured
5. The results from the analysis
6. Author(s) own conclusions

There are several ways to group the different analyses. Here they are sorted by the system of institutions (i.e. by country) being evaluated. This was done so that the reader could draw quick comparisons between different evaluations of the same system. A list of the various studies reviewed for this paper is provided in Table 1.

Studies of higher education efficiency are generally done using one of three unit-of-analyses: 1) the institution, 2) the academic department, or 3) non-academic or auxiliary units within institutions. While a number of studies have been carried out using (2) or (3) as the unit-of-analysis, the variety of programs evaluated suggests little would be gained from any comparative analysis. Rather than exclude them completely, the interested reader can find a listing of these studies at the end of this paper in Appendix A.

Table 1: Studies of institution-level efficiency in higher education

STUDY	YEAR	TECHNIQUE EMPLOYED
<b>Australia</b>		
Coelli	1996	DEA
Avkiran	2001	DEA
Abbott & Doucouliagos	2003	DEA
<b>United States</b>		
Ahn, Charnes, & Cooper	1988	DEA
Robst	2001	Stochastic Frontier
Salerno	2002	DEA
<b>United Kingdom</b>		
Athanassopoulos & Shale	1997	DEA
Stevens	2001	Stochastic Frontier
Izadi, Johnes, Oskrochi, & Crouchley	2002	Stochastic Frontier
<b>Canada</b>		
McMillan & Datta	1998	DEA
<b>The Netherlands</b>		
Jongbloed, et al. *	1994	DEA
Jongbloed & Koelman	1996	DEA
Vink *	1997	Stochastic Frontier
Goudriaan, et al.	1998	Stochastic Frontier
Jongbloed & Salerno	2003	DEA

\*These studies also derived efficiency estimates for samples of higher education institutions in Germany and the United Kingdom

#### 4.1 Australia

The earliest study found to assess the efficiency of Australian higher education is that by Coelli (1996). Using 1994 data collected from the *Australian Department of Employment, Education, Training and Youth Affairs* (DEETYA), he formulated three models of university performance for 36 universities: one to evaluate the university as a whole, one to evaluate academic aspects, and one that looked at university administration. For each model, two inputs and two outputs were specified. These are listed in Table 2.

Table 2: Input and output measures used in Coelli (1996) study

	(1)	(2)	(3)
	<b>Universities Model</b>	<b>Academics Model</b>	<b>Administration Model</b>
Outputs	Student numbers <sup>13</sup>	Student numbers	Student numbers
	Publications index	Publications index	Total staff numbers
Inputs	Total staff numbers	Academic staff Numbers	Administration Staff (AU\$)
	Non-staff expenses	Other expenses <sup>14</sup>	Other administration Expenses

For each model a variable returns to scale (VRS) DEA analysis was conducted. A secondary analysis on the overall performance model disaggregated technical efficiency in order to also estimate scale efficiency. Finally, a sensitivity analysis using modified input and output measures was also done.

He reported a mean technical efficiency score for the university model of 95.2% and a mean scale efficiency of 96.6%. When the sources of scale inefficiency were explored, he found the overwhelming majority of inefficient institutions (21 of 26) were operating at decreasing returns to scale (DRS). In the academics model, the mean technical and scale efficiency scores were found to be slightly lower than the universities model, 92.6% and 93.4% respectively. Nineteen scale inefficient institutions were shown to be operating at DRS and nine (9) institutions at increasing returns to scale (IRS). In the third model (administration) the mean technical efficiency was shown to be markedly lower (87%) and the mean scale efficiency to be slightly lower (94.4%) compared to the other two models.

<sup>13</sup> Enrollments measured in terms of Equivalent Full-Time Student Units (EFTSU).

<sup>14</sup> Includes academic support expenditures like library and computing services as well as expenditures on support staff like research assistants and technical staff.

Because Coelli's study was an internal report for the University of New England (UNE), most of his conclusions expectedly center on the performance of UNE. However, based on the broader findings from the study he did suggest that, "the majority of universities appear to be achieving high degrees of technical and scale efficiency" (p. 17). He also notes that some universities were found to be relatively inefficient in more than one model, which is worthy of further analysis.

Avkiran (2001) conducted a DEA analysis of 36 Australian universities based on 1995 data collected from DEETYA. Three separate performance models were estimated: 1) overall, 2) delivery of educational services, and 3) performance on fee-paying enrollments. All three models used the same two input measures (FTE academic and non-academic staff). The output measures used in each model are listed in Table 3.

For each model a variable returns to scale (VRS) DEA analysis was conducted. A secondary analysis was then done on the overall performance model that disaggregated technical efficiency in order to also estimate the scale efficiency.

His results showed a mean efficiency score of 95.5% (S.D. = 10.1%) for the overall model, 96.7% (S.D. = 3.5%) on the delivery of services and a mean efficiency of only 63.4% (S.D. = 29.1%) in the fee-paying enrollments model. In a secondary analysis of the overall model mean scale efficiency was shown to be 94.2% (S.D. = 8.3%). Of the 23 universities found to be scale inefficient, all but 4 were shown to be operating at decreasing returns to scale (DRS).

While no benchmark is established, based on the results from the first two models Avkiran claims that Australian universities are operating at "respectable" (p. 71) levels of efficiency. In terms of the fee-paying model, he concludes that the relatively low mean efficiency score and high standard deviation are both evidence of poor performance in attracting fee-paying students. At the same time, he does state that such a low finding was expected given the recent introduction of fees for Australian nationals enrolling in postgraduate study.

Table 3: Output measures used in Avkiran (2001) study

(1) Overall Model	(2) Educational Services Model	(3) Fee-paying Enrollments Model
Undergraduate Enrollments	Student Retention Rate (%)	Overseas Fee-paying Enrollments
Postgraduate Enrollments	Student Progress Rate (%)	Non-overseas Fee-paying Postgraduate Enrollments
Research Quantum	Graduate Full-time Employment Rate (%)	

Abbott and Doucouliagos (2003) did a third study on the efficiency of Australian universities, again using 1995 data collected from DEETYA. In total they developed and present findings for four DEA models. Two were conducted using all 36 institutions and two truncated samples were also analyzed. For the latter, analysis groups were constructed based on the ratio of each institution's research to teaching output (i.e. output mix). Those with the lowest (and highest) values were then grouped for analysis. Four input measures were specified: 1) FTE academic and 2) non-academic staff, 3) expenditures on all non-labor inputs, and 4) the value of non-current assets to approximate existing capital stock. In terms of outputs they proxy education by EFTS enrollments and research by the Research Quantum.<sup>15</sup> In the case of the output-ratio groupings, only two inputs were used: the combined FTE staff and expenditures on non-labor inputs. For each model a VRS DEA analysis was performed and estimates of scale efficiency calculated.

In the results from their first model (all 36 universities and using the Research Quantum as a measure of research output) they found a mean efficiency score of 94.6% though they do not report the standard deviation. They also report a mean scale efficiency score of 96.7%. The second model, in which the Research Quantum was replaced by the variable "medical and non-medical research income," the mean

<sup>15</sup> In the paper the authors indicate that in the course of their study they performed multiple analyses using different measures of research and education output but that the results did not vary considerably.



efficiency score they reported was slightly higher (96.7%) but mean scale efficiency did not change.<sup>16</sup> In the second set of DEA analyses based on output mix the mean efficiency score for the low-ratio universities was 96.4% and for the high-ratio it was 93%. The scale efficiency scores were shown to be similar to the first model at 96.9% and 94.6% respectively.

By and large the authors suggest that, overall, efficiency appears high and that “Australian universities are performing very well” (p. 96). No implications for higher education policy are offered, though they do admit improvements in efficiency cannot be ruled out. At the same time they conclude that there is a high degree of homogeneity in the system yet, because DEA only provides relative efficiency scores, it may be the case that the entire system is under-performing. Finally, in light of the fact that Australia competes heavily for overseas students, they state that no conclusions can be drawn about how efficient Australian universities are compared to institutions in other systems.

## 4.2 United States

The first contemporary study seeking to evaluate the efficiency of US higher education institutions was that by Ahn, Charnes, and Cooper (1988). The estimated technical and scale efficiency for 161 public and private doctoral-granting institutions, while controlling for the presence of medical schools in 1984-85. In total 12 models were estimated. Institutions were first grouped according to whether or not they had medical colleges and then subsequently divided into groups of public and privates. The data used came from the National Center for Education Statistics (NCES). Three cost-based input measures were used: 1) instructional expenditures, 2) physical investments, and 3) overhead expenditures. The three output measures they included were FTE undergraduate and graduate enrollments and federal research grants expenditures as a proxy for research output.

When results were compared by institutional control, they found public institutions without medical schools to be more technically efficient than their private counterparts (mean = 70% and 64% respectively). Roughly the same number of publics and privates were found to be fully efficient, yet given that there were approximately twice as many public institutions it would seem that privates were more likely to construct part of the efficient frontier than publics. When comparing institutions with

---

<sup>16</sup> This is not to say that each institution had the same scale efficiency score in both models. In most cases, the scores for individual institutions differed, albeit marginally.

medical schools, they found the same amount of dispersion between mean efficiency scores, though the nominal values were notably higher (mean = 84% and 77% respectively). However, when testing whether the mean scores differ for the two types of institutions, they were only able to verify differences at the  $\alpha = .10$  significance level. In the analysis of scale efficiencies they estimated a mean efficiency score for universities without medical facilities of 65% and a mean score for universities with medical facilities of approximately 79%.

Based on their analysis, the authors conclude that the mean efficiency scores for both clusters of universities are “certainly worthy of further research including a search for possible interventions” (p. 267). Because the main goal of the study was to test hypotheses about public versus private sector behavior, no policy implications were offered. In addition they note that their study only focuses on a single year and that such a study would stand to benefit from a multi-year analysis.

Robst (2001) conducted an institutional cost efficiency study on 440 public colleges and universities in the US between 1991 and 1995. Using data from the NCES’ Integrated Post-Secondary Education Data System (IPEDS) database, he estimated a series of four trans-log SFE cost functions. Different models were run based on different specifications of where inefficiencies were expected to emerge.<sup>17</sup> The dependent variable in his model was education and general expenditures (note: all cost data in the study was expressed in thousands of US dollars). The explanatory variables he specified included three outputs (FTE undergraduate and graduate enrollments and research expenditures), a dummy variable for Carnegie Classification status, compensation (as an input price), and two measures of institutional revenue (tuition plus state appropriations and the percentage of state appropriations to total institutional revenue).

In the first model he found a positive relationship between university revenues and inefficiency ( $\beta = .1837$ ). This same relationship was also evident in the second model though here he also found that increases in both graduate enrollments ( $\beta = 4.525$ ) and research expenditures ( $\beta = .5882$ ) were positively related to cost inefficiency. At the same time, inefficiency was shown to decline with increases in undergraduate enrollments ( $\beta = -6.000$ ). He was not able to identify any statistically

---

<sup>17</sup> Model (1) assumed inefficiency emerged in the revenue institutions received. Model (2) further allows for inefficiencies to emerge in the production of outputs. Finally, model (3) allows for inefficiencies to arise based on type of institution.

significant relationships between institution types and inefficiency. Robst also tested whether *cost inefficiency* was likely to vary over time. Using state share of revenue as the specification variable he found no statistically significant relationship during 1991 and 1992 but did find significantly negative relationships for 1993-95, suggesting that where state appropriations increased, efficiency rose. Though the study sought to understand the relationship between public funding and cost efficiency, where such findings certainly have relevance in political arenas, he provides no further discussion on the implications of the findings.

The most recent study of higher education efficiency in the US is that by Salerno (2002). He used DEA to assess the relative efficiency of 183 research and doctoral granting institutions in 1993. The data used came from several sources, including the National Science Foundation (NSF), IPEDS, and the Institute for Scientific Information's citation indexes. Three different models were estimated. Institutions were first grouped into two quality tiers<sup>18</sup> and separate technical efficiency analyses were conducted. Both used the same input and output measures and were estimated using a VRS DEA model. Scale efficiency was estimated and he also examined whether efficiency scores differed by type of institutional control and the presence of medical facilities. The third model was a constant returns to scale (CRS) DEA analysis of 35 public universities from the 183 from which estimates were made of overall, technical, and allocative efficiency. Three labor inputs were specified (all FTEs): 1) faculty members, 2) graduate teaching assistants, and 3) graduate research assistants. On the output side, he included three education outputs (lower-level undergraduates, upper-level undergraduates, and graduate students) measured by FTE enrollments and publication counts as a proxy for research. In the cost analysis, input prices were also included.

From the technical efficiency analyses he found mean efficiency scores of 93% (S.D. = 9.8%) for the high-quality tier and 86% (S.D. = 12.8%) for the other tier. In terms of scale efficiency, the mean scores for each tier were 95% (S.D. = 6.5%) and 90.6% (S.D. = 11%) respectively. For both the technical and scale efficiency scores, the difference between tiers was shown to be statistically significant. Further analysis of the scale efficiency measures showed that inefficient institutions were more likely to be operating at increasing rather than decreasing returns to scale. Form of institutional control was shown to not affect technical efficiency scores. However, when the presence of medical

---

<sup>18</sup> The high quality tier (Tier 1) consisted of 68 universities and the other (Tier 2) of the remaining 115.

facilities was considered,<sup>19</sup> institutions without medical facilities were found to be almost twice as likely to be considered technically efficient than those with. In the subsequent analysis of scale efficiency, publics were shown to be just as likely to be regarded as scale efficient as privates in both tiers. Again though, when the presence of medical facilities was considered, only one medical institution in Tier 2 was classified as scale efficient (3% of all medical institutions) versus 16 without (20% of all nonmedical institutions).

In the cost analysis of the 35 public institutions, he found a mean overall efficiency score of 76.9% (S.D. = 11.1%). When overall efficiency was disaggregated into technical and allocative components, he estimated mean efficiency scores of 88.7% (S.D. = 10.9%) and 86.9% (S.D. = 10.1%) respectively suggesting only about half of the overall inefficiency was due to technical inefficiency.

The author does not address whether the mean efficiency scores he calculated can be regarded as necessarily high or low though he does maintain that the findings are consistent with economic theories of university behavior. While no policy implications are derived, he does suggest that the scale inefficiency found is likely due to the rapid expansion of academic research in the 1980s. The main conclusion he reaches is that input quality and competition positively influence productive efficiency and that public and private research universities should be analyzed jointly in such circumstances.

### **4.3 United Kingdom**

Athanassopoulos, and Shale (1997) have used DEA to evaluate the efficiency of 45 “old” universities<sup>20</sup> in the United Kingdom during 1992-93. Data was collected from several sources including the 1992 Research Assessment Exercise (RAE) and publications by the Universities’ Statistical Record. Two general models were estimated, one seeking to estimate cost efficiency and another to estimate outcome efficiency. Both models included the same three output measures: 1) numbers of successful leavers, 2) number of higher degrees awarded, and 3) weighted research rating. The different inputs specified for each model are listed in Table 4.

---

<sup>19</sup> Faculty and students in medical programs were excluded from the input and output measures. Any differences were therefore assumed to arise simply from the presence of medical facilities on campus.

<sup>20</sup> The use of the term “old” here is meant to distinguish these institutions from the many higher education institutions who were officially given university status after 1992.

In the cost efficiency model, separate CRS and VRS DEA analyses were performed for institutions that were grouped according to whether they were science-only (N=12) or balanced universities (N=25). A third set of analyses was also done that included all 45 institutions. For the outcome efficiency model, again CRS and VRS efficiency scores were estimated, though the sample was not disaggregated. In addition, three additional analyses were done to show how efficiency scores would change by imposing different sets of value judgments<sup>21</sup> on the input and output measures.

Table 4: Input measures used in Athanassopoulos & Shale's (1997) study

(1)	(2)
Cost Efficiency Model	Outcome Efficiency Model
General Academic Expenditures	FTE Undergraduates
Research Income	FTE Postgraduates
	FTE Academic Staff
	Mean A-level entry scores
	Research Income
	Expenditures on Library and Computing

In the cost efficiency analysis the science universities were shown to have higher mean efficiency scores compared to the balanced universities under both CRS and VRS (in science, 90.4% and 95.4%, and in balanced = 81.1% and 88.3% respectively). However, since there were only 12 universities in the science sample compared to 25 in the balanced sample, it may be the case that the higher efficiency mean efficiency scores in the former were due to comparing significantly fewer institutions (Zhang and

<sup>21</sup> Value judgments involve placing restrictions on the shape of the constructed frontier like the one presented in Figure 5. For example, it may be the case that  $X_2$  in Figure 5 should never be given less weight than  $X_1$ . This type of value judgment, or weight restriction as it is referred to in the literature, would be equivalent to not letting the DEA program construct the QA part of the isoquant in Figure 5. See Pedraja-Chaparro, *et al.* (1997) for a good review of weight restrictions in DEA.

Bartels, 1998). When all institutions were considered together the mean CRS efficiency was shown to be 71.4% and mean VRS efficiency 83.1%.

The findings from the outcome model without any value judgments showed relatively high mean CRS and VRS efficiency scores (97.2% and 98.2%) and a disproportionately large number of institutions were found to be fully efficient: 27 of the 45 in the CRS model and 31 in the VRS model.<sup>22</sup> The results from the three different value judgment models all provide lower mean efficiency estimates than the value-free model. The lowest reported score was 93.8% (value model one) and the highest was 96% (value model three). In the same way that mean efficiency declined when value judgments were incorporated, so too did the number of institutions found to be fully efficient. Only 14 universities were 100% efficient in the first value model, 19 in the second, and 23 in the third.<sup>23</sup>

The authors make no judgment in the paper as to whether the estimated efficiency scores are relatively high or low. In their conclusions one of the key findings they point to from their study is that cost efficient universities producing high output levels do not generally equate to lower unit costs. Their other main finding is that many inefficient universities were particularly “over-resourced” in the process of producing research. From this they question whether directing resources for research based on the RAE exercise maximizes value added from additional funding.

The most recent effort to estimate the efficiency of UK higher education institutions is that by Izadi, Johnes, Oskrochi, and Crouchley (2002). Based on an earlier study by Johnes (1997), they estimate a constant elasticity of substitution (CES) SFE cost function for 99 universities during 1994-95. Two non-linear maximum likelihood (ML) models were estimated (with and without an efficiency component). The data came from several publications produced by the Higher Education Statistical Agency (HESA) in the UK. The dependent variable in the analysis was total institutional expenditures, which was regressed against four explanatory variables: 1) undergraduate student load in arts subjects, 2) undergraduate student load in science subjects, 3) postgraduate student load, and 4) value of research grants and contracts received.

---

<sup>22</sup> By construction, VRS DEA models can never generate efficiency scores lower than that in a CRS DEA model.

<sup>23</sup> The complexity behind the different value judgment schemes would be quite detailed and beyond the scope of this study. It is sufficient here to note that incorporating such restrictions invariably reduces efficiency and that such reductions are a function of the number and severity of the restrictions.

In broadly comparing the fit of the two models they were not able to reject their null hypothesis at the 5% level, suggesting that the SFE estimator did not fit the data better than the standard ML estimator. As such, the authors caution against putting too much emphasis on the technical efficiency estimates derived and not they are likely to have high standard errors. Based on their reported results, average efficiency for institutions in their sample was 87.6% (S.D. = 9.8%). Forty five percent (45%) of the sample were shown to be operating at or above 90% efficiency and nearly 88% above 80% efficiency. No effort was made, however, to identify the sources of potential inefficiencies.

In terms of whether their findings are particularly high or low, the authors conclude that the inefficiency identified in their study is “fairly modest” and “on the margins of statistical significance” (p. 70). It is suggested though that it may be possible to realize further efficiency gains by performing additional benchmarking studies. In terms of policy implications, they suggest that increases in efficiency stemming from funding councils’ funding mechanisms are likely to be limited.

Arguably the most comprehensive effort at assessing efficiency in UK higher education is that by Stevens (2001), whose study specifically sought to not only estimate efficiency but also the sources of potential inefficiencies. Using data from HESA and the *Times Higher Education Supplement*, he estimated a series of quadratic SFE cost functions from 1995 to 1998 for 80 universities in England and Wales. Three different models were estimated, each of which presents a more sophisticated treatment of input and output quality. All three models employed the same dependent variable (total expenditures), three education outputs (science undergraduates, arts undergraduates, postgraduates), one research output (research income), and one input price (average staff costs in 1995). Two quality variables were also included: 1) average A-Level score requirement for each institution, and 2) the percentage, per institution, of students achieving firsts and upper seconds. Finally all three models incorporated a vector of 13 characteristic variables, as well as a time trend, to explain the sources of potential inefficiencies. These characteristics are listed in Table 5.

In all three models, the amount of random error attributable to efficiency was found to be statistically significant, which suggests that economies of scale and scope estimates in traditional cost functions may be biased. In the simplest model 63.6% of the total variance in the error term could be attributed to inefficiency while in the most comprehensive model it was as high as 81.2%.

Table 5: Variables in Stevens (2001) used to identify sources of IHE inefficiency

FTE Staff (proportion of)	Students (proportion of)
Staff aged > 50	Students aged >= 25
Female	Female
Non-white	Non-white
Professors	Students from lower SES <sup>24</sup>
Senior Lecturers	Other EU Countries
RAE Active	Non-EU Countries
	In Arts Disciplines

Based on the first two models (the simpler quality specifications), when the sources of inefficiency were examined the parameter on the time trend was shown to be statistically significant in all three models, indicating that efficiency progressively increased during the time period. Estimates of staff over fifty, non-white, and female suggest increases in these categories resulted in lower institutional efficiency. At the same time, higher proportions of professors, senior lecturers, and research active staff are shown to be positively related to efficiency. In terms of student characteristics, institutional efficiency is shown to be higher as the proportions of students older than 25 and from lower social classes increase. In addition, efficiency is also shown to rise where there are large proportions of arts students. Finally, efficiency is also shown to be positively related to the number of students receiving first and upper-second degrees (FIRST). In the most sophisticated model, which takes into account both input and output quality as well as interactive effects, the parameter on FIRST remained statistically significant but the sign changed indicating that, “universities that produce a large number of high-achieving students tend to be less efficient, when one accounts for the direct effect of teaching quality on costs” (p. 22).

<sup>24</sup> SES (Socio-economic Status). Stevens uses the term “lower classes,” which he defines as students having their prime parent or guardian employed in the Standard Occupational Classification groups 8 (plant and machine operatives) and 9 (other occupations), or who are unemployed.



Changes in inefficiency over time were also examined. His findings showed that changes in inefficiency depended negatively on initial inefficiency. In other words, institutions having high levels of inefficiency in 1995 became more efficient more rapidly than those with less inefficiency in 1995, though average increase in efficiency scores was shown to be approximately 6%. He does note, however, that a fairly large adjustment occurred in 1997, which was the last year before tuition fees were introduced. Finally, the dispersion of efficiency scores (as measured by the standard deviations for the efficiency estimates) was shown to decline in every time period suggesting that overall cost efficiency has become less variable among English and Welsh universities.

The main policy implication Stevens reached was that the introduction of tuition fees may have led to a “shake-up” (p. 25) that motivated less efficient universities to become more efficient. Though the mean efficiency score in all three models was approximately 81% no judgment was made as to whether these scores could be characterized as exceptionally high or low.

#### **4.4 Canada**

McMillan and Datta (1998) used DEA to estimate the efficiency of 45 Canadian universities in 1992-93. The data they used came from the Canadian Association of University Business Officers (CAUBO) and the Association of Universities and Colleges of Canada (AUCC). A series of 9 DEA models was estimated, three of which examined cost efficiency. The models were formulated using different combinations of the aggregate input and output measures they specified. This was done in order to evaluate the sensitivity of their findings. Table 6 lists the input and output measures used for each model. In each specification a VRS DEA analysis was performed on all 45 universities, though the results were reported by grouping institutions into one of three categories: 1) 15 comprehensive universities with medical schools (CUMED), 2) 11 comprehensive universities without medical schools (CUnoMED), and 3) 19 primarily undergraduate universities (UGU).

The findings from the DEA analyses showed the mean efficiency scores for the CUMED group varied between 91% (models 7 and 8) and 98% (model 5) with an *average* mean efficiency score of 94%. For the CUnoMED universities, mean efficiency was shown to vary between 91% (model 7) and 97% (models 5 and 6) with an average of 95%. In the UGU group the highest mean efficiency score was 98% (model 5) and the lowest was 89% (models 2 and 7) with an average mean of 93%. In analyzing

scale inefficiency they found average scale efficiency in the UGU group to be 94.3% and for the combined comprehensive universities, 98%.<sup>25</sup>

In the authors’ own interpretation of their findings they claim that efficiency scores are, overall, relatively high, though they do point out that aspects of input and output quality are poorly accounted for and that efficiency may be artificially high due to the small sample size. They also suggest that there is evidence to support the idea that geographic competition, program specialization, and total enrollment levels may increase efficiency. Finally, while the findings are shown to be stable across different input/output combinations, models (3) – (5) and (7) – (9) are regarded to be the most suitable choices for evaluating higher education institutions.

Table 6: Input/Output measures & models in McMillan & Datta (1998) study<sup>26</sup>

	Model								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<b>Outputs</b>									
<i>Undergraduate Teaching</i>									
UG	x	x	x				x		
UG in Sciences				x	x	x		x	x
UG in Other				x	x	x		x	x
<i>Graduate Teaching</i>									
Graduate Students	x								
Master's		x	x	x	x	x	x	x	x
Doctoral		x	x	x	x	x	x	x	x
<i>Research</i>									
Research Income	x	x	x	x	x		x	x	x
%SSHCC					x	x			x
%MRCNSE					x	x			x

<sup>25</sup> In addition to the DEA analysis a secondary regression analysis was also done in order to explore the causes of potential inefficiencies. The findings are not reported here as the authors indicate they were “relatively unsuccessfully in identifying further determinants of inefficiency” (p. 508).

<sup>26</sup> SSHCC is the Social Sciences and Humanities Research Council and MRCNSE is a composite of the Medical Research Council (MRC) and the Natural Sciences and Engineering Research Council (NSERC). Enrollments are measured in FTEs.

	Model								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<b>Inputs</b>									
<i>Faculty</i>									
Total Faculty	x	x							
Faculty in Sciences			x	x	x	x			
Faculty in Other			x	x	x	x			
<i>Other</i>									
Other expenditures	x	x	x	x	x	x			
<i>Financial Input (only)</i>									
Total Expenditures							x	x	x

#### 4.5 The Netherlands

Jongbloed *et al.* (1994) used DEA to derive relative cost efficiency estimates for universities and professional education institutions in the Netherlands and Germany, and for universities in the UK using data from the 1989 to 1991.<sup>27</sup> Because it was not possible to derive a set of uniform input and output measures across all three countries, universities were only compared to others in their own country. In addition, universities in the Netherlands and UK were disaggregated into arts, sciences, and medical clusters to increase the homogeneity of the samples. For Germany, universities were analyzed together and then separately evaluated based on those with and without medical schools. Data for each country came from multiple sources, though particularly from national statistical agencies.<sup>28</sup> For each country multiple analyses were conducted using various input and output combinations. Here we only report the findings from the basic models specified.

In the analysis of each country two input measures were used, labor and material expenditures. For outputs, the measures differed due to the availability of data. For Dutch universities a single education

<sup>27</sup> Different years reflect the availability of data in different countries. Thus, Germany was evaluated using 1989 data whereas the UK data was from 1990 and for the Netherlands, 1991.

<sup>28</sup> In the Netherlands, data sources included the Central Bureau of Statistics, the Ministry of Education, Culture, and Science (OCW), the Association of Universities in the Netherlands (VSNU) and institutions' own financial statements. For Germany, data was taken from the German Statistical Bureau (Statistisches Bundesamt or StBA). Data for the UK came from the Department for Education (DfE) England, the Universities' Statistical Record, and the Higher Education Funding Council for England (HEFCE).

output variable was constructed by weighting full-time, part-time, and auditing students while research outputs consisted of the numbers of dissertations and all other publications. For Dutch professional education institutions (HBOs), the same two inputs were used and the two outputs specified were full-time and part-time students. Research was not included as an output in this sector though research income was subtracted from institutional expenditures to avoid double counting. For the UK, two education outputs (FTE undergraduates and postgraduates) and one research output (income from grants and contracts) were used. For Germany, education and research outputs in universities were measured by numbers of students (not FTE) and what is referred to in Germany as *Drittmittel*.<sup>29</sup> Inputs and outputs for the professional education institutions (*Fachhochschulen*) were specified in the same way as the Dutch HBOs. For each set of institutions a VRS DEA analysis was conducted and measures of scale efficiency and their source were calculated.

The findings for the Netherlands<sup>30</sup> showed university arts clusters to be, on average, 97% cost efficient and 96% scale efficient with half of the scale inefficient, the source of the latter being increasing returns to scale (IRS) for all inefficient clusters. In the case of science clusters, all were shown to be 100% efficient under VRS and 99% scale efficient. Of the scale inefficient clusters, one was shown to be operating at decreasing returns to scale (DRS) and three at IRS. For the medical clusters, mean cost and scale efficiency were both shown to be 96%. Where scale inefficiencies were present, three were operating at IRS and one at DRS.

For the HBO sector,<sup>31</sup> when all institutions were analyzed together mean efficiency was found to be 91% (S.D. = 11%) and 50% of the institutions were shown to be fully efficient. Mean scale efficiency was 92% with approximately the same number of scale inefficient institutions operating at DRS and IRS. When only “arts” HBOs were analyzed mean efficiency was found to be 90% (S.D. = 13%) and mean scale efficiency to be 97%. In the analysis of general HBOs, mean efficiency was 96% (S.D. = 7%) and over 60% were reported to be fully efficient. Mean scale efficiency was shown to be 93% though the majority of scale efficient institutions were operating at DRS (70% of institutions were reported to be scale inefficient of which only 9% were operating at IRS). The analysis of social science

---

<sup>29</sup> Non-core research funding from government, research enterprises, and private foundations.

<sup>30</sup> The sizes of the different university clusters were as follows: Arts = 10, Sciences = 10, and Medicine = 8.

institutions showed a mean efficiency score of 88% (S.D. = 12%) and mean scale efficiency of 91%. In contrast to the findings for the general HBOs, the preponderance of scale inefficiency was attributable to institutions operating at IRS (71% of all institutions).

For the UK universities,<sup>32</sup> the mean cost (scale) efficiency score for arts clusters was 91% (98%). When the sources of scale inefficiencies were evaluated, 24% of the arts clusters were shown to be operating at IRS and 48% operating at DRS. For the sciences clusters, the reported mean cost (scale) efficiency was 94% (89%). A significant number of clusters here were shown to be operating at DRS (70%) while only 11% were shown to be operating at IRS. In the UK medical cluster, mean cost efficiency was reported to be 94% while mean scale efficiency was slightly higher at 96%. Again a large number of medical clusters were shown to be operating at DRS (58%) and only a relatively small number at IRS (10%).

When all German universities<sup>33</sup> were analyzed together, the findings indicated a mean cost efficiency of 91% (S.D. = 13%). While mean scale efficiency was reported to be zero, analysis of the individual institutions showed 30% to be operating at DRS and 18% of all institutions operating at IRS. When only institutions with medical schools were evaluated the mean cost efficiency was shown to be slightly higher at 93% (S.D. = 12%). Mean scale efficiency was found to be 94%; just over half of the universities (53%) were shown to be operating at IRS and 20% at DRS. Finally, in the analysis of the non-medical institutions, mean cost efficiency was found to be 86% (S.D. = 18%) and mean scale efficiency was almost 99%. Here, 44% of the cluster was operating at DRS and only 22% at IRS.

In the overall analysis of the German *Fachhochschulen*,<sup>34</sup> mean cost efficiency was found to be 83% (S.D. = 17%) and mean scale efficiency was 93%. Most institutions were found to be operating at DRS (75%). When specialized institutions were removed, analysis of the remaining “general” institutions showed mean efficiency rose to 89% (S.D. = 13%) while mean scale efficiency remained constant

---

<sup>31</sup> The sizes of the different HBO clusters were as follows: Arts = 7, General = 23, Social Science (gamma) = 30, and Total = 60.

<sup>32</sup> The numbers of institutions represented in each cluster are as follows: Arts = 58, Science = 63, and Medical = 31.

<sup>33</sup> The numbers of institutions in the different clusters are as follows: All = 33, Medical = 15, and non-Medical = 18.

<sup>34</sup> The total number of *Fachhochschulen* was 48. The number of “general” institutions, those left after removing specialty providers, was 33.

(93%). The sources of inefficiency, however, were shown to change mildly. The proportion of institutions operating at DRS declined to approximately 61% while the numbers of scale efficient and IRS institutions rose to 21% and 18% respectively (compared to 13% shares for each in the larger sample).<sup>35</sup>

In another study, Vink (1997) estimated a series of SFE cost functions as part of a comparative analysis of higher education costs in the Netherlands, Germany, and the United Kingdom in 1990.<sup>36</sup> The data he obtained for each country came primarily from the same sources as that in Jongbloed, *et al.* (1994). The dependent variable in each analysis was current costs, which was regressed against three primary variables: 1) enrollments, 2) publication counts, and 3) an index based on average personnel costs per FTE staff member. In alternative models both the enrollment and publication variables were disaggregated into arts and science clusters. He also included variables for the percentage of full-time and post-graduate enrollments, the percentage of students and publications in sciences programs and journals, and a dummy variable for the presence of academic hospitals. Finally a sensitivity analysis was done by publication counts with research income and enrollments with degrees granted.

Vink estimated no less than 24 SFE cost functions using a variety of explanatory variable combinations and functional forms. Because such a large number of estimates were reported and no overall, or main model was specified I do not report the results here. Instead I report his general conclusions and encourage the interested reader to give close attention to the fifth chapter in his book.

In his conclusions, he found a positive correlation between enrollments and cost efficiency (.324) for the HBO sector. A similar finding was shown between enrollments and efficiency (.506) and between publications output and efficiency (.363) in the analysis of UK universities. When scale efficiency was evaluated, the findings indicated that in both the UK and Dutch HBO sectors, average cost efficiency was higher where institutions had above average enrollments. In addition, he also found scattered

---

<sup>35</sup> Jongbloed and Koelman (1996) conducted a small, follow-up study in which they examined changes in efficiency between 1990 and 1992.

<sup>36</sup> The samples he used for each country included: 1) 70 HBO institutions in the Netherlands, 2) 20 German universities from the Federal states of Bavaria and Lower Saxony and 18 *Fachhochschulen*, and 3) 73 Universities in the United Kingdom.

evidence to support the idea that institutions only producing output in single clusters were more likely to have higher cost efficiency scores than those producing in multi-cluster institutions.

Interestingly, though he did not estimate an SFE cost function, Vink did estimate a “traditional” cost function using a pooled cross-national sample of data. While the findings themselves are not particularly relevant (as no accounting is made of institutional efficiency), he did explore whether the different national samples could be pooled for a joint analysis using analysis of variance techniques. His findings suggest that the HBO and *Fachhochschulen* sectors could be analyzed jointly and there is partial evidence to support comparisons between the Dutch and German universities and between the German and UK universities (p. 218).

Goudriaan, Jongbloed, and van Ingen (1998) estimated a series of translog SFE cost functions for both the university (N = 13) and HBO (N = 54) sectors for the period 1990 to 1994. The data they used came from the same sources as that in Jongbloed, *et al.* (1994). The dependent variable in the university and HBO analyses was institution’s variable costs, which was regressed against a number of output variables. These are listed in Table 7. In addition a sensitivity analysis was done on both sectors using modified explanatory variables.

The findings from the cost function estimations indicated a high proportion of the total variance was due to inefficiency: 75.5% in universities and 53.7% in HBOs. From the results of the university analysis, mean cost efficiency across the entire time period was shown to be 93%, with the highest mean in 1991 (95%) and the lowest in 1992 and 1994 (92%). No institutions were shown to be operating below 84%. For the HBO sector mean efficiency was reported to be 92% and fairly constant across all years. Compared to the university findings the dispersion of efficiency was shown to be greater in the HBO sector; 4% of the institutions had efficiency scores below 80%. When HBOs were grouped according to whether they were single- or multi-cluster institutions, mean cost efficiency was shown to be 91% and 92% respectively. At the same time, the dispersion of efficiency scores was slightly greater in the single-cluster group, where the minimum efficiency score (76%) was nearly 5% lower than in the multi-sector institutions (83%). Finally, estimates of scale efficiency for the two sectors suggest the average university in 1994 was scale efficient while the average HBO was shown to be slightly larger than optimal size (97% scale efficient).

Table 7: Explanatory variables used in Goudriaan, et al. (1998) study

University Model	HBO Model
Hedonic <sup>37</sup> education output	Hedonic education output
% of enrollments in beta and medical programs	% of enrollments in education programs
% of full-time enrollments	% of enrollments in health-related programs
Hedonic research output	% of enrollments in technical-type programs
% of publications generated by 2nd stream funding	% of enrollments in social science programs
% of publications generated by 3rd stream funding	% of part-time enrollments  contract activities

The main conclusion reached from their analysis is that average cost efficiency in HBOs was relatively high compared to other countries and that university efficiency was shown to be comparable with similar estimates of mean cost efficiency for British and German universities reported by Vink (1997). The authors do caution however, that the small sample size in the university sector and no accounting for other outputs than teaching and research was likely to bias the results. In terms of policy findings they suggest that contract activities in the HBO sector were, on average, not profitable.

The most recent effort at assessing the efficiency of higher education institutions in the Netherlands is that of Jongbloed and Salerno (2003). In that study they used the used DEA to allocate shared expenditures to education in order to derive a series of per-student cost estimates for both the university and higher professional institution (HBO) sectors. Similar to the Jongbloed et al (1994) study, each university was subdivided into arts, sciences, and medical program clusters. In the HBO sector, where many institutions only provide education in narrow disciplinary fields, institutions were grouped into

<sup>37</sup> The hedonic variables correct for the composition of student bodies (i.e. the numbers of students in different program clusters).



one of seven areas depending on which programs were offered.<sup>38</sup> The two inputs that were used were personnel expenditures and non-personnel expenditures. In terms of outputs, for the HBO institutions education was approximated by FTE enrollments and research by the amount of third-stream funding<sup>39</sup> the institution received. In the university clusters, education was approximated by nominal enrollments (data was not available to create FTEs) while research was proxied by the number of FTE researchers per cluster. For each subgroup of HBOs and universities, a CRS “shared resources” DEA model<sup>40</sup> (Mar-Molinero, 1996) was estimated for the years 1996 to 2001.<sup>41</sup> In each year and for each group, three cost efficiency measures were calculated, including the efficiency at: 1) producing education, 2) producing all non-education outputs, and 3) producing all outputs.

Based on the university analyses, arts clusters in 2000 were shown to be, on average, 81.5% (S.D. = 18.8%) cost efficient at producing education and 80.3% (S.D. = 13.3%) cost efficient overall. For the sciences cluster, mean cost efficiency in the provision of education and overall were shown to be 75.2% (S.D. = 22%) and 70.5% (S.D. 70.5%) respectively. In both the arts and sciences clusters, mean overall cost efficiency was shown to decline over the period by approximately 6% in each cluster. However, while the cost efficiency at producing education was also shown to decline in the arts cluster (from 87.2% in 1996), in the sciences cluster education efficiency rose slightly (from 72.3% in 1996).

For the HBO sector, the mean cost efficiency of producing education in arts institutions during 2001 was reported to be 80.5% (S.D. = 15.9%) and overall cost efficiency was shown to be 80.4% (S.D. = 15.8%). In both cases mean cost efficiency increased by approximately 3.5% between 1996 and 2001. In the “technical” HBO institutions mean efficiency scores (education and overall) were 86.1% (S.D. = 12.8%) and 83.5% (S.D. = 7.7%) respectively and the scores remained stable over time. The mean cost efficiency of education and overall for institutions specializing in education during 2001 was

---

<sup>38</sup> The HBO institutions were clustered into one of the following categories: 1) arts institutions, 2) education institutions, 3) social science (i.e. economics) institutions, 4) institutions offering technical programs, 5) institutions offering courses in all areas (5-cluster), 6) institutions offering courses in all but arts areas (4-cluster), and 7) institutions offering programs in social science, education, and technical programs.

<sup>39</sup> Third-stream funding is funding received for contract research, which does not include funding from national research councils and government appropriations.

<sup>40</sup> The model was estimated under the assumption of CRS only because the sample size was too small to derive efficiency estimates under VRS.

<sup>41</sup> For the university sector data was only available up to and including 2000.

86.6% (S.D. = 9.5%) and 82.1% (S.D. = 7%). In addition, the mean efficiency scores in both cases increased by approximately 10% from 1996 values.

In HBO institutions offering the widest array of programs (5-cluster) mean cost efficiency at producing education and overall in 2001 were 87.9% (S.D. = 8.2%) and 84.9% (S.D. = 8.1%) respectively. These mean scores were shown to have declined by approximately 5% each from 1996 levels. The efficiency scores of 4-cluster institutions, in contrast, remained relatively stable over time, with the mean efficiency in education being 86% (S.D. = 8%) and overall being 82.5% (S.D. = 5.5%). Finally, the mean efficiency scores for the 3-cluster institutions were the lowest among all HBO clusters. In 2001, mean cost efficiency was shown to be 77.5% (S.D. = 10.4%) and mean overall cost efficiency to be 68.8% (S.D. = 11.9%). Moreover, the mean scores for education were shown to have declined by 11% and overall declined by 18% since 1996.

The authors devoted only marginal attention to interpreting the efficiency findings and even went so far as to not call them efficiency scores but instead “measures of group homogeneity.” Scores were reported only to give readers a broader perspective on the cost variation across Dutch higher education institutions rather than assess efficiency. The reason cited for this was that DEAs use in the study was not for the purpose of estimating efficiency but instead as a novel way to apportion shared costs in order to derive more accurate per-student cost estimates. As such, they recognized among other factors that not adequately accounting for quality, performing analyses on small samples, and using the more restrictive CRS DEA model were all likely to limit the ability to derive meaningful conclusions about cost efficiency.

## 5. Discussion

At the beginning of this paper I observed that an increasing number of empirical studies on higher education efficiency have been conducted on a multi-national scale. As the last section (combined with the list in Appendix A) clearly shows one can confidently say that, today, efficiency may be regarded as a fully-fledged subtopic within the broader higher education literature base; so much so that the number of empirical studies are even beginning to rival the numbers devoted to estimating production and cost relationships. At the same time, it is also apparent that a high degree of diversity is present in both the methodologies used to estimate efficiency in these studies and the findings they generate. That said, in this last section I return to the questions posed at the beginning of the paper with three goals in mind: 1) establishing the extent to which commonalities and differences are evident in the different types of efficiency analyses, and 2) establishing guidelines under which relatively reasonable inferences about institutional efficiency can be made.

### 5.1 Across country comparison

One of the most striking findings from reviewing different higher education efficiency studies is that, where judgment is made about the level of system efficiency, most researchers suggest that technical and/or cost efficiency is relatively high. This is puzzling in that it seems to contradict economic theories of nonprofit behavior, especially where higher education has been analyzed, that suggest inefficiency to be much more prevalent than in for-profit firms. Yet in none of the studies did the author express any concern that inefficiency was pervasive. While some passed no judgment at all, when opinions were presented the most stringent admonishment was that “room for improvement was possible” or should not be ruled out. Most, however characterize system efficiency as “high,” “performing very well,” “respectable,” or that inefficiency was “fairly modest.”

There also seems to be a degree of technique preference across different countries. In Australia, for example, all three studies use DEA<sup>42</sup> while SFE seems to be the approach of choice in the UK. Only in case of the Netherlands do we see a more balanced choice of techniques. No pattern seems to emerge in the US which is rather surprising given: 1) the large number of institutions and, 2) ease of access in

---

<sup>42</sup> As a side note, one has to wonder why SFE was not used in the Australian case as most of the SFE studies use a computer program (Frontier version 1.4) for estimating such functions developed by Tim Coelli at the University of New England.

obtaining rather detailed statistical data. What we do see in the US though is that both of the DEA studies focus less on estimating efficiency and more on testing whether theoretical predictions about efficient organizational structure and production are evident.

The studies of higher education efficiency in the Netherlands are unique for several reasons. First is that a relatively large number of efficiency studies have been conducted in a relatively short time span: five since 1994. What makes it so interesting though is that, in principle, it should be very difficult to perform analyses such as these where the entire university sector consists of only 13 institutions. Such small samples inherently generate model bias and, in all probability, push efficiency scores artificially high in the Dutch university sector, which several of the authors do note. Second, it is the only country where any effort is made to examine efficiency in a cross-national context. In addition to estimating cost efficiency in the Netherlands, the two earliest studies (Jongbloed, *et al.*, 1994 and Vink, 1997) repeated this exercise for Germany and for the UK. Though neither analyzes cost efficiency directly (i.e. through the use of a pooled sample similar to that in Vink's study) they do partake in a partial analysis in an attempt to provide a more cohesive picture.

It would seem that all higher education efficiency studies are relatively uniform with respect to the measures they use. The "typical" model specifies two outputs,<sup>43</sup> education and research (though several variables may be used for each), and two inputs, labor and non-labor. In most studies sensitivity analyses were also done. Education output is usually expressed in terms of FTE enrollments and, depending on the study, may be split out into undergraduate and graduate students or by broad disciplinary groups such as the arts and sciences. In only two studies reviewed for this paper (Avkiran's "performance on the delivery of educational services" model and Athanassopoulos and Shale's "outcome efficiency" model) did the researcher not proxy education output by student enrollments. The most common measure of research output used is "research grants and contract income," though in some cases publication counts or some other physical unit is specified. Inputs are most often expressed in physical units (i.e. numbers of FTE academic and non-academic staff) though cost efficiency studies obviously employ expenditure measures (in some cases studies also use a mixture of physical and cost-based measures). In terms of the latter, some studies distinguish between labor and non-labor expenditures (e.g. Jongbloed *et al.*, 1994; Jongbloed and Salerno, 2003) whereas

---

<sup>43</sup> Public service is rarely accounted for, even indirectly.

others consider expenditures in aggregate (e.g. McMillan and Datta) or by function, such as in Ahn *et al.* (1988) (instructional and overhead expenditures plus physical investments) or Athanassopoulos and Shale's (1997) (general expenditures and research income). Where a mix of physical and cost inputs is specified, labor is invariably expressed in FTEs while "other expenditures" are used as a proxy for all other institutional inputs.

Yet these similarities seemingly belie the extent to which meaningful comparisons between different countries' systems can be derived. As a rule, it is highly likely that very little can be gleaned from such analyses. Below I briefly outline what I see as the major obstacles to making cross-system inferences about higher education institutional efficiency.

The most glaring problem, given that the preponderance of efficiency studies uses DEA, is that it is simply not possible to directly compare relative efficiency scores found in different countries. Even under relatively ideal circumstances (e.g. using identical input and output measures and comparing like systems) finding that the mean technical efficiency for one system is 95% while in another it is only 90% does not mean that the latter is less efficient than the former. It may very well be that if the two systems were analyzed jointly we would find that the most efficient institutions in the former system are, in fact, *inefficient* when compared to the latter. Unfortunately, this problem is a byproduct of the underlying mechanics behind DEA itself. The strongest aspect of DEA, its capability to estimate efficiencies for multi-input/output institutions, depends directly on estimating efficiency relative to others. As a result, not only is it difficult to draw meaningful comparisons between aggregate efficiency scores (i.e. those calculated in VRS or CRS models) but also scale efficiency as well. Remarkably out of all the studies presented in the last section only Abbott and Doucouliagos (2003) explicitly state that, though efficiency appears high, it is not possible to determine whether the system *writ large* is underperforming, especially in comparison with other institutions in other higher education systems (p. 96).

At the same time, many methodological problems arising in traditional cost and production function estimates come are evident. I am specifically talking about how to account for aspects of input and output quality, long regarded to be the Achilles' heel of higher education cost and production studies. First where quality is accounted for, it is often only done partially and on outputs (only

Athanassopoulos and Shale attempt to control for input quality).<sup>44</sup> For example, two of the Australian studies use the Research Quantum as a quality-controlled measure of research output though neither attempts to account for education quality. Another example is Stevens' (2001) analysis of UK universities where he used variables like "average A-level scores" and the "percentage of students receiving firsts and upper-second" (in secondary education) to account for education quality but simply used "research income" as a proxy for research output. The one possible exception is Salerno's (2002) study where institutions were sorted into broad quality tiers and analyzed separately. However, the sort criterion he used was a variable called "mean scholarly quality of (an institution's) program faculty,"<sup>45</sup> which arguably gives too much emphasis to research.

This inability to adequately control for quality gives rise to another issue that must be considered when evaluating the efficiency estimates in the different studies: institutional comparability. Because there is a wide degree of variance between what is and what is not included in the different input and output measures for each institution, one must bear in mind that institutions found to be inefficient may not, in fact, be inefficient at all. That is, in the absence of quality control, efficiency scores will likely include differences attributable to quality.

A closely related concern to that above is sample homogeneity. For example comparing universities with medical schools to those without penalizes the former because medical schools are both labor and cost intensive: one would be comparing "apples with oranges" so to speak. Such a notion has not been lost on the studies examined here. In most cases efficiency is calculated based on like comparisons: research universities are not compared to liberal arts institutions and, especially in the case of the Netherlands, science clusters are not compared with arts clusters. The DEA-oriented studies tend to group institutions by type while in the SFE studies all tend to account for it through the use of dummy variables.

In the end, quality creates problems on several levels that further serve to limit cross-country comparisons. Because there is little consensus amongst researchers about how to appropriately account

---

<sup>44</sup> Two reasons for emphasizing output, rather than input, quality are readily apparent. The first is data availability. The second reason is that in cost function estimations, the explanatory variables are outputs.

<sup>45</sup> This measure was reported in the National Research Council's (NRC) survey of research and doctorate programs in the United States (1993).

for quality in higher education (Nelson and Hevert, 1992), even in the unlikely case where two studies both control for quality in the same way (that is, on the same variables), the different approaches researchers use to accounting for quality in various country studies will still distort comparisons. Unfortunately, collecting quality data has proven to be a costly task in terms of time and financial resources. As a result, what data is available is almost exclusively collected by government statistical agencies, which means it is generally not collected on an annual basis and usually underscored by the political forces of the day. The extent then, to which quality is accounted for in any economics-related studies of higher education institutions, depends heavily on the availability and reliability of existing data and is very likely to be country-specific.

The last major issue I would like to bring attention to here is the comparability problems arising when inputs are measured in terms of expenditures. As legislators increasingly express an interest in ensuring public funds are used productively and efficiently, it is logical that they take a broader perspective, consider the entire system as the unit-of-analysis, and ask where it stands relative to other countries. On its face, expressing inputs in a relatively independent unit of measurement seems to be the most promising way in which to compare institutions or systems at the international level. Obviously this requires first converting different countries' expenditures into a common measurement unit like purchasing power parity (PPP), but that is easily surmountable and will become even less of an issue, at least for Europe, with broader adoption of the Euro.

Where the real problems lie is in the different accounting frameworks organizations like higher education institutions use, not only in different countries, but even across institutions *within* countries. Research by Winston (2000; 1997; 1994) for example has shown that, even within the United States, the various ways in which universities report expenditure data (particularly to the National Center for Education Statistics) significantly underestimate institutions' capital expenditures and make it difficult to capture important economic concepts like opportunity costs. As such, he argues that it would be better to use expenditure data from institutions' own financial statements and, in his work, outlines an algorithm for computing such costs.

The problems, however, with this approach are that: 1) it can be very cumbersome to obtain the financial statements for large numbers of institutions (like in the US or UK), and 2) the algorithm

provided by Winston is, self-admittedly, complex. Nonetheless, if Winston is correct then it would be difficult to draw comparisons between the different efficiency studies surveyed in this paper as the institutional expenditure data used comes almost exclusively from national data sets. The only instance I found where researchers obtained expenditure data from institutions' financial statements was in the DEA study of Dutch higher education institutions by Jongbloed, *et al.* (1994), but in that case only 13 institutions were being analyzed.

Another major problem with comparative cost efficiency analysis is that, more often than not, what is included in a given expenditure category differs by country. Personnel expenditure categories can be ambiguous in terms of whether they include salaries of administrative or support staff, or may or may not also include fringe benefits. Administrative expenditures may or may not include both labor and non-labor expenditures. Most problematic, and pervasive, though is the "other expenditures" category, which is used in many of the efficiency studies presented in the last section (Abbott and Doucouliagos, 2003; Jongbloed and Salerno, 2003; McMillan and Datta, 1998; Athanassopoulos & Shale, 1997; Coelli, 1996; Jongbloed, *et al.*, 1994; Ahn, *et al.*, 1988). Depending on what is included, this category may include expenditures on: libraries, computing, laboratory equipment, student support services (e.g. counseling and health centers), and administrative staff costs.

This ambiguity strongly comes to bear when interpreting estimates of cost efficiency. First, even within a given system, a more thorough accounting of costs combined with inadequate quality measures is likely to result in greater variation with respect to efficiency scores.<sup>46</sup> Second, direct comparison between two independent efficiency studies from different countries is not possible, even if PPP expenditure measures are computed. Finally, where institutions from multiple systems are compared together, institutions in those countries accounting for a greater proportion of their costs would be penalized in an efficiency analysis; estimates would be distorted by incompatible accounting techniques.

---

<sup>46</sup> As I have argued in other places (Jongbloed and Salerno, 2003), where quality cannot be satisfactorily controlled for, more of the variation in institutional efficiency scores will be explainable by differences in quality rather than productive or cost inefficiency. This was the basis for our decision to label the efficiency tables "degrees of institutional homogeneity" in our paper.



## 5.2 Policy implications

In the final analysis then we can confidently draw several, albeit sobering conclusions about higher education efficiency that are salient in a more practical setting. Foremost, though a number of different studies have been conducted across several different countries it is not possible to determine from the available evidence whether any given system is more (less) efficient than any other. While having such information would be highly useful from a policy perspective, particularly where decisions about aggregate funding levels must be made, it would be a serious mistake to draw inferences or develop comparative benchmarks on the currently available information.

Second, even though the mathematical and statistical tools currently available are flexible enough to estimate higher education efficiency, a number of logistical problems still persist that make it very difficult to draw reasonable inferences about efficiency even within any particular system. I am specifically talking about notions of input and output quality which, as I discussed earlier, are still grossly neglected in most empirical efforts. It is probably safe to assume there is a one-to-one relationship between the validity of estimated efficiency scores and the extent to which quality is properly accounted for. In other words, as efforts to account for quality in the study increase one can be more confident that any inefficiencies found actually reflect differences attributable to inefficiency and not other factors like quality. However, as the findings of this survey reveal, a good deal of the inefficiency reported in the different studies reviewed here is heavily confounded by quality factors and it is important, especially for the policymaker, to keep such factors in mind when using the results to effect real system changes.

Finally, though one could easily get lost in the sheer volume of efficiency scores and various supporting statistics, the majority of the studies do little in the way of explaining *why* inefficiencies occur. In this regard it is almost pleasing that most efforts seem to identify a high degree of institutional efficiency; if not, policymakers would have to rely on little more than intuition when developing efficiency enhancing policies. An ideal higher education efficiency study would develop hypotheses about where inefficiencies might arise and then employ statistical inference to test the validity of such claims. Even in the absence of statistically significant findings, such efforts would go a long way toward helping policymakers identify and implement intelligent policies for enhancing productivity and efficiency. Unfortunately however, to-date such efforts have been the exception and

not the rule. Importantly this means that one must take considerable care when formulating efficiency-enhancing policies based on the available research.

### 5.3 A modest proposal

Clearly many of the structural barriers to comparative analysis expressed earlier in this section show no signs of abating and most will likely remain at the forefront of methodological considerations for years to come. That said, I conclude this paper by laying out guidelines for comparative efficiency analysis that, at the least, will allow researchers to draw more realistic and methodologically grounded inferences about cross-national efficiency.

Foremost it is highly unlikely that any meaningful analysis could take place between more than two countries. There are just too many factors, even at the system level, such as national educational objectives and internal labor market conditions that researchers would at least have to make an effort to control for. Evidence for this can be found in Vink's variance analysis, which showed statistically significant differences across the countries he examined. Even for only a three-country analysis, the number of ill-defined factors such as these would grow exponentially. Thus the first consideration is finding suitable, or comparable, countries for the analysis. By this I mean identifying two countries or even regions sharing strong similarities in areas like economic development (in terms of technology use, research and development, and levels of industry) and education participation rates in the population. Some good organizations having data that could be used for this would include the OECD, the World Bank, and the Asian Development Bank. Most importantly the compared countries should be as similar as possible in terms of higher education structure. Here I am talking about issues like relatively similar *time to completion rates* for obtaining a first degree, systems of institutions with comparable missions (e.g. professional education institutions, research universities, and teaching institutions), and common degree structures (e.g. the bachelor/master framework in Anglo-Saxon countries). This is not to say that countries failing to meet such criteria are incomparable; it just means the closer two systems are the easier it will be to take stock of the macro-level "intangibles" the limit such analysis.

Once the comparator countries have been established, the next consideration is how to develop some set of outputs and inputs for which the data can be collected for institutions in both countries. How, for example, can the output “research” be specified so that a country whose institutions do relatively more contract-based or applied research be compared against another with a more traditional “publishing” orientation? As we saw in the review of prior efficiency studies, enrollments are probably the best proxy currently available for education. Do the available data allow the researcher to distinguish between full- and part-time students? Can data for “comprehensive” universities be subdivided into enrollments in the arts and sciences? This would be critical where technical universities or teaching colleges co-exist with multi-purpose universities.

As for inputs, on the one hand, if the object is to estimate cost efficiency then is it possible to obtain the data directly from institutions’ financial statements? Though this may be time-intensive or even require the assistance of an accountant in an analysis, it does give the researcher greater control over *how* expenditures can be categorized, which is the critical task in such a study. More importantly though, it permits a more accurate accounting of *what* is included in different expenditure categories: the primary limitation I identified above. If, on the other hand, the objective is to estimate technical efficiency then emphasis should be placed on identifying the most disaggregated data for which analyses can be conducted. For example, suppose staff numbers are available. Is it possible to distinguish between full- and part-time employees, academic and non-academic staff, or between researchers and educators? If not, is it possible to indirectly make these specifications? These are the types of questions that need to be addressed and what determine how deep of an analysis can be.

Realistically, it is probably very difficult, if not impossible to do any detailed efficiency analysis. To expect such flexibility though misses the point. In a comparative study, the focus shifts dramatically from imposing controls on the model to maintaining consistency in the model.

This leads to a third consideration. Earlier I suggested that lack of attention to issues like quality is likely to distort estimates of institutional efficiency. However, in a comparative analysis *simplicity* is key. For that reason alone it would be more instructive to eschew efforts at accounting for various aspects of quality, which I have already pointed out has the potential to be politically biased, and instead focus on straightforward measures of inputs and outputs. After the initial efficiency analysis

has been completed, the researcher can then step back, impose broad quality considerations in a more qualitative framework, and evaluate how efficiency may differ between countries in terms of quality. Seeing as the best one can do is speculate about the influence of quality on an international level anyway, fewer problems are likely to arise in an *ex post facto* qualitative analysis than in trying to make relative precise efficiency estimates in a quantitative analysis with falsely precise controls on quality.

Finally, the last major point that needs considering is *how* the analysis should be done. In the third section of this paper, I explored the two most popular approaches to assessing the efficiency of higher education institutions: stochastic frontier estimation and data envelopment analysis. I also discussed how the weak points to each type of analysis are strong points of the other. Given that the inputs and outputs specified can readily be incorporated into either type of analysis, using both techniques to calculate efficiency can help to validate the findings of the other.

## References

- Abbott, M. & Coucouliagos, C. (2003). The efficiency of Australian universities: A data envelopment analysis. Economics of Education Review 22, p. 89-97.
- Ahn, T., Charnes, A., & Cooper, W.W. (1988). Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. Socio-economic Planning Sciences 22(6), 259-269.
- Aigner, D., Lovell, C.A.K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6, p. 21-37.
- Athanassopoulos, A.D, & Shale, E. (August,1997). Assessing the Comparative Efficiency of Higher Education Institutions in the UK by Means of Data Envelopment Analysis. Education Economics 5(2), 117-134.
- Avkiran, N.K. (2001). Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. Socio-Economic Planning Sciences 35. 57-80
- Banker, R.D., Charnes, A., & Cooper, W.W. (1984). Models for estimating technical and scale efficiencies in data envelopment analysis. Management Science 30(9).
- Brinkman, P.T. (1990). Higher education cost functions. In S.A. Hoenack & E.I. Collins (Eds.), The economics of American universities. Buffalo, NY: State University of New York Press.
- Charnes, A., Cooper, W.W., & Rhodes, E.L. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2(6). 429-44.
- Coelli, T., Prasada Rao, D.S., & Battese, G.E. (1998). An Introduction to Efficiency and Productivity Analysis. Boston: Kluwer Academic Publishers.
- Coelli, T. (1996). Assessing the performance of Australian universities using data envelopment analysis. Internal report, Centre for Efficiency and Productivity Analysis, University of New England.
- Cohn, E., Rhine, S.L.W., & Santos, M.C. (1989). Institutions of higher education as multi-product firms: Economies of scale and scope. The Review of Economics and Statistics 71 (2), 284-290.
- Cooper, W.W., Seiford, L.M., & Tone, K. (2000). Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software. Boston, MA: Kluwer Academic Publishers.

- de Groot, H., McMahon, W., & Fredericks Volkwein, J. (1991). The cost structure of American research universities. The Review of Economics and Statistics 73 (3), 424-431.
- Dundar, H., & Lewis, R.D. (1995). Departmental productivity in American universities: Economics of scale and scope. Economics of Education Review 14 (2), 119-144.
- Farrell, M.J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A, 120(3). 253-290.
- Glass, J.C., McKillop, D.G. & Hyndman, N. (1995). Efficiency in the provision of university teaching and research: An empirical analysis of UK universities. Journal of Applied Econometrics, 10(1), p. 61-72.
- Goudriaan, R., Jongbloed, B.W.A. & van Ingen, D.C. (1998), *Kostendeterminanten en Doelmatigheid van het Nederlandse Hoger Onderwijs*, Beleidsgerichte Studies Hoger Onderwijs en Wetenschappelijk Onderzoek 57, Ministerie van Onderwijs, Cultuur en Wetenschappen. Den Haag: SDU.
- Hopkins, D.S.P. (1990). The higher education production function: Theoretical foundations and empirical findings. In S.A. Hoenack & E.I. Collins (Eds.), The economics of American universities. Buffalo, NY: State University of New York Press.
- Izadi, H., Johnes, G., Oskrochi, R., & Crouchley, R. (2002). Stochastic frontier estimation of a CES cost function: The case of higher education in Britain. Economics of Education Review, (21), p 63-71.
- James, E. (1990). Decision processes and priorities in higher education. In S.A. Hoenack & E.I. Collins (Eds.), The economics of American universities. Buffalo, NY: State University of New York Press.
- James, E., & Rose-Ackerman, S. (Eds.). (1986). The nonprofit enterprise in market economics. New York: Harwood Academic Publishers.
- James, E. (1978). Product mix and cost disaggregation: A reinterpretation of the economics of higher education. Journal of Human Resources 13(2), 157-186.
- Johnes, G. (1997). Costs and industrial structure in contemporary British higher education. Economic Journal, 107, p. 727-737.
- Johnes, J. & Johnes, G. (1995). Research funding and performance in U.K. university departments of economics: A frontier analysis. Economics of Education Review 14(3), 301-314.

- Jondrow, J., Lovell, C.A.K., Materov, I.S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, *19*, p. 233-238.
- Jongbloed, B.W.A., Koelman, J.B.J., Goudriaan, R., de Groot, H., Haring, H.M.M. & van Ingen, D.C. (1994), *Kosten en Doelmatigheid van het Hoger Onderwijs in Nederland, Duitsland en Groot-Brittannië*, Beleidsgerichte Studies Hoger Onderwijs en Wetenschappelijk Onderzoek 35, Ministerie van Onderwijs, Cultuur en Wetenschappen. Den Haag: SDU.
- Jongbloed, B.W.A. & Koelman, J.B.J. (1996), *Universiteiten en Hogescholen vergeleken*, Beleidsgerichte Studies Hoger Onderwijs en Wetenschappelijk Onderzoek 38, Ministerie van Onderwijs, Cultuur en Wetenschappen. Den Haag: SDU.
- King, W.D. (1997). Input and output substitution in higher education. Economic Letters, *57*, 107-111.
- Mar Molinero, C. (1996). On the joint determination of efficiencies in a data envelopment analysis context. Journal of the Operational Research Society *47(10)*, p. 1273-79.
- McMillan, M.L. & Datta, D. (1998). The relative efficiency of Canadian universities. Canadian Public Policy, *24(4)*, 485-511.
- Mettas, R.D., Vargas, V.A., & Whybark, D.C. (2001). An investigation of the sensitivity of DEA to data errors. Computers & Industrial Engineering, *41*, p. 163-171. National Research Council. (1993). *Research-Doctorate Programs in the United States*. Washington: National Academy Press.
- Nelson, R., & Hevert, K.T. (1992). Effect of class size on economies of scale and marginal costs in higher education. Applied Economics *24*, 473-482.
- Nicholson, W. (1995). Microeconomic theory: Basic principles and extensions. New York: The Dryden Press.
- Pedraja-Chaparro, F., Salinas-Jimenez, J., & Smith, P. (1997). On the role of weight restrictions in data envelopment analysis. Journal of Productivity Analysis, *8*, p. 215-230.
- Robst, J. (2001). Cost efficiency in public higher education institutions. Journal of Higher Education, *72(6)*, p. 730-750.
- Salerno, C.S. (2002). On the technical and allocative efficiency of research-intensive higher education institutions. Unpublished doctoral dissertation. The Pennsylvania State University: University Park, PA.
- Shepherd, R.W. (1953). Cost and Production Functions. Princeton: Princeton University Press.

- Stevens, P.A. (2001). The determinants of economic efficiency in English and Welsh universities. National Institute of Economic and Social Research, Discussion paper number 185.
- Varian, H.R. (1992). Microeconomic Analysis. New York: W.W. Norton
- Vink, M.J.C. (1997). Efficiency in higher education: A comparative analysis on sectoral and institutional level. Twente, The Netherlands: Universiteit Twente.
- Winston, G.C. (2000). A guide to measuring college costs. In M.F. Middaugh (Ed.), Analyzing Costs in Higher Education: What Institutional Researchers Need to Know. New Directions for Institutional Research, 106, (Summer).
- Winston, G.C. (1997). College costs: Subsidies, intuition, and policy. In the National Commission on the Cost of Higher Education (NCCHE) report, Straight Talk About College Costs and Prices. Phoenix, AZ: ACE-Oryx Press, p. 117-127.
- Winston, G.C. (1994). A note on the logic and structure of global accounting. Williams Project on the Economics of Higher Education, Discussion Paper no. 23, Williamstown, MA: Williams Project on the Economics of Higher Education.
- Worthington, A.C. (2001). An empirical survey of frontier efficiency measurement techniques in education. Education Economics, 9(3), 245-268.
- Zhang, Y., & Bartels, R. (1998). The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand. Journal of Productivity Analysis, 9, 187-204.



## Appendix A

### Selected bibliography of sub-institutional efficiency studies

- Bessent, A.M., Bessent, E.W., Charnes, A., Cooper, W.W., & Thorogood, N.C. (1983). Evaluation of education program proposals by means of DEA. Educational and Administrative Quarterly 19(2), p. 82-107.
- Beasley, J.E. (1995). Determining teaching and research efficiencies. Journal of the Operational Research Society 46(4), p. 441-52.
- Beasley, J.E. (1990). Comparing university departments. Omega International Journal 18(2), p. 171-183.
- Chen, T. (1997). A measurement of the resource utilization efficiency of university libraries. International Journal of Production Economics 53(1), 71-80.
- Johnes, J. & Johnes, G. (1995). Research funding and performance in U.K. university departments of economics: A frontier analysis. Economics of Education Review 14(3), 301-314.
- Johnes, J. & Johnes, G. (1993). Measuring the research performance of UK economics departments: An application of data envelopment analysis. Oxford Economic Papers 45, p. 332-347.
- Mar Molinero, C. (1996). On the joint determination of efficiencies in a data envelopment analysis context. Journal of the Operational Research Society 47(10), p. 1273-79.
- Stern, Z.S., Mehrez, A. & Barboy, A. (1994). Academic departments efficiency via DEA. Computers and Operations Research 21(5), p. 543-556.
- Tomkins, C. & Green, R. (1988). An experiment in the use of data envelopment analysis for evaluating the efficiency of UK university departments of accounting. Financial Accountability and Management 4(2), p. 147-164.