Journal of the Operational Research Society (2011) 62, 1282–1297

www.palgrave-journals.com/jors/

Costs and efficiency of higher education institutions in England: a DEA analysis☆

E Thanassoulis[1*], M Kortelainen[2], G Johnes[3] and J Johnes[3]

[1]*Aston Business School, Aston University, Birmingham, UK;* [2]*University of Manchester, Manchester, UK; and* [3]*Lancaster University Management School, Lancaster, UK*

As student numbers in higher education in the UK have expanded during recent years, it has become increasingly important to understand its cost structure. This study applies Data Envelopment Analysis (DEA) to higher education institutions in England to assess their cost structure, efficiency and productivity. The paper complements an earlier study that used parametric methods to analyse the same panel data. Interestingly, DEA provides estimates of subject-specific unit costs that are in the same ballpark as those provided by the parametric methods. The paper then extends the previous analysis and finds that further student number increases of the order of 20–27% are feasible through exploiting operating and scale efficiency gains and also adjusting student mix. Finally the paper uses a Malmquist index approach to assess productivity change in the UK higher education. The results reveal that for a majority of institutions productivity has actually decreased during the study period.

Journal of the Operational Research Society (2011) **62,** 1282–1297. doi:10.1057/jors.2010.68

Published online 21 July 2010

Keywords: higher education; data envelopment analysis; performance measurement; productivity; cost function

Introduction

The last 20 years have been a time of rapid change in the UK's higher education sector. Many former polytechnics have gained university status and student numbers have expanded substantially in response to various policy changes. These have included the introduction of student loans for maintenance in 1990, and the subsequent introduction of tuition fees. In an environment of expanding student numbers, it is vital for the government to understand the cost structures that underpin provision in this sector as well as to find out the potential for improved performance of higher education institutions (HEIs). However, although it is known that addressing key policy issues in UK higher education requires research on cost structures, little recent information is available about the costs and performance of HEIs.

This paper draws on a study commissioned by the Department for Education and Skills, now Department for Innovation, Universities and Skills. The aim of the study was to investigate the structure of costs in the UK higher education in the period 2000/2001–2002/2003 in light of the fact that the UK government at the time wanted to increase substantially the number of students attending university. The commissioned study used both econometric and Data Envelopment Analysis (DEA) methods to study the cost structure and addressed a number of issues including cost per student by type, economies of scale and scope, and productivity change over time. The findings on subject-specific unit costs and returns to scale and economies of scope based on parametric regression methods are reported in Johnes *et al* (2008). This paper reports the findings on subject-specific unit costs and on returns to scale using DEA, finding a large measure of agreement on the results given by the two different approaches. The paper then goes further by examining inefficiency of HEIs and by analysing the performance improvement potential existing in the sector.

In evaluating costs and performance of HEIs, it is generally important to account for the multi-product nature of educational production. This has been done in a number of previous studies in the higher education sector; see for example Stevens (2005), for a review. The studies measuring performance of HEIs have typically used either DEA or stochastic frontier analysis (SFA) to evaluate efficiency of institutions. For SFA applications using the UK data, see for example Izadi *et al* (2002),

☆This paper draws on research carried out on behalf of the UK Department for Education and Skills, now Department for Innovation, Universities and Skills, DIUS. The views expressed in the paper are those of the authors and no representation is being made that they are necessarily shared by the DIUS.

*Correspondence: E Thanassoulis, Operation & Information Management Group, Aston Business School, Aston University, Birmingham, B4 7ET, UK. E-mail: E.Thanassoulis@aston.ac.uk

Stevens (2005) and Johnes and Johnes (2009); and for DEA applications in the UK, see Johnes and Johnes (1995), Athanassopoulos and Shale (1997), Flegg *et al* (2004), Beasley and Wong (1990), Beasley (1995) and Johnes (2008). Although the performance analysis of HEIs in the UK has been the subject of several previous studies, most of them have some limitations. First, HEIs have been traditionally treated as a homogenous group, although there is a lot of variety between HEIs. For example, cost and output profile is quite different in traditional universities to those observed in former polytechnics that were granted university status in 1992. Second, only three outputs (ignoring any possible subject disaggregation) have been considered: undergraduate and postgraduate (PG) teaching and research. Apart from a few exceptions, the so-called third mission activities of HEIs have not been included as output, even though they have an increasingly important function in society by encompassing, *inter alia*, the provision of advice and other services to business, the storage and preservation of knowledge, and the provision of a source of independent comment on public issues (see eg Verry and Layard, 1975). Importantly, Johnes *et al* (2008) accounted for these limitations by estimating separate parametric cost functions for distinct HEI groups and by including in the analysis more disaggregated teaching outputs as well as a variable measuring the third mission output. However, they stop short of an in-depth efficiency analysis, their main emphasis being on analysing cost structures of various HEI groups and calculating estimates for economies of scale and scope.

This paper uses a 3-year panel data set of 121 HEIs in England in order to analyse the performance of institutions and evaluate the potential for efficiency improvements of HEIs. We follow Johnes *et al* (2008) by estimating separate models for distinct HEI groups and by using more disaggregated teaching outputs, including a variable measuring the third mission output. Besides estimating subject-specific unit costs and inefficiency scores with DEA, we use several different DEA models to study potential gains that could be produced by achieving most productive scale size (MPSS). In addition, we examine potential augmentations in student numbers without additional costs, including ways of exploring alternative mixes of student numbers. Finally, by utilising the panel structure of the data, we estimate a Malmquist productivity index and its components separately for different HEI groups. This permits technology or the efficient boundary to vary in different years (in each group) and allows us to decompose productivity change into efficiency change and boundary shift components.

The rest of the paper is organised as follows. The following section outlines methodologies used in the study. The third section discusses the variables used and presents an empirical analysis of costs in the higher education sector in England based on DEA. Finally, the last section presents conclusions from this research.

Methodology

Data envelopment analysis

The methodology we use in this study is DEA (Farrell, 1957; Charnes *et al*, 1978), which is a well-known linear programming method for measuring the relative efficiencies of Decision Making Units (DMUs) such as bank branches or universities. DEA is an alternative method to SFA (Aigner *et al*, 1977; Meeusen and van den Broeck, 1977), which is an econometric technique for efficiency analysis based on regression analysis. Generally, DEA and SFA are the two main methods of choice for modelling cost structures and more generally measuring efficiency of organisational units. The two approaches are mathematically quite different, each one having its own advantages and drawbacks. The main advantage of SFA is that it allows for noise in the data and makes possible stochastic inferences, while DEA basically assumes that data are noise-free. However, SFA requires strong parametric assumptions for the functional form linking output and inputs (or costs) and (usually) also distributional assumptions for noise and inefficiency, whereas DEA does not require any kind of parametric assumptions and is thus non-parametric. Nevertheless, more recently both statistical tests and bootstrapping methods for confidence intervals on DEA efficiencies have been developed (see eg Banker and Natarajan, 2004; Simar and Wilson, 2008). Regarding the application considered here, one relevant virtue of DEA is its flexibility; it is quite straightforward to estimate DEA models that treat some or all outputs as endogenous. A further advantage of DEA in the present application is that it can yield specific information about targets, benchmarks etc for each unit in turn, which can be used to examine possible savings in cost or output augmentations in the sector as a whole or at specific HEIs under alternative policies for efficiency and productivity gains.

Apart from measures of efficiency, in which the production context permits non-constant returns to scale, DEA makes it possible to identify whether a unit operates under increasing (IRS), decreasing (DRS) or constant returns to scale (CRS). It also makes it possible to identify the *MPSS* at which a unit could operate. Note that returns to scale under DEA correspond to ray economies of scale under SFA and other parametric methods in that they concern maintaining the mix of inputs and outputs and simply changing scale size. A statement of the models we have used can be found in the Appendix while a fuller introduction to DEA can be found in Thanassoulis *et al* (2008).

Typically, we are not just interested in identifying the type of returns to scale at a particular unit, but also how far it is from MPSS. By estimating efficiencies under both constant (CRS) and variable returns to scale (VRS) models, it is possible to determine the scale efficiency for a unit. The scale efficiency score for an individual HEI can be simply calculated as a ratio of its efficiency score under the CRS model to that under the VRS model (see the Appendix). The scale efficiency of a unit measures the extent to which a unit can lower its costs by changing its scale size to the MPSS. In the analysis here, we will maintain VRS as our basic assumption, which is consistent with the findings of non-constant ray economies of scale for the same data set in Johnes *et al* (2008). We will determine returns to scale properties for efficient units in each group of HEIs and also examine scale inefficiency at the group level. Later, however, we will use the assumption of CRS to estimate potential savings and/or output augmentations were HEIs to attain MPSS. (We did not test statistically for differences between CRS and VRS efficiencies as we ran both CRS and VRS assessments, and irrespective of any test results a DEA assessment under VRS would be needed to compare our results with those in Johnes *et al*, 2008. More recent literature allows for hybrid returns to scale in which some inputs and outputs follow constant while others VRS (see Podinovski, 2004). This type of returns to scale was, however, not appropriate for our input–output set.)

Features of DEA models used

In the empirical analysis here, we treat the outputs (specified below in terms of teaching, research and the third mission) as exogenously fixed and attempt to estimate the minimum cost at which a HEI could have handled the output levels that it had. This means that we adopt an input orientation. To complement the input-oriented analysis, we will also later alter the orientation to estimate maximum output levels, keeping expenditure constant. This helps us to examine if one or more of the outputs can be increased further without incurring additional costs.

It is worth emphasising that our estimates of efficient levels of costs (or outputs) are relative rather than absolute. That is to say, each time we take a full set of HEIs or some subset, we identify benchmark HEIs in that set that offer the lowest total operating cost for their mix and absolute levels of output. Those units that are not on the frontier have scope for cost savings relative to the benchmarks. Benchmark units themselves may have scope for cost savings relative to some absolute standard, which is not known to us. Thus, a drawback with DEA is that we could be identifying a unit as an efficient benchmark simply because there are no suitable comparators for its mix of outputs and/or scale size. On the other hand, the strength of DEA is that when we do identify a unit as inefficient, the benchmarks will clearly indicate why that unit is deemed inefficient.

Initially, we shall assess efficiencies by treating all HEIs in the sector over the 3 years as a coherent set, operating the same technology in terms of how costs are driven by the outputs captured in our model. This analysis will give a broad-brush view of relative efficiencies, but the set is not used in subsequent analyses as the group of all HEIs represents too diverse a set. Instead, more reliable results are sought by grouping HEIs into more homogeneous subsets by objectives and operating context. Specifically, we group HEIs into more uniform subsets in technology consisting, respectively, of four groups as explained below.

Note that by estimating a DEA model for the whole sample or for the subgroups using 3 years' pooled data, we make no assumptions as to whether or not a HEI has changed efficiency over the 3 years. Our analysis merely assumes that the technology of delivering education over the 3 years concerned has not changed, in the sense that if a cost level (after adjusting for inflation) could support a given bundle of outputs in 1 year it could have done so in any one of the 3 years. However, since it is possible that efficient boundaries are different for different years, we will also conduct a separate analysis, in which we allow boundaries to shift and also measure productivity change in the sector.

Malmquist index approach

To examine whether there have been changes in technology during the assessment period, we will relax the assumption of no change in technology by evaluating productivity changes and boundary shifts year on year using DEA. Our approach is based on the Malmquist productivity index that was introduced as a theoretical index by Caves *et al* (1982) and has been used since then in a large number of empirical studies. The DEA-based approach to estimate the Malmquist productivity index and its components was developed by Färe *et al* (1994a, b). The basic idea behind this approach is to use DEA to estimate separate efficient boundaries for different periods, and then decompose total factor productivity (TFP) change into two subcomponents: *efficiency catch up* and *boundary shift*, which, respectively, measure the extent to which productivity changes are due to changes in efficiency and technology. For details of how to compute the Malmquist Index and its components, the interested reader is referred to Thanassoulis (2001, Chapter 7) and for alternative decompositions of the Malmquist index to Lovell (2003).

Assessing efficiency and productivity of HEIs in England

Input-output variables used

Our analysis uses data on all HEIs in England, covering ancient universities, such as Oxford and Cambridge, traditional universities (in the pre-1992 sector), new universities (mainly former polytechnics that were granted university status in 1992), and colleges of higher education (members of GuildHE, which is an association of colleges of higher education that do not have university status). We focus on England in order to avoid complications that arise from spatial differences in the higher education system arising from devolution of powers to Scotland and Wales. We exclude a small number of institutions on the grounds that they have acquired medical schools during the period under consideration, and hence have moved from one group of institutions to another. The data have all been provided by the Higher Education Statistics Agency (HESA). In common with Johnes *et al* (2008), we use panel data that relate to the years 2000/2001 through 2002/2003.

We use a single input, that of *total operating costs*, net of residence and catering costs, adjusted for inflation. Our outputs, detailed in Table 1, reflect full-time equivalent (FTE) undergraduate student load by subject area, FTE PG students, value of research grants and third leg activities. These input and output variables are the same variables as used in the analysis of the same data by parametric models in Johnes *et al* (2008). As teaching outputs, these types of studies have usually employed the number of undergraduate and PG students. However, here more disaggregated variables for undergraduate students are used, as we allow for distinct categories for medicine and dentistry students as well as science students and non-science students. For PG students we use total number of students across all disciplines.

We are mindful of the fact that student numbers as used here do capture the quantity but not the quality of output on teaching. One approach to capture quality on teaching would be to break down student numbers by degree class awarded (first class, upper second class, and so on). There is in the UK system a degree of comparability of degree classifications across universities, as universities appoint external examiners who scrutinise results for maintaining standards across universities. However, there are a number of practical difficulties in adopting this approach. First, the model will also then have to control for the academic level of students on entry as prior attainment is generally a strong predicting factor for attainment on exit from university. This is not easy to do for students who come in with a variety of qualifications not only from the UK but also from all over the world. Standardising across such a range of qualifications would require too many subjective assumptions. Even if all this could be done on a conceptual level, in practice such data on qualifications on entry are not available for the period covered by this study. Second, for PG students we have too coarse a classification of attainment, namely simply fail, pass or pass with distinction. Notwithstanding this, we still face the problem of allowing for their attainment on entry given that a very high proportion of PG students in the UK come in with non-UK entry qualifications. Finally, allowing for breakdown for quality of students both at entry and exit, unless reduced to a single indicator each, would lead to considerable loss of degrees of freedom, and yet reducing to single indicators, multiple exit and entry quality indicators would require subjective assumptions. For all these reasons, unfortunately, we could not reflect teaching quality in this study.

Table 1 Definition of variables used in the analysis

Type of variable	*Variable*	*Description*
Input	*TOPCOST*	Total operating cost (£000) in constant prices. This figure is inclusive of depreciation*
Outputs	*UGMED*	Full-time-equivalent (FTE) undergraduates in medicine or dentistry (000)
	UGSCI	FTE undergraduate science students (000). Summation of subjects allied to medicine, veterinary, biological, agriculture, physical sciences, maths, computing, engineering and architecture
	UGNONSCI	FTE undergraduate non-science students (000). Summation of social economics, law, business, librarianship, languages, humanities, creative arts and education
	PG	FTE postgraduate students in all disciplines (000)
	RESEARCH	Quality-related funding and research grants in constant prices (£m)
	THIRD MISSION	Income from other services rendered in constant prices (£m)

*Total operating costs does not include 'hotel' costs related to catering and student accommodation. We decided to exclude hotel costs, because these are unrelated to the core education function of institutions. Instead, the total operating cost measure does include depreciation, since we wish to include the cost of capital in our estimates of costs.

Regarding research, we use research funding as a proxy for research activity fully appreciating that there are hazards implicit in this approach. Nevertheless, since research funding is based on (i) peer reviewed research proposals that are linked to specific project output and (ii) the outcome of the research assessment exercise, we consider this to be an adequate proxy. An alternative would have been to use research assessment scores aggregated to institution level (see, for instance, http://web.archive.org/web/20030102041040/http://www.gla.ac.uk/rae/ukleague2001.xls); it is known that the degree of correlation between these scores and funding is extremely high (Johnes and Johnes, 2009).

It is generally known that the third mission activities have nowadays an increasingly important function for HEIs in the UK, involving the provision of advice and other services to business and regional development, the storage and preservation of knowledge, and the provision of a source of independent comment on public issues. Despite the importance of the third mission for society, excluding Johnes *et al* (2008), previous studies have not included the third mission activities as output primarily due to data limitations. For example, De Groot *et al* (1991) note that: 'We realize the importance of public sector for many universities. There is, however, very limited nationwide information of output of this type'. We address this deficiency of previous studies by incorporating into our DEA analysis a variable measuring the amount of the third mission work. Although published data do not allow the extent of such activities to be measured very precisely, the income from 'other services' identified in the HESA data provides one possible measure. In the absence of a better alternative, this is what we have used in the analyses that follow.

Descriptive statistics for the chosen input and output variables can be found in Table 2. In order to make values of monetary variables in different years comparable, deflated variables are used. Thus, based on the retail price index, monetary values within the data were adjusted to 2002/2003 prices using inflators of 1.0366 and 1.0294 for 2000/2001 and 2001/2002, respectively. These deflators may be compared with, and are close to, those produced by Universities UK for non-pay expenditure in higher education.

Interestingly, we note in Table 2 that there are some considerable variations in the values of input and output variables depending on the type of HEI. On the input side, the range of total operating costs across institutions is large, reflecting the large differences not only in scale but also in HEI type. For example the minimum cost for a pre-1992 university with a medical school is higher than the maximum cost for a GuildHE college. Even larger variations among various HEI groups can be found in the number of undergraduates and PGs and research income. For example, research income is on average more than 10 times and more than 100 times higher for traditional institutions than for post-1992 institutions and for GuildHE colleges, respectively. Note that the diversity of the specified groups results mainly from the historical development of the institutions. Some institutions within the traditional university sector, for example, have developed from Colleges of Advanced Technology, and, as such, the subject mix that is provided by these institutions is heavily skewed towards the sciences.

Owing to considerable diversity across HEIs in the English higher education sector, it seems reasonable to group HEIs by type for the purpose of efficiency analysis.

Table 2 Descriptive statistics for the variables in the data set*

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
All institutions					
TOPCOST	363	84144.29	88612.72	1372	462530
UGMED	363	0.207	0.544	0	2.724
UGSCI	363	2.552	2.243	0	7.719
UGNONSCI	363	3.388	2.615	0	12.616
PG	363	1.733	1.447	0	6.068
RESEARCH	363	21.653	42.512	0	213.689
3RD MISSION	363	4.263	5.273	0	29.946
GuildHEs					
TOPCOST	114	17274.36	12729.09	1372	51047
UGMED	114	0	0	0	0
UGSCI	114	0.539	0.643	0	2.310
UGNONSCI	114	1.726	1.371	0	5.621
PG	114	0.441	0.514	0	2.429
RESEARCH	114	0.435	0.558	0	2.397
3RD MISSION	114	0.701	1.512	0	8.512
Post-1992 HEIs					
TOPCOST	99	86907.6	21948.76	42805	133524
UGMED	99	0	0	0	0
UGSCI	99	4.371	1.468	1.163	7.464
UGNONSCI	99	5.971	2.169	2.590	12.616
PG	99	2.132	0.866	0.768	4.078
RESEARCH	99	4.711	3.014	0.171	12.547
3RD MISSION	99	4.746	2.479	0.498	12.800
Pre-1992 universities medicine					
TOPCOST	54	251138.1	106265.2	52103	462530
UGMED	54	1.395	0.575	0.077	2.724
UGSCI	54	4.784	1.755	0.286	7.719
UGNONSCI	54	4.531	2.269	0	11.223
PG	54	3.954	1.968	0.283	6.068
RESEARCH	54	103.907	60.198	19.84	213.688
3^{RD} MISSION	54	12.012	6.604	0.378	29.946
Pre-1992 universities non-medicine					
TOPCOST	96	66768.66	33134.16	9277	136116
UGMED	96	0	0	0	0
UGSCI	96	1.813	1.745	0	5.506
UGNONSCI	96	2.056	1.815	0	6.027
PG	96	1.607	1.220	0.110	5.658
RESEARCH	96	18.051	12.525	0.319	45.256
3RD MISSION	96	3.637	4.749	0	24.498

*Units of measurement are specified in Table 1.

To account for diversity, Johnes *et al* (2008) used in their estimations three groups of institutions: GuildHE colleges, new universities, and traditional universities. In this analysis, we will use four subsets: GuildHE colleges, new universities, traditional (pre-1992) universities with medical schools and traditional universities without medical schools. It seems well-founded to separate traditional universities into those with and those without medical schools as their cost structures are generally quite different as can be seen in Table 2 (see Agasisti and Salerno, 2007, for a similar point).

Identification of outliers

As noted above, DEA is a deterministic frontier method as it does not allow random noise in the data generating process. As a result, the efficient boundary in DEA can be sensitive to extreme data points. Such data points can impact significantly the location of the efficient boundary and yet their isolated position raises doubts as to whether the data are genuine or the result of random noise or other error. We shall attempt therefore to identify and remove such observations before we carry out the analysis of performance. We shall refer to such observations as *outliers*. It should be noted that outlier observations here are simply those showing exceptionally 'high efficiency' relative to the rest of the observations rather than being outliers in a statistical sense, in which very low cost efficiencies could also feature as outliers. Outlier observations of poor performance are not of concern in DEA as they do not impact the location of the efficient boundary, which in turn forms the reference plane for all efficiencies estimated.

Figure 1 illustrates the adopted approach in respect of identifying outlier HEIs. It depicts HEIs that use a single input [I] to secure a single output [O]. The left panel in Figure 1 depicts the efficient boundary for the full set of HEIs.

To identify outliers, we adapt the procedure used by Thanassoulis (1999). We first identify the units with exceptional achievements by using the concept of 'super-efficiency' introduced by Andersen and Petersen (1993). The central idea in measuring the super-efficiency of a HEI (say B in Figure 1) is to assess it relative to the efficient boundary drawn on the remaining HEIs, that is excluding HEI B as shown on the right panel of Figure 1. Thus, in Figure 1 the super-efficiency (input-oriented) of HEI B is given by the ratio AU/AB, which is clearly larger than 1. The further B is from the remaining data points the larger its super-efficiency. Thus, we can use the super-efficiency measure to judge how far a data point is from the rest of the data and thereby decide whether it is to be treated as an outlier or not.

Following Thanassoulis (1999), we adopted a threshold difference of super-efficiency of 10 percentage points to identify outliers. That is to say any subset of HEIs that had super-efficiency over 100% and were separated from other less efficient units by a gap of 10 percentage points or more were deemed to be outliers. For instance, if we had super-efficiencies ordered 110, 112, 123, 124 and 125%, the units with super-efficiency 123% or more were deemed to be outliers. Once a set of outliers was removed the super-efficiencies were estimated again until either there was no gap of 10 percentage points in super-efficiency or 5% of the sample had been identified as outliers. This means no unit in the final set lies more than 10 percentage points in efficiency further away than some other unit or 5% of the sample exceed in efficiency the final boundary used. Once the outliers were identified we did not permit them to influence the position of the efficient boundary, but retained them with their data adjusted to sit on the boundary mapped out by non-outlier units.

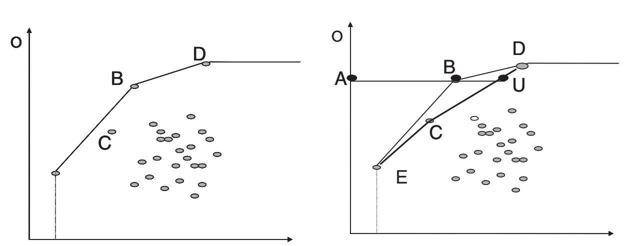

Figure 1 The identification of outliers.

Efficiencies and unit costs for the full sample of HEIs

Using the above procedure we identified five outliers in the full sample of HEIs. The outliers in the pooled sample were 'ordinary' HEIs not known to specialise in any way on provision or mission. This suggests they simply in the years concerned had very low expenditure relative to the bundle of outputs we capture. This could to an extent be the result of moving expenditure from one year to the next. (In contrast within the subgroups of institutions, the outliers were predominantly specialist institutions by way of only offering a limited subset of the curriculum 'normal' institutions offer.) Table 3 summarises the results obtained for all 3 years together and for year 3 separately. Having calculated by DEA the results for year 3 separately allows a comparison to be made between the efficiencies derived through DEA and those estimated using SFA taken from Johnes *et al* (2008), where the results relate to year 3 only. The DEA efficiencies exhibit a higher mean and narrower range than the SFA efficiencies. Spearman's rank correlation coefficient between the DEA and SFA ranks on efficiency is 0.60, which is significant at the 1% level. While highly significant, this correlation is not particularly high,

Table 3 Summary of efficiencies (%) (all HEIs excluding outliers)

	N	*Min*	*Q1*	*Mean*	*Median*	*Q3*	*Max*	*Std. Dev.*
DEA	358	27.5	79.3	86.3	91.2	99	100	15.8
DEA year 3 only	118*	27.6	78.9	85.4	90.5	98.7	100	16.7
SFA year 3 only	121	6.0	67.0	74.7	83.7	89.7	98.7	22.9

*The number of observations for year 3 in DEA is three short of that in SFA due to outliers identified in DEA. The outliers do feature at 100% efficiency when computing mean results by DEA.

which is as expected, given that all groups of HEIs are aggregated here into one overall sample.

We can evaluate efficiency at the sector level if we divide the aggregate efficient level of expenditure by the corresponding aggregate observed expenditure across all HEIs. This ratio is 0.924, suggesting that HEIs could have saved about 7.6% of their total expenditure if they had all been performing at the level of the benchmark HEIs. Given the noise in the data, this does not in itself suggest there is a great scope for savings at sector level. There is, however, scope for quite considerable savings at some HEIs as can be deduced from the lower quartile efficiencies that are below 80%.

Turning now to marginal costs, in DEA we have a different set of marginal costs per unit output at each efficient segment (or facet) of the boundary such as EC and CD in the right panel of Figure 1. In order to summarise the information we can attempt a parametric description of the DEA boundary. This is possible in this case because we have a single input. It involves projecting the units on the efficient boundary so that in effect inefficiencies have been eliminated. (For instance, project all inefficient units in the right panel of Figure 1 to the efficient frontier ECD). We can then use OLS regression on the 'efficient' input–output profile of each HEI to derive an equation for the boundary. (For further details of this and related approaches to estimate sets of unit costs with DEA jointly with other methods, see Thanassoulis, (1996).) As the boundary by construction is piece-wise linear we shall attempt a linear model for it.

The best fit equation estimated after dropping some eight of the least efficient (relative to the line being estimated) observations is

$$\begin{aligned} TOPCOST = & 13121 \times UGMED + 5657 \\ & \times UGSCI + 4638 \times UGNONSCI \\ & + 3828 \times PG + 1376 \times RESEARCH \\ & + 1537 \times 3RDMISSION \end{aligned}$$

with statistically significant regressors. (The coefficients and their standard errors and p-values are reported in Table 6.) This equation fits the 'efficient' data well offering an R^2 of 0.995. Note that in regressions of the boundary of this type R^2 is typically quite high, as the variation of the original data that was attributable to inefficiency has been eliminated through projecting the data to the efficient boundary. The equation was forced through the origin because the regression constant is not statistically significant. In essence, therefore, we are estimating an approximation to the VRS boundary that matches the part of the boundary in which CRS hold. The unit output costs therefore will reflect better the more productive of the HEIs (those enjoying CRS) rather than those operating under increasing or DRS.

Table 4 compares unit costs produced by DEA (and OLS) with the parametrically derived unit costs reported in Johnes *et al* (2008). When interpreting these results, one must be aware of the following two points that complicate somewhat the straightforward comparison of the DEA and the parametric unit output costs. First, we are using different definitions of unit output costs between DEA and the parametric methods. In the case of the parametric methods, we are using the cost function estimated to compute average incremental costs (AICs), which reflect 'the cost on average for a unit of output' were a HEI to go from zero to an average level of that output while keeping the rest of the outputs at average levels. In contrast, in DEA we are estimating a 'best fit' set of unit costs for the estimated DEA efficient input-output levels of the HEIs. Thus, the unit costs reported here need to be seen as broad brush rather than precise estimates.

Table 4 shows that DEA agrees with the other methods regarding the observation that medical undergraduates, on average, cost more than their science counterparts, who in turn cost more than their non-science counterparts. Interestingly, all methods yield similar costs for science undergraduates. However, DEA estimates medical and PG students at lower (more efficient) level than the parametric methods, while the opposite is the case for non-science undergraduates. Note that the monetary estimates mean that it is more than three times costlier to educate medical than PG students according to the DEA results, whereas parametric methods do not give so large a difference between the costs associated with these two types of

Table 4 Units costs by DEA and AICs by parametric methods*

	$N=358$**	$N=363$		
	DEA 'Mean' Unit costs(£)	*Stochastic frontier AIC (£)*	*GEE pa AIC (£)*	*Random effects AIC (£)*
UGMED	13121	15973	16132	17769
UGSCI	5627	5506	5258	5079
UGNONSCI	4638	3665	3046	3217
PG	3828	6979	9643	9569

*See Johnes *et al* (2008)for the discussion of the GEE and random effects methods.

**Note that 358 here reflects the number of observations that were available to determine the DEA frontier, as the remaining five units being outliers were simply adjusted to sit on the frontier the rest of the units could determine.

students. Taking into account estimates from previous studies, it seems that DEA is likely to underestimate unit costs for PG students. On the other hand, as we might expect, the results of the SFA (which, like DEA, evaluates the position of an efficient boundary) are the ones that are closest to the results of DEA for all student groups. On the whole, given the totally different assumptions underlying DEA and parametric methods and the fact that in all methods we are estimating summary ('average') unit costs the degree of agreement between the methods is quite remarkable.

Efficiencies and unit costs by HEI group

As the HEIs are very different in terms of objectives, history and operating practices, we focus our attention next on assessing HEIs in more homogeneous subsets. As explained earlier, estimations are implemented separately for four subgroups: traditional universities (pre-1992) with and without medical schools, new universities (post-1992) and GuildHE colleges. For compactness and for ease of comparison, where applicable, the results for all groups are presented jointly in Tables 5, 6 and 7. We will comment on the results by group in the ensuing subsections.

Pre-1992 universities without medical schools

This set consists of 32 HEIs over 3 years making a total of 96 observations. Three outliers were identified. The estimated efficiencies are reported in Table 5, the OLS estimates and standard errors in Table 6, and the unit output costs given by DEA and parametric methods in Table 7. We have generally high efficiency in the sector (median efficiency 98.91% as can be seen in Table 5)—though there are some individual HEIs that have quite low efficiencies as the minimum value of 39.65% and the relatively high standard deviation suggest. Were all HEIs to have operated at the benchmark level they could have saved on average 6% from their total expenditure, implying that the efficiency of this subset is on average just under 94%. Again this is a remarkably high level of efficiency. Of course it should be recalled that this merely suggests performance is fairly uniform on cost relative to output levels. We have no way of knowing through this type of comparative analysis whether the institutions are cost efficient in some absolute sense.

As in the case of the full sample, we estimated a mean level of costs per unit of output by projecting all HEIs to the efficient boundary and then estimating the boundary through OLS. The parameters of the resulting equation can be seen in Table 6 (Pre-1992 no medical schools columns). The equation was forced through the origin as the regression constant was not significant. The DEA-based unit output costs from this equation are contrasted with unit output costs derived using parametric methods in Table 7 (Pre-1992 no medical schools rows). Here we have reasonably close agreement between all the methods on unit costs except that DEA estimates unit PG costs much higher than do the parametric methods. Note that this is contrary to the full sample results, in which DEA unit cost was much lower for PG students (£3828 *versus* £12369 at 2002/2003 prices). A probable explanation for this variability is the fact that the full sample is too diverse for DEA to give reliable estimates. Yet, it is hard to say whether DEA unit cost is closer to true value than parametric estimates for this group. In any case, the results give us more affirmation that, for pre-1992 universities without a medical school, it is more than two times costlier to educate a PG student than a science undergraduate student and that non-science undergraduate students have the lowest unit costs.

Post-1992 universities

This subset consists of 33 HEIs over 3 years making a total of 99 observations. Our preliminary analysis did not identify any outliers and so all HEIs in principle can be used to form the efficient boundary of this subset. Once again while the range of efficiencies is some 26 percentage points wide, there is generally uniform performance on efficiency among post-1992 institutions with over 75% of institutions having efficiency at the 88% level or better (Q1 = 88.79%). Taking on board the fact that we have not allowed for noise in the data, we have relatively little scope for cost savings in this subset too, but as we will see later more scope for output augmentation, keeping costs as they are. The efficient level of expenditure for this subset is 93.5% of the actual expenditure, which again reflects a remarkable level of uniformity of efficiency. It should be also noted that efficiencies in Table 5 are not comparable across groups as the efficient boundary used is different for each subset of units.

Table 5 Summary of DEA efficiencies (%) for HEI groups

Subgroup	*N*	*Min*	*Q1*	*Mean*	*Median*	*Q3*	*Std. Dev.*
Pre-1992 HEIs without medical schools (three outliers)	96	39.65	91.06	92.61	98.91	100	13.63
Post-1992 universities (no outliers)	99	73.65	88.79	93.67	96.5	100	7.352
GuildHE colleges (three outliers)	114	27.55	78.88	85.99	90.5	100	16.85
Pre-1992 HEIs with medical schools (no outliers)	54	87.97	97.16	98.23	100	100	3.16

Table 6 OLS estimated approximations to the DEA efficient boundary

	All		Pre-1992, no medical schools		Post-1992		GuildHE		Pre-1992 with medical schools	
	$\hat{\beta}$	Std. error (p-value)	$\hat{\beta}$	Std. error (p-value)	$\hat{\beta}$	Std. error (p-value)	$\hat{\beta}$	Std. error (p-value)	$\hat{\beta}$	Std. error (p-value)
UGMED	13121	1095 (0.000)	—	—	—	—	—	—	10631	3670 (0.006)
UGSCI	5657	287.9 (0.000)	4655	612.7 (0.000)	6006	273.3 (0.000)	7046	340.6 (0.000)	3992*	904.7 (0.000)
UGNONSCI	4638	246 (0.000)	3047	580.8 (0.000)	2714	183 (0.000)	3070	177.9 (0.000)	3992*	904.7 (0.000)
PG	3829	539.6 (0.000)	12369	899 (0.000)	7504	582.3 (0.000)	6273	447 (0.000)	7572	2676 (0.07)
RESEARCH	1376	16.20 (0.000)	1052	64.94 (0.000)	675	183.1 (0.000)	2243	344.2 (0.000)	1470	46.67 (0.000)
THIRDMISSION	1537	107 (0.000)	1846	220 (0.000)	1131	113.1 (0.000)	1073	117.9 (0.000)	1051	409.7 (0.013)

*For pre-1992 with medical schools UGSCI and UGNONSCI have not been separated and have the same coefficient.

We again estimated a mean level of costs per unit of output in this subset by using the approach outlined earlier. After dropping six observations that were the least efficient relative to the line being estimated, we obtained the OLS model detailed in Table 6 from which we draw the DEA-based unit output costs for this subset. Unlike the preceding two cases, we obtained a significant set up cost in the form of a positive regression constant. This means the costs we are estimating on this occasion are more in line with the part of the efficient boundary in which we have non-CRS. Importantly, there is a considerable level of agreement between all methods on the unit costs for this subset as can be seen in Table 7, with the exception that DEA estimates the unit cost of an undergraduate science student to be considerably higher than do the parametric methods. Nonetheless, the results show that all methods agree that also in this group average cost is higher for PG students than for undergraduate science students who in turn cost more than their non-science counterparts.

GuildHE colleges

This set consisted of 38 units observed over 3 years making a total of 114 observations. Following the procedure outlined earlier two institutions were identified as outliers. Here we have greater variation in efficiency than is the case with either pre- or post-1992 universities. This is as we might expect given the greater diversity of types of GuildHE colleges ranging from very specialised to those offering a 'full' range of university-type courses. The efficient level of expenditure for these colleges is 90% of the observed expenditure. Although this is down on the types of university modelled earlier, it is still a good level of efficiency in comparison with those found in studies of other sectors. However, as we will see later, there is in relative terms much more scope for output augmentation in the GuildHE colleges, particularly if we focus on simply raising student numbers.

Using the procedure outlined earlier we estimated the linear regression model for the DEA efficient boundary for GuildHE colleges whose parameters appear in Table 6. The unit costs from this model are compared with parametrically derived unit costs in Table 7. Looking at the relevant part in Table 7, we see that the unit output costs for this subgroup, as estimated by DEA, are considerably higher than the estimates obtained using parametric methods in which UG science students and PG students are concerned, and lower for UG non-science students. The DEA estimates here are likely to reflect better the situation than the parametric ones. This is because the parametric AICs, as we saw, assume mean levels on all bar the output whose mean incremental cost is being estimated. However, GuildHE colleges tend to specialise in specific outputs and so the assumption of mean levels on all

Table 7 Unit output costs estimated for HEI groups with DEA and parametric methods

Subgroup	*Method*	*UGSCI*	*UGNONSCI*	*PG*	*UGMED*
Pre-1992 HEIs without medical students	DEA	4655	3047	12369	
	SFA	4935	3981	8133	
	GEE	4300	2487	8877	
	Random eff.	4320	2423	8956	
Post-1992 HEIs	DEA	6006	2714	7504	
	SFA	4465	2725	7680	
	GEE	4229	2884	7373	
	Random eff.	4204	2863	7345	
GuildHE colleges	DEA	7046	3070	6273	
	SFA	5604	4808	2030	
	GEE	5760	5069	2891	
	Random eff.	5660	5096	3158	
Pre-1992 HEIs with medical students	DEA	3992	3992	7572	10631
	SFA	2805	4778	4607	17079
	GEE	5305	3773	4753	12350
	Random eff.	4093	3930	5982	15268

bar one output is not safe. In contrast, DEA by its nature permits a unit to give maximum weight (ie estimated unit cost) to the outputs on which its performance is best relative to other HEIs. Thus, given that GuildHE colleges tend to specialise in small subsets of our outputs, DEA would estimate the 'maximum' cost at which that college could attain its best possible efficiency level relative to other colleges, assuming in general negligible unit costs for those outputs on which the college has low or even zero level. Thus the DEA basis for estimating unit costs is closer to reality in the case of GuildHE colleges compared to the AIC approach.

Pre-1992 universities with medical schools

This subset consists of 18 HEIs over 3 years making a total of 54 observations. There were no extreme observations in the form of outliers as defined earlier and so all observations have been used in the assessment. We have little discrimination here on performance due to the relatively small sample and the large number of variables and the fact that we take scale as exogenous. The efficient level of expenditure for this subset is 98.4% of the observed total expenditure thus being remarkably high. Again, the picture changes if we switch from cost minimisation to output augmentation in which we can identify significant scope for raising output numbers.

Using the approach outlined earlier of projecting inefficient HEIs to the boundary and then using OLS regression, we obtain DEA-based unit cost estimates that may be compared with those obtained using parametric methods. The OLS model appears in Table 6, last two columns on the right. As can be seen in the relevant part of Table 7, so far as science UG students are concerned, clearly the SFA unit cost estimate is low; indeed being lower than that for non-science students it is counter-intuitively so. It is also much lower than the estimated cost for science undergraduates in other groups of universities. Although the unit cost of PG students as evaluated by DEA is higher than in parametric estimates, it is actually more in line with unit costs for such students in pre-1992 universities without medical schools, and generally closer to the estimates for unit costs for PG students obtained by all methods in pre-1992 universities without medical schools. In view of this the DEA, and perhaps the random effects model, estimates are the most plausible, DEA perhaps underestimating the cost of a medical student while random effects overestimating it. This picture is reversed where PG students are concerned. In all cases, as we might expect, the results confirm that on average it is much more expensive to educate medical than any other students.

Looking at the results collectively the following summary points can be made so far:

- DEA shows scope for cost savings at sector level of the order of 5–10% of the observed expenditure; however, the potential gains through efficiency are considerably higher at some HEIs.
- Unit costs estimated by parametric and non-parametric methods here need to be used only as rough indications. We have a complex set of institutions operating at different scale sizes and different output mixes. Naturally they experience varying costs and our methods offer no more than a broad-brush summary of the complex underlying structure of unit costs.

- We have imposed no restriction on the weight an institution places on any one of the outputs in arriving at the estimates of efficiency. However, if either a given institution, or the funding body for that matter, wishes to adhere to some preferences structure over the value of raising alternative outputs (eg favouring student numbers over research output or the other way round), then the DEA models solved would need to be adjusted to reflect this. One variant of this has been implemented below for the case when student number increases are to be prioritised over research output. However, additional models for imposing weights restrictions in estimating efficiencies can be found in Thanassoulis *et al* (2004) and for imposing alternative preference structures when estimating targets in Thanassoulis and Dyson (1992).

Returns to scale and potential savings

We next examine the efficient units mapping out the boundary in each one of the subgroups modelled in order to get a sense as to the type of returns to scale predominating in each case. Table 8 shows the type of returns to scale identified at the efficient units in the various sets we have modelled. (The full set of 121 $\times$ 3 has not been computed here as it is too diverse to offer reliable returns to scale estimates. For example, we could be benchmarking a university with medical school on a GuildHE college with few disciplines.) The indications are that, *on the frontier*, in all but one subset of the sector returns to scale can be characterised as predominantly constant or decreasing. Only in post-1992 universities do we mostly have constant or IRS.

Table 8 is complemented by Table 9, which gives a measure of the savings that are possible, in principle, were HEIs to eliminate diseconomies of scale as distinct from eliminating technical inefficiency given their scale size. Table 9 suggests that there is relatively little room in pre-1992 universities with medical schools for either scale or operating efficiency gains. In contrast, pre-1992 universities without medical schools can, on aggregate, gain about 6% through operating efficiency improvements and a further 6% through scale efficiency improvements. GuildHE colleges can gain the most, in total over 15% on aggregate, two-thirds of it through operating efficiency and one-third through scale efficiency gains. There is relatively little to be gained in post-1992 HEIs through ray scale efficiency adjustments. However, as we will see, more gains can be made if we refocus our priorities from cost savings to output expansions.

So far our attention has been input-oriented. That is to say we have sought to estimate the minimum cost at which each HEI could operate given its output levels. However, we can also estimate the augmentation of output levels, notably student numbers that would be feasible at current levels of expenditure if inefficiencies were to be eliminated.

We computed the augmented 'efficient' levels for all outputs using the output-oriented model that scales all outputs equiproportionately maintaining the mix of all outputs (students, research and third mission). The potential output augmentations based on this model are presented in Table 10(a). As can be seen from the results, for given inputs, across the sector there is scope for about 10% rise in undergraduate science, 15% in non-science undergraduates and 17% in PG student numbers. About two-thirds of these gains are possible through the elimination of technical inefficiency and the remainder through the additional elimination of scale inefficiencies. Looking at the different types of institution the largest rise in student numbers possible in relative terms is to be found

Table 8 Returns to scale holding at efficient units

	IRS	*CRS*	*DRS*	*Total number efficient*
Pre-1992 without medical schools ($N=96$)	3	20	21	44
Pre-1992 with medical schools ($N=54$)	1	18	17	36
Post-1992 universities ($N=99$)	10	21	3	34
GuildHE colleges ($N=114$)	1	24	12	37

Table 9 Decomposition of potential savings through eliminating technical inefficiency and scale size diseconomies

	Percent of actual expenditure attributable to technical inefficiency	*Percent of actual expenditure attributable to scale inefficiencies*	*Percent of actual expenditure recoverable through operating and scale efficiency gains*
Pre-1992 no medical schools ($N=96$)	6.02	6.49	12.51
Pre-1992 with medical schools ($N=54$)	1.65	2.65	4.30
Post-1992 Universities ($N=99$)	6.51	2.28	8.80
GuildHE colleges ($N=114$)	10.66	4.94	15.60

at GuildHE colleges ranging from 20% for undergraduate science to 36% for PG students through a combination of scale and efficiency gains.

Clearly, more sophisticated analysis than that reported in Table 10(a) is possible if we vary the priorities for output expansion. For example, we may modify the models to favour expansion of say science undergraduates. Further, priorities over output expansion can be varied by type of institution favouring, say medical student rises in universities with medical schools, science undergraduates in say post-1992 universities and so on. Indeed priorities can be varied at HEI level offering the HEI the option to set its own priorities for student expansion perhaps within broad national guidelines. Finally, investigations can be carried out permitting additional investment beyond the observed level of expenditure to identify efficient output levels at the new level of expenditure either varying or maintaining output mix.

In order to examine the differences in results that can be obtained when the priorities for output expansion are not uniform across all outputs, we estimated alternative DEA models in which only student numbers are expanded giving virtually zero weight to the rise in research and the third mission output. The results appear in Table 10(b). Comparing Tables 10(a) and (b) we see that there are many remarkable changes when only students are targeted to increase. Looking at the rows labelled 'Total' and for the case in which both technical and scale inefficiencies have been eliminated, we see that the percentage rise in science undergraduates doubles from 11% to 22% and there is a 10 percentage point rise in the number of PG students from 17.52% to 27.16%. The least change is in undergraduate non-science students, in which the percentage gain rises from 15.26% to 19.81%.

Looking at individual types of institutions in Table 10(b), we see even greater potential for student number increases. For example, pre-1992 universities without medical schools can recruit about 33% and 25% more undergraduate science and non-science students, respectively, by simply eliminating technical inefficiencies. These percentages nearly double when scale inefficiencies are additionally eliminated. GuildHE colleges can virtually double their PG students—albeit from a low base—when both scale and technical inefficiencies are eliminated.

These are large potential gains and it is instructive to see how the findings come about. We have used a DEA model that maximises the total gains in student numbers at each HEI without the need for additional expenditure or any decrease in research and third mission activity. The model has sought for each HEI to raise those student numbers in which the maximum gain in absolute terms can be made, unconstrained by the need to maintain the mix of outputs. In some cases the model suggests only one type of student be augmented (eg at one university only science students rise), because that is where the maximum

Table 10 (a) Potential output augmentation maintaining current levels of expenditure and output mix; (b) Potential output augmentation maintaining current levels of expenditure but targeting only student numbers to rise to best advantage at each HEI

	Percent rise through eliminating technical inefficiency			*Percent rise through eliminating technical and scale inefficiency*		
	UG SCIENCE	*UG NON-SCIENCE*	*PG*	*UG SCIENCE*	*UG NON-SCIENCE*	*PG*
(a)						
Pre-1992 without medical schools ($N=96$)	7.71	13.32	8.78	12.67	26.02	21.62
Pre-1992 with medical schools*($N=54$)	2.09	2.33	2.34	8.4	5.6	9.35
Post-1992 Universities ($N=99$)	10.05	11.34	13.27	11.22	13.5	18.48
GuildHE colleges ($N=114$)	13.64	13.21	24.5	20.62	22	36.73
Total	7.63	10.15	9.32	11.33	15.26	17.52
*Medical students: after eliminating technical inefficiency 4.64%; after eliminating technical and scale inefficiency 9.93%.						
(b)						
Pre-1992 without medical schools ($N=96$)	33.33	24.85	9.84	64.74	57.53	20.30
Pre-1992 with medical schools*($N=54$)	2.83	1.59	4.44	11.23	0.69	15.72
Post-1992 Universities ($N=99$)	8.38	13.45	22.85	10.25	17.92	27.11
GUILDHE colleges ($N=114$)	19.16	6.67	55.33	30.84	11.63	98.36
Total	12.17	11.83	15.97	22.00	19.81	27.16

*Medical students: after eliminating technical inefficiency 9.93%; after eliminating technical and scale inefficiency 37%.

potential for gain in student numbers lies. In this sense the results in Table 10(b) represent the potential for gains not only by eliminating scale and technical inefficiency, but also by eliminating 'allocative' inefficiency in the sense of maximising aggregate student numbers by altering the mix of students where appropriate. This explains to a large extent the substantial potential for gains in student numbers at no extra cost. We must, however, when looking at these apparent possible gains, also be mindful of the fact that our models have not discriminated between different types of science or non-science students. For example, there may be a substantial cost differential between educating say mathematics and biology students, yet the model treats both types as simply science students. As the model by its nature would tend to use the cheapest type of science student as benchmark, it may be over-estimating potential gains at HEIs that have a larger proportion of the more expensive type of student within each one of our three overarching categories of science, non-science and PG students.

It is recalled that our model in its outputs reflects quantity but not quality of teaching. We need to ensure that the increased numbers of students estimated here can be catered for without detriment to quality of teaching. As we have not included variables on quality of teaching we cannot, in principle, be certain that the increased student numbers will not necessarily mean a deterioration of teaching quality. However, the estimated targets can still be used as follows. We know from our analysis the benchmarks on which the estimated higher student numbers are based for each one of the institutions that are not benchmark themselves. We can, in respect of each non-benchmark institution, assess outside the DEA framework teaching quality as, for example, it may reflect on student outcomes relative to quality of students recruited. If the teaching quality is deemed at least of the same levels as that of the non-benchmark institution then we can use the estimated raised student numbers as targets for the non-benchmark institution. Otherwise a judgement needs to be made whether the benchmark HEIs do provide acceptable quality of student outcomes even if not to the same standard as the non-benchmark institution before the targets are accepted.

The foregoing caveats are specific to the particular output variables adopted here and the data limitations. They are not generic to the methodology being used. DEA can cope with any break down of students, including variables on quality and quantity of students or indeed research by category, provided we have the necessary data and sufficient observations to carry out the analysis.

Productivity change between 2000/2001 and 2002/2003

The foregoing assessments have treated the 3 years from 2000/2001 to 2002/2003 as a single cross section. This is compatible with assuming that in the 3 years involved, 'technology of production' has not changed substantially so that whatever output levels were feasible for a given level of expenditure in any one of the 3 years in the cross section will also be so in any other year within the cross section, once of course we adjust for cost inflation. In this section we drop this assumption and instead check whether there has been any productivity change at HEI level, and if so to what extent and at which HEIs. Further, we check whether in each subset the efficient boundary has moved and if so whether that was towards a more productive location.

We implemented the foregoing approach separately for each one of the four subsets of HEIs measuring productivity change over the 2-year period from 2000/2001 to 2002/2003. We excluded outliers from the subsets as identified in each case earlier.

The results on total factor productivity change are summarised in Table 11 for all four subsets of institutions. The median TFP change as reflected in the Malmquist Index is 0.98 both for pre-1992 universities without medical school and for post-1992 universities. Thus on average these HEIs have registered little change in productivity, which is perhaps not surprising given the short time period the data cover. In contrast, the median TFP change for pre-1992 universities with medical school and the GuildHE HEIs is 0.94 suggesting they have suffered an average 6% loss in productivity over the 2 years. This may be partly a consequence of above-inflation increases in costs faced by HEIs, particularly in the latter 2 years of our study (see the HEPPI, Higher Education Pay and Prices Index, produced by Universities UK available at http://www.universitiesuk.ac.uk/Newsroom/Facts-and-Figures/Documents/heppi_guide_2005.pdf). The bulk of the TFP change estimates are between 0.9 and 1.15 suggesting that the majority of HEIs registered anything from a loss of 10% to a gain of 15% in productivity. There is a good size minority of post-1992 HEIs that show a tendency to have the higher productivity gain. GuildHE colleges have a wider range of productivity change even after dropping two of their extreme values. This is indicative of the wider diversity of type of HEI within the GuildHE definition. The maxima values of the Malmquist index at 1.34 and 1.74 for pre-1992 HEIs without medical schools and GuildHE, respectively, and for that matter the minima values for GuilHe in Table 11, should be treated with caution and can be the result of inconsistent year on year data reporting as changes in productivity of that level within 2 years would be unlikely.

Turning to the components that make up the productivity change, Table 11 presents also descriptive statistics for the efficiency changes over the 2-year period modelled and also for the boundary shift. Note that we are presenting efficiency change relative to the CRS boundary—not the VRS boundary that we used earlier. (See Grifell-Tatjé and

Table 11 Summary of the results on productivity change

	Min	*Q1*	*Median*	*Geometric mean*	*Q3*	*Max*
Malmquist index						
Pre-1992 without medical schools ($N=32$)	0.78	0.91	0.98	0.99	1.04	1.34
Pre-1992 with medical schools ($N=18$)	0.78	0.90	0.94	0.95	1.01	1.12
Post-1992 Universities ($N=33$)	0.87	0.95	0.98	1.00	1.07	1.14
GUILDHE colleges ($N=38$)	0.05	0.88	0.94	0.89	1.06	1.74
Efficiency change						
Pre-1992 without medical schools ($N=32$)	0.68	0.95	1	0.99	1.01	1.38
Pre-1992 with medical schools ($N=18$)	0.92	1.00	1.00	1.01	1.05	1.08
Post-1992 Universities ($N=33$)	0.83	0.91	0.97	0.96	1	1.13
GUILDHE colleges ($N=38$)	0.06	0.97	1.00	0.94	1.07	1.38
Boundary shift						
Pre-1992 without medical schools ($N=32$)	0.91	0.96	0.98	1	1.02	1.34
Pre-1992 with medical schools ($N=18$)	0.81	0.91	0.93	0.94	0.97	1.12
Post-1992 Universities ($N=33$)	0.93	1	1.04	1.04	1.08	1.18
GUILDHE colleges ($N=38$)	0.81	0.89	0.93	0.95	0.97	1.32

Lovell (1995) and Herrero and Pascoe (2004) about the bias introduced in measuring productivity change using VRS technology specifications.) The efficiency change values in Table 11 reflect whether each HEI has moved closer to or further from *MPSS* for its output mix over the 2 years rather than closer to the boundary given its scale *size*. Given that most values are around 1, we find that there has been little change in distance from the MPSS at HEI level, the exception being GuildHE HEIs that show a considerable range of changes in distance from the MPSS. Of the remaining HEIs, a large number of post-1992 HEIs appear to have moved somewhat further from MPSS in 2002/2003 compared to 2000/2001 as suggested by the quartile 3 value of 1 and a median value of 0.97 for the efficiency change component.

Finally, the bottom third of Table 11 shows whether the MPSS at each HEI's mix of outputs has moved to a more or less productive position in the form of 'boundary shift'. Here we have a clear tendency for the boundary of post-1992 HEIs to have become more productive both over time and relative to the other types of HEI. That is to say the *most efficient* of the post-1992 HEIs, which are the ones that define the boundary and operate at local CRS, have improved productivity over time, more so than have the corresponding efficient HEIs in the remaining three types of HEI. In contrast, generally for the other three types of HEI, the most efficient HEIs in each case are less productive in 2002/2003 compared to 2000/2001 as can be deduced from the median values of 0.93 and 0.98. Indeed, pre-1992 HEIs with medical schools and GuildHE HEIs have quartile 3 values of 0.97 suggesting that 75% of the boundary projection points are less productive in 2002/2003 than 2000/2001, the projections being on the CRS boundary.

In sum, over the 2-year period that we have analysed, we find gains in productivity for a considerable minority but not a majority of HEIs. More specifically, the percentage of HEIs that show overall productivity gain are as follows: pre-1992 HEIs with medical school 28%, pre-1992 HEIs without medical school 45%, post-1992 HEIs 40% and GuildHE colleges 33%. Further, the results show that most HEIs keep up with their efficient boundary, but that boundary generally became less productive over our period of study, the exception in this being post-1992 HEIs in which the mix of outputs appears to have shifted to more productive configuration over time for most HEIs. Note, however, that these results should be interpreted with caution given the short time period covered by the study.

Conclusions

Our analysis based on DEA reaffirms the conclusion of Johnes *et al* (2008) that the higher education sector in England cannot be analysed as a unitary set. Evidently, using more homogeneous subsets of institutions by objectives and operating environment will lead to more reliable and robust results. DEA provides estimates of subject-specific unit costs that are in general similar to parametric estimates of those same unit costs provided the institutions have a truly multi-product profile. Where institutions have specialised output profiles so that certain institutions produce only certain outputs, then DEA appears to offer better unit cost estimates because of the flexibility (piece-wise linear) in the 'cost function' that it actually fits to the data.

Besides comparing the results of DEA and parametric methods, we have examined potential cost savings and output augmentations in different HEI groups using various DEA models. Interestingly, our analysis shows that there is substantial scope for gains in student numbers at no additional cost, if all efficiency gains are directed to raising student numbers, permitting each HEI to raise numbers in areas in which it has itself the largest scope for gains. It must be recalled that the efficiency gains estimated here are relative to the best observed performance among the HEIs in the comparative set used. Further gains may be possible in absolute terms, but these can only be identified by going beyond observed practice reflected in the comparative data used.

The reported results are mainly based on static DEA models, which assume that the technology of delivering education over the 3 years concerned has not changed (progressed or regressed), in the sense that if a cost level could support a given bundle of outputs in 1 year it could do so in any one of the 3 years. To allow technology or the efficient boundary to vary in different years, we also used DEA to calculate the Malmquist index of productivity change that enables one to measure productivity change and decompose it further into efficiency change and boundary shift components. An interesting finding was that, with the exception of post-1992 institutions, the efficient boundary became less productive during the sample period. Nevertheless, average changes in productivity and its components at the group level have not been large.

Although we ran our assessments using four distinct HEI groups, one should recall that there is still some heterogeneity within these groups that can affect the results presented. However, with the data set used in the paper it was not possible to use smaller and more uniform subgroups due to the lack of cross-sectional observations and the short time period. In future work, data for a longer run of years could be used. A longer data panel would offer the possibility of investigating factors such as subject mix and scale size associated with higher productivity growth rates, which can be disseminated for the benefit of the sector. Future work could also address the issue of quality of teaching so that both quantity and quality of teaching are reflected in the assessment, provided the necessary data to capture teaching quality (eg student outcomes on exit and quality of students on entry) would be available.

Further research could also extend methodologies used here in at least two different ways. First, one could explore the determinants of inter-institutional differences in efficiency by looking at potential explanatory factors such as staff–student ratios, administrative structures and other academic policy parameters in the way the institutions function. Second, it would be potentially fruitful to employ recently developed semi- and non-parametric SFA techniques to higher education, as these methods have not yet been applied in this area. In particular, it would be interesting to apply the 'stochastic nonparametric envelopment of data' (StoNED; see Kuosmanen and Kortelainen, 2007), which allows a non-parametric functional form for the cost function and is therefore more flexible than parametric SFA. As the method combines the main characteristics of DEA and SFA in a unified framework, it could provide an important benchmark for the DEA and SFA results reported here and in Johnes *et al* (2008). Last, but not least, our findings on unit costs, efficiencies, targets and productivity change are naturally specific to the data set we have used. We have used this particular data set because it facilitates comparison between results from parametric (Johnes *et al*, 2008) and non-parametric methods. Analyses of this type need to be updated as new data become available over time both to test the stability of the findings and monitor the performance of HEIs over time.

References

Agasisti S and Salerno C (2007). Assessing the cost efficiency of Italian universities. *Educ Econ* **15**: 455–471.

Aigner DJ, Lovell CAK and Schmidt P (1977). Formulation and estimation of stochastic frontier production function models. *J Econometrics* **6**: 21–37.

Andersen P and Petersen NC (1993). A procedure for ranking efficient units in data envelopment analysis. *Mngt Sci* **39**: 1261–1264.

Athanassopoulos AD and Shale E (1997). Assessing the comparative efficiency of higher education institutions in the UK by the means of data envelopment analysis. *Educ Econ* **5**: 117–134.

Banker R and Natarajan R (2004). Statistical tests based on DEA efficiency scores. Chapter 11. In: Cooper WW, Seiford L and Zhu J (eds). *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers: Norwell, MA, pp 299–321.

Banker R, Charnes A, Cooper WW and Rhodes E (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Mngt Sci* **30**: 1078–1092.

Beasley J (1995). Determining teaching and research efficiencies. *J Opl Res Soc* **46**: 441–452.

Beasley J and Wong Y-HB (1990). Restricting weight flexibility in data envelopment analysis. *J Opl Res Soc* **41**: 829–835.

Caves DW, Christensen LR and Diewert WE (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica* **50**: 1393–1414.

Charnes A, Cooper WW and Rhodes E (1978). Measuring the efficiency of decision making units. *Eur J Opl Res* **2**: 429–444.

De Groot H, McMahon WW and Volkwein JF (1991). The cost structure of American research universities. *Rev Econ Stat* **73**: 424–431.

Färe R, Grosskopf S, Lindgren B and Roos B (1994a). Productivity developments in Swedish hospitals: A Malmquist output index approach. In: Charnes A, Cooper WW, Levin AY and Seiford LM (eds). *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer Academic Publishers: Boston, MA, pp 253–272.

Färe R, Grosskopf S, Norris M and Zhang Z (1994b). Productivity growth, technical progress and efficiency change in industrialized countries. *Am Econ Rev* **84**: 66–83.

Farrell MJ (1957). The measurement of productive efficiency. *J R Stat Soc Ser A* **120**: 253–290.

Flegg AT, Allen DO, Field K and Thurlow TW (2004). Measuring the efficiency of British universities: A multi-period data envelopment analysis. *Educ Econ* **12**: 231–249.

Grifell-Tatjé E and Lovell CAK (1995). A note on the Malmquist productivity index. *Econ Letters* **47**: 169–175.

Herrero I and Pascoe S (2004). Analysing the effect of technical change on individual outputs using modified quasi-Malmquist indexes. *J Opl Res Soc* **55**: 1081–1089.

Izadi H, Johnes G, Oskrochi R and Crouchley R (2002). Stochastic frontier estimation of a CES cost function: The case of higher education in Britain. *Econ Educ Rev* **21**: 63–71.

Johnes G and Johnes J (2009). Higher education institutions' costs and efficiency: Taking the decomposition a further step. *Econ Educ Rev* **28**: 107–113.

Johnes G, Johnes J and Thanassoulis E (2008). An analysis of costs in institutions of higher education in England. *Stud High Educ* **33**: 527–549.

Johnes J (2008). Efficiency and productivity change in the English higher education sector from 1996/97 to 2004/05. *Manch Sch* **76**: 653–674.

Johnes J and Johnes G (1995). Research funding and performance in UK university departments of economics: A frontier analysis. *Econ Educ Rev* **14**: 301–314.

Kuosmanen T and Kortelainen M (2007). *Stochastic nonparametric envelopment of data: Cross-sectional frontier estimation subject to shape constraints*. Discussion Paper No: 46, Department of Economics and Business Administration, University of Joensuu.

Lovell CAK (2003). The decomposition of Malmquist productivity indexes. *J Prod Anal* **20**: 437–458.

Meeusen W and van den Broeck J (1977). Efficiency estimation from Cobb-Douglas production function with composed error. *Int Econ Rev* **8**: 435–444.

Podinovski V (2004). Bridging the gap between the constant and variable returns-to-scale models: Selective proportionality in data envelopment analysis. *J Opl Res Soc* **55**: 265–276.

Simar L and Wilson PW (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In: Fried H, Lovell CAK and Schmidt S (eds). *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press: New York.

Stevens PA (2005). A stochastic frontier analysis of English and Welsh universities. *Educ Econ* **13**: 355–374.

Thanassoulis E (1996). A data envelopment analysis approach to clustering operating units for resource allocation purposes. *Omega: Int J Mngt S* **24**: 463–476.

Thanassoulis E (1999). Setting achievement targets for school children. *Educ Econ* **7/2**: 101–119.

Thanassoulis E (2001). *Introduction to the Theory and Application of Data Envelopment Analysis: A Foundation Text with Integrated Software*. Kluwer Academic Publishers: Boston.

Thanassoulis E and Dyson RG (1992). Estimating preferred input-output levels using data envelopment analysis. *Eur J Opl Res* **56**: 80–97.

Thanassoulis E, Portela MCAS and Allen R (2004). Incorporating value judgments in DEA. In: Cooper WW, Seiford LW and Zhu J (eds). *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers: Norwell, MA, pp 99–138.

Thanassoulis E, Portela M and Despic O (2008). The mathematical programming approach to efficiency analysis. In: Fried H, Lovell K, Schmidt S (eds). *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press: New York pp 251–420.

Verry DW and Layard PRG (1975). Cost functions for university teaching and research. *Econ J* **85**: 55–74.

Appendix

Mathematical presentation of DEA under VRS and CRS

The original DEA model of Charnes *et al* (1978) assumes constant returns to scale (CRS) under which the DEA-derived input- and output-oriented measures of efficiency for DMU are identical. The CRS assumption can be relaxed and the DEA model can be easily modified to incorporate variable returns to scale (VRS) (Banker *et al*, 1984). The set of DMUs identified as inefficient under VRS will be the same whether an input- or output-oriented approach is taken. In contrast to the CRS framework, however, the actual values of the efficiency scores for the inefficient DMUs vary with the orientation adopted.

In practice, DMUs may produce many outputs from their resources, in which case programming techniques have to be used to identify the piecewise linear frontier joining up all efficient DMUs. Suppose DMUs use m inputs to produce s outputs. Under VRS the following linear programming problem must be solved for each of the n DMUs ($k=1, \ldots, n$):

Output-oriented (VRS):	*Input-oriented (VRS):*
Maximise ϕ_k (A.1) Subject to	Minimise θ_k (A.2) Subject to
$\phi_k y_{rk} - \Sigma_{j=1}^{n} \lambda_j y_{rj} \leqslant 0 \; r=1, \ldots, s$	$y_{rk} - \Sigma_{j=1}^{n} \lambda_j y_{rj} \leqslant 0$ $r=1, \ldots, s$
$x_{ik} - \Sigma_{j=1}^{n} \lambda_j x_{ij} \geqslant 0 \; i=1, \ldots, m$	$\theta_k x_{ik} - \Sigma_{j=1}^{n} \lambda_j x_{ij} \geqslant 0$ $i=1, \ldots, m$
$\Sigma_{j=1}^{n} \lambda_j = 1, \; \lambda_j \geqslant 0 \; \forall_j = 1, \ldots, n$	$\Sigma_{j=1}^{n} \lambda_j = 1, \; \lambda_j \geqslant 0$ $\forall_j = 1, \ldots, n$

Overall efficiency of DMU k is measured by $E_k = 1/\phi_k$ in the output-oriented framework or $E_k = \theta_k$ in the input-oriented framework. Further, scale efficiency can be identified by calculating the following ratio for DMU k:

$$SCE_k = \frac{E_{k,CRS}}{E_{k,VRS}}, \qquad \text{(A.3)}$$

where the numerator and denominator include efficiency scores calculated under CRS and VRS, respectively. Note that the CRS efficiency score can be calculated simply by deleting the constraint $\Sigma_{j=1}^{n} \lambda_j = 1$ from model 1 (A.1) or (A.2).

Received October 2008;
accepted April 2010 after two revisions