Efficiency, costs, rankings and heterogeneity: the case of US higher education

Tommaso Agasisti[a]* and Geraint Johnes[b]

[a]*Politecnico di Milano, Department of Management, Economics and Industrial Engineering, p.za Leonardo da Vinci, 32, 20133 Milan, Italy;* [b]*Lancaster University Management School, Lancaster LA1 4YX, UK*

Among the major trends in the higher education (HE) sector, the development of rankings as a policy and managerial tool is of particular relevance. However, despite the diffusion of these instruments, it is still not clear how they relate with traditional performance measures, like unit costs and efficiency scores. In this paper, we estimate a variety of models to evaluate costs in US higher education institutions. A particularly innovative feature of our approach involves the estimation of latent class and random parameter stochastic frontier models of a multiproduct cost function: this allows us fully to accommodate both the heterogeneity across institutions and the presence of technical inefficiencies. Such methodological strategy is essential in analyzing the US HE system, which is characterized by a strong internal differentiation. Our main findings are two. First, on a public policy ground, the estimates suggest that global economies could be achieved by effecting a reduction in the number of institutions providing undergraduate instruction, while increasing the number of institutions engaged in postgraduate activity. Second, the current existing rankings turn out as coherent with ratings provided by the calculation of efficiency scores.

Keywords: costs; efficiency; rankings; stochastic frontier; latent class; random parameter

1. Introduction

Higher education institutions (HEIs) are complex organizations, characterized as they are by their multi-product nature.[1] This feature, whereby institutions jointly produce teaching and research across different subjects and different levels renders evaluation of their performance difficult, since an institution that performs well in one dimension may fare worse in another. Arguably this is even more the case for higher education (HE) than for other public entities – the assessment of which already poses methodological challenges (Gutiérrez-Romero, Haubrich, and McLean 2008). At the same time, despite these difficulties, judging the performance of HEIs is highly desirable and relevant in the policy context: they typically receive a large amount of public funding, and they provide crucial services for students, families, and the wider public interest.

Unsurprisingly in this context, rankings of institutions have developed that are used both to inform the (quasi-) market and to influence policy. In many countries,

*Corresponding author. Email: tommaso.agasisti@polimi.it

independent agencies provide a list of indicators to assess institutions' quality, prestige, productivity and other relevant dimensions (Dill and Soo 2005; EUA 2011; Hazelkorn 2011). The rankings produced by agencies such as the Academic Ranking of World Universities (ARWU) or Times Higher Education have become a widely cited source of information about the 'best universities in the world': public agencies and governments use them in public debates, and students and their families as a basis for their decisions about educational investments.

However, such rankings are sensitive to the methods used in their construction (Butler 2007; Jeremic et al. 2012), and it is appropriate therefore to assess how these rankings provide information that is congruent with more objective measures of the performance of HEIs. In posing this question, we shall focus on measures such as unit costs and efficiency scores.

In particular, we start by asking the following question: do similar universities have the same cost structure? This is a crucial question for both theoretical and practical reasons. The policy implications of the answer are important for policy makers, especially in periods when pressures on public funds call for reflection on the organization and financing of such strategic sectors as HE. For instance, different cost functions imply different scale and scope effects, which can justify the application of different policies to clusters of institutions in the sector, and hence the application of funding models that make allowance for structural differences across types of institution. An example is the efficiency of institutions in using available resources matters for decisions about the distribution of public funds. Moreover, if the cost functions attached to different institutions are different, then this has implications for the evaluation of their performances, including efficiency, expenditures and other dimensions usually included in the indicators used for rankings.

This topic is particularly relevant with reference to the HE system in the USA, where there is a multitude of institutions of HE, with different missions, activities, strategies, and organizational structure. Most importantly, this problem has been only partially faced in some previous empirical literature. For instance, De Groot, McMahon, and Volkwein (1991, 424) state that they 'account for research productivity differences between institutions'. Harter, Wade, and Watkins (2005, 371) study the costs of four-year US colleges, and acknowledge that 'expenditures per student showed significant variance across institutions having the same mission and control'. Sav (2004) has conducted a specific analysis to check whether there is a difference between the cost structure of public and private universities, finding evidence of marked differences. Nevertheless, heterogeneity amongst US HEIs is well established; see, for example, Hoxby (1999). Despite the awareness of the structural heterogeneity of the US HE system, no study has fully accounted for it in previous academic literature about HEIs' performances and costs, and this paper attempts to fill this gap.

Heterogeneity across HEIs has been a major focus of research also in Europe. In this context, universities have often been deemed to be similar in character to one another, and a strong differentiation across institutions (typical of the US system) does not occur. In a break with this tradition, G. Johnes et al. (2005) study the HE sector in UK, and they detect significant differences across different subsectors. However, it is only recently that the problem of correctly identifying the heterogeneity across universities' cost functions has been systematically tackled using newly developed estimation technology. The idea proposed by G. Johnes and Johnes (2009) is that heterogeneity occurs not only (or not necessarily) between universities in different subsectors, but also it is possible between universities that belong to the

same subsector of HE. This analysis has been made possible by using the new techniques proposed by Tsionas (2002) and Greene (2005); these involve a random parameter specification of models, which will be discussed later in this paper. After this study, similar analyses were conducted for other European countries: Italy (Agasisti and Johnes 2010), Spain (G. Johnes and Salas-Velasco 2007), and Germany (G. Johnes and Schwarzenberger 2011). In all these countries, HEIs are conventionally thought of as being very similar to one another, but the analysis showed that relevant differences exist and are in some cases substantial. This is evidenced also by a recent study conducted by Bonaccorsi and Daraio (2007), focusing on different strategies pursued by European universities. These authors conceptualized university strategy as '(…) an emergent pattern of configuration of university outputs that depend on (relatively) autonomous decision making by universities, supported by appropriate combinations of resources (inputs)' (Bonaccorsi and Daraio 2007, 11).

Coming back to the main focus of this paper, the HE sector in the US is traditionally characterized by a perception of strong differences across institutions. Part of the difference is due to variation in unit costs. Harter et al. (2005) conducted an analysis of US four-year public colleges for the period 1989–1998, showing a great variation of cost per student – where students have been measured on a FTE (full-time equivalent) basis. This analysis showed that, for colleges with an enrollment of between 3000 and 10,000 FTEs, the range of real unit costs in 1989 was between $8144 (for comprehensive public colleges without a medical school) and $17,538 (for research colleges with medical schools). Another source of difference is in the output mix, that is the type of outputs produced: some universities are more focused on research, others on undergraduate teaching, others on postgraduate teaching. Since 1970 the Carnegie Foundation, aware of such heterogeneity in the mission of HEIs, has proposed a classification of colleges according to different dimensions (http://www.carnegiefoundation.org/classifications/): Undergraduate Instructional Program, Graduate Instructional Program, Enrollment Profile, Undergraduate Profile, Size and Setting. A basic classification is also proposed considering the following categories of colleges: associates' colleges, doctorate-granting universities, masters' colleges and universities, baccalaureate universities, special focus institutions, and tribal colleges. Differences across institutions (in terms of performance, output levels and reputation) have become critical in determining the choice of students and families, not least because of the proliferation of public rankings of institutions such as the US News and World Report. The 2009 Report states: '(…) the rankings provide an excellent starting point because they offer the opportunity to judge the relative quality of institutions based on widely accepted indicators of excellence. (…) Schools are categorized by mission, based on the basic Carnegie classification and, in some cases, by region' (US News and World Report 2009, 82). It is reasonable to assume that such an accepted heterogeneity across institutions could impact also on their cost structure – as a pure analysis of unit costs indeed confirms. What is less clear is how much the evaluations based on costs, efficiency (the ability to reduce unit costs while producing a given amount of output), or alternative performance dimensions are congruent with each other: inconsistency across these different measures can raise doubts on their usefulness in terms of policy and managerial use. Of particular relevance is whether rankings are able to reflect also differentials in the costs sustained by universities for producing their outputs; indeed, a fair performance evaluation should take this potential discrepancy into account.

The goal of this paper is to analyze the cost structure of the HE sector in the USA. As discussed above, it represents a particularly interesting case study, both because of the large number of universities and colleges that are operating in the market and the huge differences between them – in terms of dimension, types of education and research provided, ownership, subject mix, and so on. We pursue this first objective in a comprehensive and innovative way, by using empirical approaches that exploit institutional heterogeneity in the calculation of the parameters of the cost function (section 4). Also, the methodology adopted in this paper allows us jointly to analyze costs and efficiency. The methods we employ also exploit the panel nature of the dataset, thus enabling us to make allowance for differences between institutions that cannot be captured by the data. This is achieved by introducing a shift parameter (and, more generally, other parameters) that may vary across institutions but which, for each institution, must be fixed over time. Hence, costs in each institution are allowed to vary for reasons that are not observed, but which are associated with that institution in each period of time. This has the considerable advantage of allowing for differences, for example, in quality of provision across institutions. We use two approaches: random parameter and latent class models; the former offering greater allowance for heterogeneity than the latter. This is the straightforward way to proceed, as the cost function represents the boundary that describes the lowest cost at which it is possible to produce a given vector of outputs. In this context, frontier estimations permit us to describe the cost structure and to compute efficiency scores simultaneously. We also compare the efficiency scores obtained with rankings provided by a well-regarded US magazine, namely US News and World Report, which is one of the major sources of information for prospective HE students in the USA. For the reasons discussed above, testing the coherence between efficiency-based evaluations and public rankings of HEIs can help in better judging the robustness of the latter.

The paper is organized as follows. In the next section, a literature review is provided. Section 3 describes the background for the subsequent analysis. Section 4 concerns the data and methodology used for this study. Section 5 presents the results and a discussion. Section 6 concludes.

2. Literature review

Recent theoretical work has shown that universities, in a context of limited resources and competition, tend to specialize (Del Rey 2001), and that the equilibrium of the HE market converges towards different steady states, where mass production of both research and learning is just one of four possible strategies of universities (the others are: full-time teaching, full-time research, selective teaching and research). Similar results have been obtained by De Fraja and Valbonesi (2008). In light of these findings, it is natural to suppose that a number of different types of HEI might exist, and that these should be characterized by distinct cost structures.

The empirical investigation of the cost structure of universities is not a new theme in the USA. Early work is surveyed by Brinkman and Leslie (1986), and points to the existence of widespread scale effects. The first paper to introduce the concept of universities as multiproduct organizations, and hence to focus upon more sophisticated measures of economies of scale and scope, is Cohn et al. (1989). After this pioneering work, several contributions aimed to provide further empirical evidence about the cost structure of the HE sector.

In general terms, the approach adopted is to estimate a cost function of the following type:

$$TC_i = f(x_i), \tag{1}$$

where TC_i is the total cost for the ith university, and x_i is a vector of outputs. It is widely recognized that universities typically produce three kinds of output: teaching, research and social service (the last of these typically taking the form of knowledge transfer, organization of cultural events, consultancy, and the like)[2,3].

The contribution of Cohn et al. (1989) employs a multi-output cost function measuring both teaching and research outputs. To be specific, their study measures full-time-equivalent enrolments as teaching output, and uses measures of research grant income as a proxy for research output. Faculty salaries are included as a price factor. Their sample comprises 1887 HEIs. The results of this pioneering study are that: (i) institutions can benefit from scope economies by raising the output level; (ii) comprehensive institutions are less costly than specialized ones; and (iii) very small institutions are more costly than average ones. All of these findings suggest that an important role is played by scale economies in the HE sector.

De Groot, McMahon, and Volkwein (1991) considered 147 doctorate-granting universities. The variables considered are FTE enrolments of undergraduate and graduate students, and the number of research publications. Their results indicate the presence of sizeable economies of scale for the average institution, as well as scope economies associated with the joint production of undergraduate and graduate education. At the same time, they did not find evidence of any significant impact of ownership and state regulation on costs. The results have been validated also through a sensitivity analysis by substituting enrolments with degrees awarded.

Koshal and Koshal (1999) focus on comprehensive institutions. Their sample contains 158 private and 171 public HEIs. FTE enrolments and dollars spent for research activity are used as indicators for teaching and research output. Their analysis is unusual in that they include a measure of quality for the teaching and learning domain, that is the average total scores on the Scholastic Aptitude Test (SAT) of entering freshmen.[4] Product-specific economies of scale are detected for undergraduate education but are absent in the case of graduate education; the authors also find that economies of scope remain unexploited.

Laband and Lentz (2003) build upon the Cohn et al. (1989) framework, estimating cost functions for 1492 private and 1450 public HEIs. They use enrolments and externally funded research as proxies for teaching and research output, respectively. The results indicate that the cost structures of private institutions and public institutions differ significantly from one another. There are, however, two common features across these two sectors: the presence of economies of scale, and diseconomies of scope. The straightforward policy implication of such findings is that global unit costs could be reduced by increased specialization of institutions, while raising output levels in each university.

Sav (2004) conducted a study on an extensive sample of 2189 universities and colleges. In his cost function, he includes *inter alia* a wage variable as a factor price, and a dummy for the presence of a medical school. Teaching output is captured by a measure of teaching hours, for undergraduate, graduate and professional courses; research output has been measured through the research grants. The author reports ray economies of scale in the private, but not in the public, sector. The private sector also benefits from economies of scope. Moreover, both private and public universities

show product-specific economies of scale attached to research activity. The paper derives some strong and important policy implications: that '(…) instead of trying to be everything to everyone or all taxpayers, large state-supported research institutions might move to more specialized production' (Sav 2004, 613). Also regional differences are detected, by suggesting a possible differentiation of cost functions due to the socio-economic differences of the regions (in particular, the wage factor could be affected by such differences).

Laband and Lentz (2004) devote further attention to the differences between private and public sectors, by analyzing a sample of more than 2700 universities and colleges. Also this paper uses FTE enrolments as teaching output and research grants as research output, and a measure of faculty compensation has been included as a price factor. The results confirm the hypothesis that the two subsectors actually are quite different in terms of the cost structure of the institutions. More specifically, '(…) the public IHEs (Institutes of Higher Education) produce more cheaply than the private, not-for-profit IHEs at almost all scales of output' (Laband and Lentz 2004, 438).

Overall, previous studies about US universities acknowledge the differences across institutions in terms of production costs, but suffer two major limitations: (i) they face this problem by estimating cost functions for separate groups of universities, or by including dummies for specific cost-related factors (such as medical schools), but failed to develop an empirical strategy that permits study of the cost structure of all universities simultaneously accounting for heterogeneity; and (ii) these studies did not consider the relationship between cost structures and performances (measured by efficiency, productivity or quality – i.e. rankings). Our study points directly at over-coming these two gaps in the existent literature.

3. Background

The use of random parameters for modeling the cost function of universities is quite recent, even though it was already utilized for empirical analyses in UK (G. Johnes and Johnes 2009), Italy (Agasisti and Johnes 2010), Spain (G. Johnes and Salas-Velasco 2007), and Germany (G. Johnes and Schwarzenberger 2011). The idea behind these studies is that colleges tend to be different, and so they each face a cost function that is distinct. In the same spirit, recent work by Bonaccorsi and Daraio (2007) demonstrates that, even if is there such a presumption of homogeneity, universities already are very different in terms of productivity, performance and efficiency. More specifically, they developed the idea of university 'strategy', whereby each university tends to specialize (voluntarily or not) in the activity in which it has a competitive advantage.

In the US it is widely accepted that universities behave in a different manner from one another, with some well-defined segments of the market leading to a number of 'types' of institution, each of which specializes to a greater or lesser extent on the provision of certain types of output. A clear source of heterogeneity in institutions' behavior is in the specialization towards teaching or research; moreover, within teaching, institutions might specialize in undergraduate or graduate activities.

While technology might explain how specialization can lead to allocative efficiencies, it is difficult in a world where HEIs face homogeneous cost structures to see how some institutions would choose to specialize in the production of some outputs while others do otherwise. A possible way out of this conundrum is to acknowledge the possible existence of heterogeneity in the cost structure of universities. Different institutions

are endowed with different characteristics, and these matter when we consider how costs are determined.

In much of the received literature, cost functions have been estimated in a parametric setting that *assumes* that all institutions face the same cost function. Yet we know from theory, and from studies that have started from the identification of prescribed categories of institution, that they do not. Put another way, the economic literature has pointed out that we should expect there to be heterogeneity across institutions, and that this should be captured in the institution-specific effects associated with their operations. So our models should allow for the possibility that different institutions have different cost functions. A fairly crude way of representing this is to use fixed-effects models, which incorporate in the cost function estimates some 'unobservable' structural differences across institutions. However, the recent development of estimators proposed by Greene (2005) and Tsionas (2002) allow us to identify variation not only in the constant, but also in the remaining parameters of the cost equation, so that each HEI faces an entirely different cost function – all within the context of a frontier model. This approach resembles in key respects the nonparametric approach of data envelopment analysis (DEA), where the weights, attached by an institution to each input and output, are allowed to differ from those applied by other institutions. It also offers the considerable virtue that inter-institutional differences in factor prices (due, perhaps, to location) are captured by allowing the parameters of the model to vary across institutions following a random parameter approach, and thus follow a specified distribution rather than – as in DEA – being determined by a very small number of observations on each HEI. The random parameters approach is well explained by Tsionas (2002, 128): '(…) production possibilities are expected to differ in a cross-section of firms, and a set of different technologies may simultaneously coexist at any given time. If that is the case, efficiency measurement cannot proceed under the assumption of common technology. (…) The relative difference in output reflects technological differences, not inferior practice.'

This argument is particularly convincing for the HE sector. While it is reasonable to assume the same *functional form* for the cost function of all institutions, it is less reasonable to impose the assumption that the *parameters* are constant across institutions. Use of panel data allows a random parameter specification to be modeled. From a methodological point of view, such an approach benefits from generality in the ability of modeling heterogeneity – in other words, 'many of the models already considered (such as fixed-effects or random-effects specifications) are special cases' (Greene 2005, 288). The assumption of the same technology for all institutions would, by way of contrast, result in a '(…) confusion between technological differences and technology-specific inefficiency' (Tsionas 2002, 128).

We expect that there do exist relevant technological differences and behavioral differences across colleges, and that these affect their production processes of teaching and research, so we estimate cost functions using a random parameter approach, and we compare the results with other, more traditional, approaches (the traditional frontier model and latent class models). Also, we investigate if these efficiency estimations are correlated with ratings included in the rankings of university quality. There are two byproducts of the described approach: (i) testing differences in the estimates of efficiency provided by different methods (which is important for policy-oriented reflections about the robustness of performance evaluation methods); and (ii) explicitly comparing efficiency-based judgments with different ones (rankings) mainly based on subjective and quality-based assessments.

4. Data and methodology

In this paper, the estimation of efficiency is focused upon the cost function, rather than the production function, because HEIs potentially produce several outputs at the same time; also, the use of a cost function allows the straightforward calculation of unit costs, which is another typical indicator often used as an (indirect) measure of performance.

All the data come from the Integrated Postsecondary Education Data System (IPEDS) dataset, provided by the National Center for Educational Statistics (NCES), and refer to three academic years: 2003–4, 2004–5, and 2005–6. All the financial variables are collected on a financial year basis, and they are matched with the academic year that starts in the same year. Hence costs in year t have been matched with students in the academic year that straddles t and t + 1.

Our sample comprises all the universities classified as four (or more) year degree-granting institutions, both public and private, making a total of 2318 institutions. For the purposes of empirical analysis, however, we drop all the observations for which there are missing data for any variables in any one of the three years considered. This results in dropping 1364 institutions, and the final sample includes 954 institutions for which we have three complete years of data, that is a total of 2862 observations.

The dependent variable (COSTS) is defined as total amount of expenses. 'Expenses may be thought of as the costs of goods and services used to produce the educational services provided' (IPEDS glossary definition).[5]

The independent variables related to the teaching and learning activities are defined as follows:

- the number of bachelor (b) degrees. A qualification at bachelor level represents the first completed level of HE, which includes degrees obtained after four or (less usually) five years of study;
- the number of postgraduate (p) degrees. This includes the number of students who obtained a first professional degree (in, for example, medicine), a masters degree, or a doctorate.

These qualifications (p) are clearly different from one another – first professional degrees are somewhat akin to unusually long bachelors programs, taking students from entry to HE through exit at higher degree level. Masters programs are typically of relatively short duration – one or two years. Doctorates, meanwhile, combine a taught component with a requirement for substantive research activity, and typically take several years to complete. In early work, we separated out these various types of qualification, but as a result of multicollinearity problems we have merged them into a single 'postgraduate' variable for the purposes of the present paper. While many studies use the number of students (a stock) as an indicator for teaching activity, here we use the number of graduates (an outflow) because it better represents the final output of the teaching. In comparing the results provided by our study with other previous contributions, this difference should be kept in mind. Our choice of output variables is driven in part by the availability of data, and we recognize that alternative measures of teaching output exist. For example, Sullivan et al. (2012) have recently proposed measures that include the number of credit hours passed.

As a proxy for research activities we use the value of grants received for conducting research. More precisely, this variable is the sum of 'all operating expenses associated with activities specifically organized to produce research outcomes, and commissioned by an agency either external to the institution or separately budgeted by an

organizational unit within the institution' (IPEDS website: http://nces.ed.gov/ipeds/). In the case of sponsored research, the institution may receive funding only on production of evidence that there has been expenditure, so there may be a particularly direct relationship between one of our explanatory variables and the dependent variable. Of course it is *desirable* that there should be a behavioral relationship between the explanatory and dependent variables, so this is not a particular concern. To the extent that research income captures *only* sponsored research activity of this kind, we would expect the linear term in research to have a coefficient of unity and the quadratic and interaction terms to be insignificantly different from zero. As is readily observed in the results that follow, this is not the case, and we can infer from this that our research variable is effective in capturing activity that is not (solely) sponsored research. Moreover, the issues surrounding the use of this type of indicator as a proxy for research output are well known and have been extensively debated in the literature (De Groot, McMahon, and Volkwein 1991; G. Johnes and Johnes 1993); while grants may be regarded as an input, they offer several advantages as a measure of research activity. Notably they provide a 'quality-adjusted' measure of the volume of research that is done, and they offer a contemporaneous measure that has advantages over more retrospective alternatives such as citations and publication counts. Grants have therefore been used to proxy research in a number of earlier studies about US HEIs, amongst which are Cohn, Rhine, and Santos (1989), Laband and Lentz (2004) and Sav (2004).

Summary statistics for our sample of institutions are reported in Table 1, and clearly show evidence of a marked heterogeneity among universities – the standard deviation is higher than the mean for all the variables considered. The financial variables (costs and research) were considered in 2005 real terms, by inflating the data through the Consumer Price Index (1.0623 in 2003, and 1.0326 in 2004 – source: Bureau of Labor Statistics).

This heterogeneity can also be observed when some simple bivariate scatterplots are constructed. Figure 1 plots the number of bachelor students against the value of research grants earned by institutions in the sample. At least three patterns emerge: (1) a group of colleges focused on teaching activities, with a level of research output that is well below average; (2) a group of colleges with high levels of both teaching and research activity; and (3) a group of colleges strongly focused on research, with low levels of bachelor enrolments. Similar evidence of institutions with heterogeneity of missions derives from the plot of the number of bachelor students against

Table 1. Descriptive statistics.

	Mean	Standard deviation	Min	Max
Costs	266.917	497.024	.491	4200.00
Bachelor	1162.20	1449.35	.000	9840.00
Postgraduate	571.809	851.284	.000	6985.00
Research	37.435	101.717	.186E-03	995.00
Private	.446	.497	.000	1.000
Medical	.164	.369	.000	1.000

Notes: Costs and Research have been reported in $million.
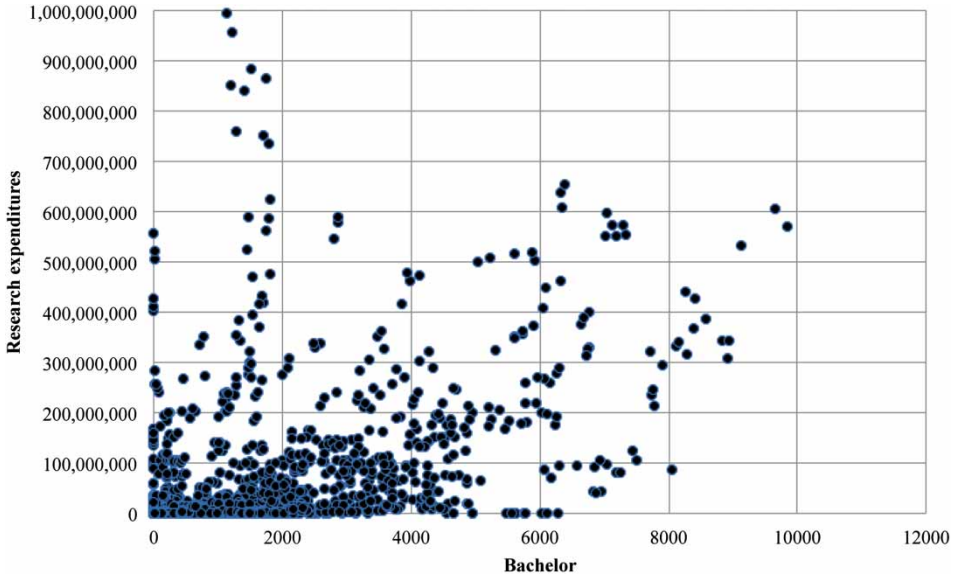
Number of observations: 2862.

Figure 1. The output mix of US colleges' – bachelor versus research expenditures.
Notes: The observations refer to all the three years (each college is included three times in the figure).

postgraduates shown in Figure 2. Again, three groups can be identified: (1) a group of colleges with a 'balanced' level of bachelor and postgraduate students; (2) a group with above-average levels of bachelors and below-average levels of postgraduates; (3) a group of colleges focused on postgraduate teaching.

Noting this heterogeneity, we proceed to employ an appropriate modeling technology, specifically using panel data techniques developed by Tsionas (2002) and Greene (2005) in the context of the evaluation of a stochastic cost frontier. The traditional frontier estimation, as developed by Aigner, Lovell, and Schmidt (1977), specifies an equation like the following (in a panel setting):

$$y_{it} = \alpha_i + \beta' x_{it} + v_{it} + u_{it}, \tag{2}$$

where v_i denotes a normally distributed residual and u_i is a non-normal residual which is supposed to capture technical inefficiency. The distribution of u_i must be specified a priori, and is usually assumed to be half-normal or exponential.

An elementary way to consider heterogeneity across institutions is to cluster observations into different groups or classes, which have some common characteristics that might explain the cost differentials. Where there are no strong a priori grounds on which to base the construction of these classes, or when there is a preference to 'let the data speak', the approach should be to estimate a latent class model (LCM) (Orea and Kumbhakar 2004). This approach divides observations, on the basis of maximum likelihood, into m classes (where m is prescribed by the analyst), and estimates distinct parameter vectors for each of the m classes. The specification of a LCM is:

$$y_{it} = \alpha_{it} + \beta_m x_{it} + v_{it,m} + u_{it,m}. \tag{3}$$

It is important to note that the calculation of efficiency is not conducted with respect to the whole sample, but conditionally on the basis of the class to which the unit belongs.
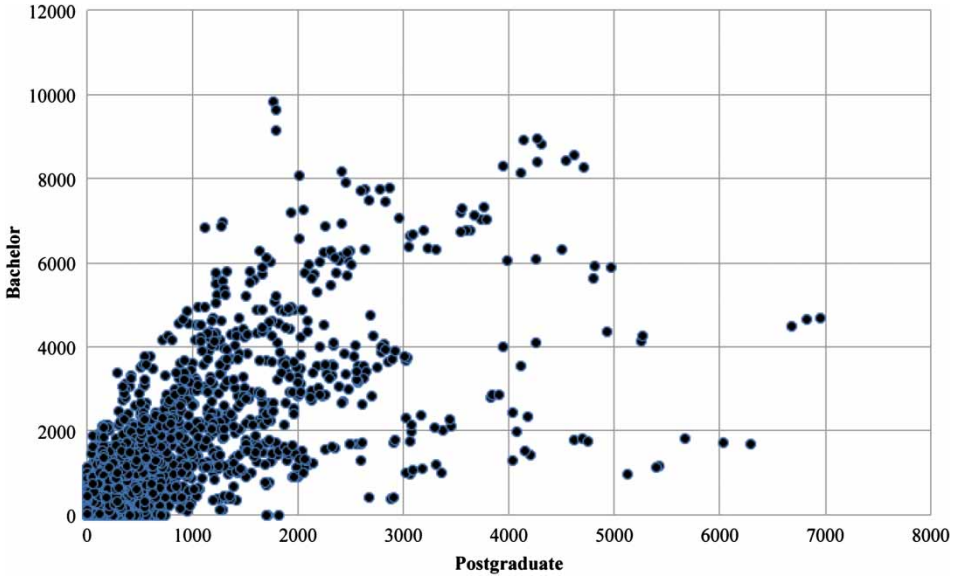
Figure 2. The output mix of US colleges – postgraduate versus undergraduate education.
Notes: The observations refer to all the three years (each college is included three times in the figure).

An intuitive description of the latent class approach may be useful at this stage. In essence, the method is based on the familiar technique of ordinary least squares (OLS) regression. OLS estimates a line of best fit through a scatterplot of data. Latent class models differ from this in that they involve a procedure whereby, using maximum likelihood methods, the observations are grouped into two or more 'latent classes', and a separate line of best fit is estimated for each of these classes. As stated above, the number of latent classes is prescribed by the analyst. This means that, in our case, institutions of a particular type – whose unobserved characteristics are similar – are grouped into the same latent class, and the estimated cost equations estimated for each class each describe the relationship of cost to a vector of outputs for similar types of institution. A second way in which the methods used here differ from OLS is that we estimate frontiers, not lines of best fit – the stochastic frontier lies below the best fit line because it makes allowance for inefficiency.

An extreme case of the LCM is one in which m equals N, the number of institutions; in this case, each institution has its own distinct cost function. Here, the vector of coefficients is random across institutions – the case of a random parameters model (RPM). In this case, parameter heterogeneity is modeled as follows:

$$\left. \begin{aligned} (\alpha_i, \beta_i) &= (\bar{\alpha}, \bar{\beta}) + \Gamma_{\alpha,\beta} w_{\alpha,\beta i} \\ \mu_i &= \bar{\mu} + \Gamma_\mu w_{\theta i} \\ \theta i &= \bar{\theta} + \Gamma_\theta w_{\theta i} \end{aligned} \right\}. \qquad (4)$$

The random variation appears as the vector $w_{j,i}$, where j denotes the constant or the slope parameter. The terms μ and θ represent the moments of the inefficiency distribution. It is possible to model the RPM either with half-normal residuals for each institution constrained to be constant across time, or with these residuals unconstrained.

An intuitive understanding of the random parameter approach involves noting that it is essentially similar to a latent class model in which there are as many latent classes as there are observations. A separate cost equation is estimated for each institution. The efficiency score obtained by an institution is defined as ratio of the observed cost to the sum of the observed cost and the value of the one-sided residual. Since, in the latent class model, a distribution of the one-sided residual is obtained separately for each of the classes, efficiency scores can, strictly speaking, be compared only within classes. In the work that we report below the residuals are unconstrained, but we note that this makes little difference to our results. In the sequel we estimate (i) a traditional frontier model; (ii) a LCM in which the number of classes is constrained to be 2; and (iii) a RPM. The results are reported and discussed in the next section.

A choice must be made about the functional form of the cost equation. The development of a theory about the functional form in the case of multiproduct organizations is due to Baumol, Panzar, and Willig (1982) who argue that the cost function of a multiproduct firm should meet a number of requirements. Firstly, it must be non-negative, non-decreasing, concave and linearly homogenous in input prices. In our analysis, we do not use input prices as an explanatory variable; earlier contributions to the literature have shown that the influence of such factor prices is typically not or weakly significant (Cohn, Rhine, and Santos 1989; Laband and Lentz 2003, 2004). We argue that allowing heterogeneity across universities' cost structures better captures such differences. Secondly, cost functions must allow sensible predictions to be made for the costs of institutions that produce zero levels of some outputs. Thirdly, the function must not be linear, because it should allow for economies of scale or scope. Following these considerations, and in line with much of the literature in this area (see, for example, Cohn and Cooper 2004), we employ a quadratic form. Mathematically, the cost function that has been estimated is:

$$y = \alpha + \Sigma \beta_i x_i + \Sigma_i \Sigma_j \varphi_{ij} x_i x_j + v + u, \qquad (5)$$

where y is the HEI's cost, and x_i and x_j are the outputs of type i and j, respectively. The quadratic terms allow for scale economies; the interaction terms allow for scope effects.

Finally, in our approach we use measures of average costs (AIC), marginal costs (MC), scale and scope effects that are coherent with the multiproduct nature of the institutions, by using the Baumol, Panzar, and Willig (1982) definitions (mathematical details are provided in Appendix 1).

5. Results

5.1. Estimates of production costs

Table 2 reports the regression results. The table contains four columns: in the first, we report the estimates for the traditional frontier model; in the second and third, the estimates for the LCM (here, two sets of coefficients have been reported due to the assumption of two classes); in the final column, the mean parameters of the RPM model are presented.

The regression results are difficult to interpret owing to the nonlinear nature of the model. A more useful picture emerges from the analysis of average incremental costs (AIC), marginal costs (MC) and product-specific scale effects calculated by using the formulae discussed in the appendix. These results are reported in Tables 3, 4 and 5.

Table 2. Regression results.

	Frontier	LC model (group 1)	LC model (group 2)	RPM model
Constant	−55.609	−83.025	−5.825	−19.623[a]
	(4.486)	(16.763)	(.732)	(1.196)
Bachelor	.072	.065	.070	.055[a]
	(.004)	(.014)	(.001)	(.001)
Postgraduate	−.004	−.004	.017	.001[a]
	(.008)	(.021)	(.002)	(.002)
Research	2.379	3.418	2.240	5.303[a]
	(.095)	(.149)	(.056)	(.014)
$Bachelor^2$	−.011	−.028	−.077	.008
	(.009)	(.028)	(.005)	(.002)
$Postgraduate^2$	.136	.107	.007	.033
	(.020)	(.063)	(.009)	(.007)
$Research^2$	−10.248	−21.037	−29.393	8.663
	(1.003)	(2.065)	(1.490)	(.252)
Bachelor × postgraduate	−.074	−.083	.221	.090
	(.027)	(.051)	(.016)	(.005)
Research × postgraduate	3.374	3.316	3.586	.977
	(.318)	(.570)	(.227)	(.068)
Bachelor × research	−.607	−.649	−.573	−5.302
	(.138)	(.255)	(.220)	(.037)
Log-likelihood	−18,294.91	−15,176.03	−15,176.03	−15,745.10

[a] mean of random parameters; standard errors in parentheses.

Table 3. AIC, MC and product-specific scale effects – traditional frontier model.

AIC	Bachelor	Postgraduate	Research
1	64,474	7742	2.46
0.75	49,819	3587	1.83
1.25	78,152	13,378	3.11

MC	Bachelor	Postgraduate	Research
1	63,168	15,545	2.43
0.75	49,085	7976	1.81
1.25	76,112	25,569	3.05

Scale	Bachelor	Postgraduate	Research
1	1.021	0.498	1.016
0.75	1.015	0.450	1.012
1.25	1.027	0.523	1.020

Table 4. AIC, MC and product-specific scale effects – latent class model.

Group 1

AIC	Bachelor	Postgraduate	Research
1	40,346	33,144	3.44
0.75	35,035	17,830	2.58
1.25	42,473	53,143	4.31

MC	Bachelor	Postgraduate	Research
1	34,040	47,170	3.17
0.75	31,488	25,720	2.43
1.25	32,620	75,060	3.89

Scale	Bachelor	Postgraduate	Research
1	1.185	0.703	1.084
0.75	1.113	0.693	1.062
1.25	1.302	0.708	1.107

Group 2

AIC	Bachelor	Postgraduate	Research
1	70,790	40,035	2.29
0.75	52,960	25,731	1.71
1.25	88,708	57,203	2.87

MC	Bachelor	Postgraduate	Research
1	64,250	40,289	2.26
0.75	49,282	25,874	1.69
1.25	78,489	57,599	2.82

Scale	Bachelor	Postgraduate	Research
1	1.102	0.994	1.014
0.75	1.075	0.994	1.011
1.25	1.130	0.993	1.018

The calculations have been made for an average institution – that is considering the means of outputs reported in Table 1 as the hypothetical output vector – and for institutions that produce 75% and 125% mean output levels.

Several features of the results are worth commenting upon. First, the average incremental cost estimates for postgraduates are estimated to be lower than those for undergraduates. This may reflect in part the long duration of bachelors programs in comparison with masters programs, and also the fact that the cost attached to the production of doctoral graduates is mitigated by the fact that students at this level often contribute to a university by providing teaching or research assistance.

The costs attached to postgraduate education vary markedly with the estimation technology. They are lowest (implausibly low, we think) in the stochastic frontier model, which does not allow for any parameter variation across institutions. The

Table 5. AIC, MC and product specific scale effects – random parameter model.

AIC	Bachelor	Postgraduate	Research
1	42,792	17,504	4.77
0.75	34,552	10,126	3.68
1.25	49,395	26,884	5.80

MC	Bachelor	Postgraduate	Research
1	43,757	19,425	4.81
0.75	35,095	11,206	3.70
1.25	50,902	29,884	5.85

Scale	Bachelor	Postgraduate	Research
1	0.978	0.901	0.993
0.75	0.985	0.904	0.995
1.25	0.970	0.900	0.992

sensitivity of average incremental costs of postgraduate provision with respect to estimation method is likely due to the fact that many institutions produce few or no postgraduates, and these institutions bias the coefficients on postgraduates in the fully parametric specification. This observation reinforces our argument that heterogeneity across institutions needs to be accommodated in the modeling procedure.

Further results of interest include the findings that there are scale economies for undergraduate teaching (but not in the RPM specification) while there are diseconomies of scale associated with postgraduate provision. Economies of scale associated with research activity, meanwhile, are virtually exhausted. A note of caution is needed here, however, in that all the calculations have been made considering either an 'average' institution (which is a college producing mean values of the output vector) or an institution producing some multiple (0.75 or 1.25) of *all* outputs produced by the 'average'. Such an institution does not exist, because the reality is a sector very differentiated where colleges tend to be more specialized towards a certain output. This being the case, our results must be interpreted as illustrating the scale effects for an average institution, while recognizing that differences from the average are the norm.

5.2. *Efficiency scores*

A further interesting result concerns the estimation of efficiency scores. The traditional frontier model postulates the existence of one efficiency frontier, and the efficiency of each college is computed as the distance from it. The latent class specification employed here considers the presence of two frontiers, one for each group of institutions. We expect in this case that the average efficiency of the sector should be higher than is the case where a single cost function is assumed to apply to all institutions, since the reference frontier is, under the latent class method, calculated with respect to similar competitors. Finally, the RPM assumption is that colleges must be compared with their own (potential) cost performance, and then the efficiency mean should be even higher than in the latent class case. In Figures 3, 4 and 5 the frequency
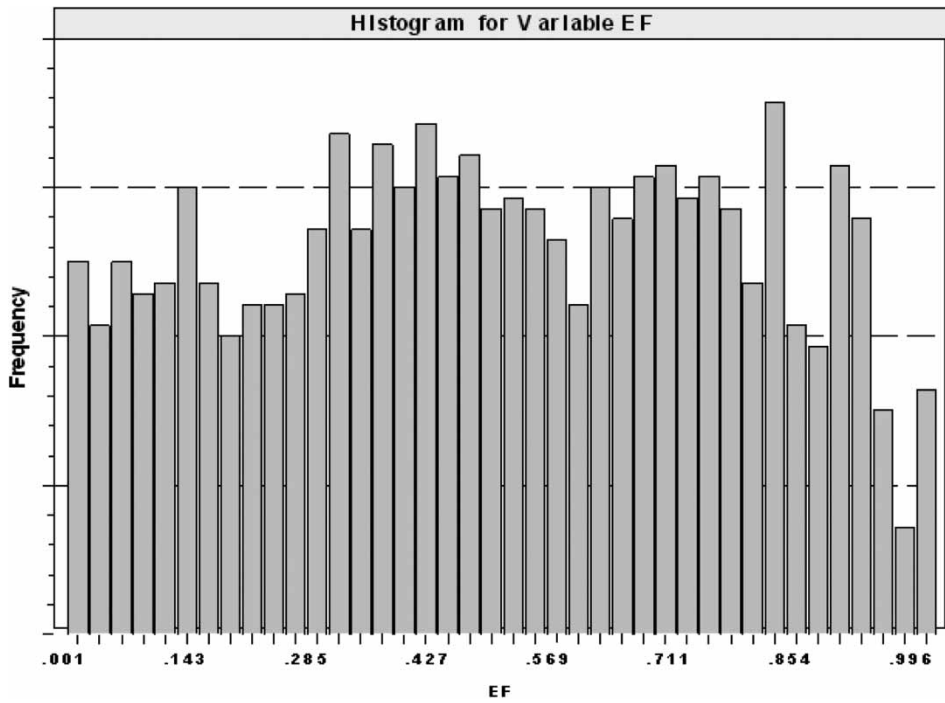
Figure 3.   Distribution of the efficiency scores – traditional frontier model.
Notes: The efficiency scores refer to all the three years (each university is included three times in the figure).
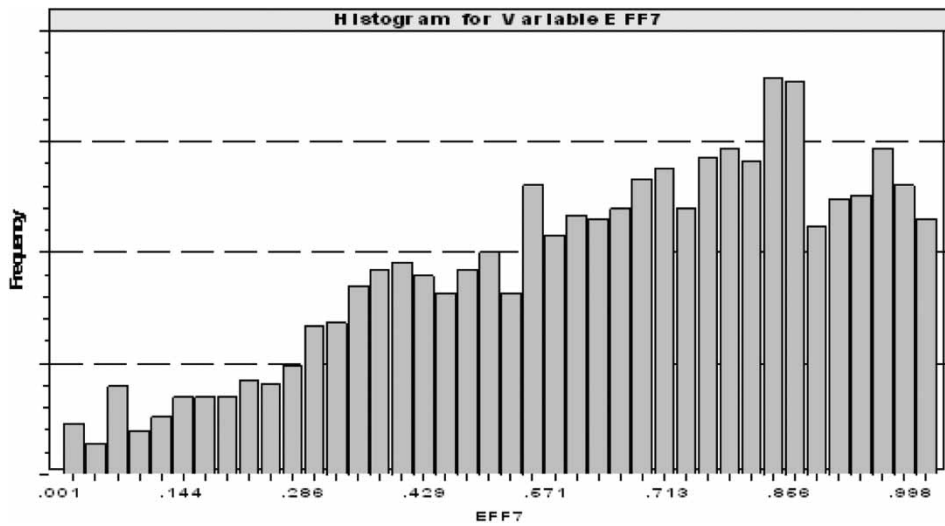

Figure 4.   Distribution of the efficiency scores –latent class frontier model.
Notes: The efficiency scores refer to all the three years (each university is included three times in the figure).
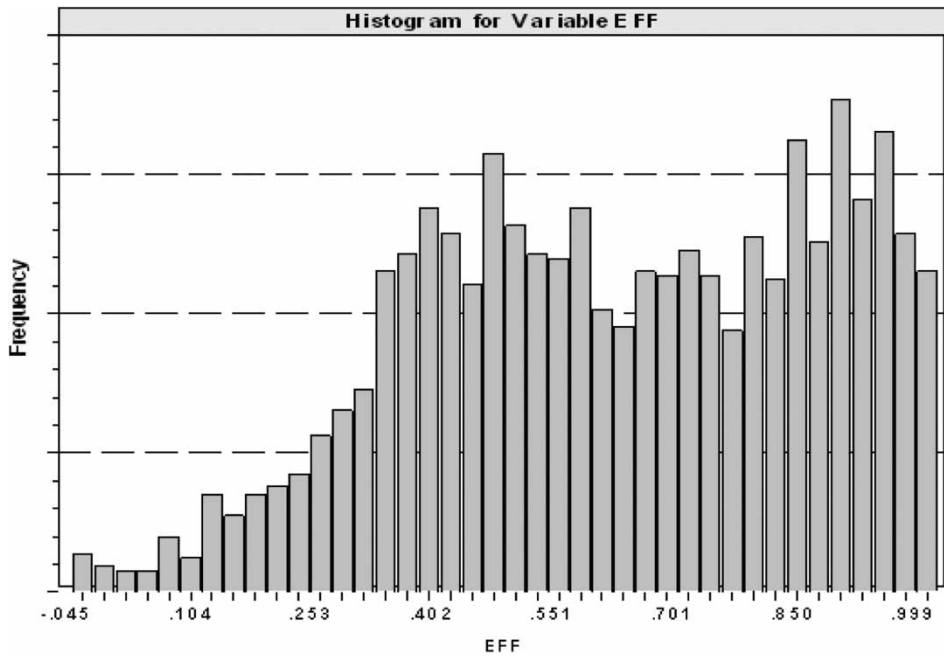
Figure 5.   Distribution of the efficiency scores – random parameter frontier model.
Notes: The efficiency scores refer to all the three years (each university is included three times in the figure).

distribution of efficiency scores are reported for each model, and they conform to our expectations.

The distribution of efficiency scores is particularly wide in the first case (the traditional frontier model), suggesting that when considering a unique frontier many colleges result are deemed to be highly inefficient. The concentration of scores below 0.5 confirms this. Moreover, the number of highly efficient institutions is very limited. The picture dramatically changes when we turn to the latent class specification. Allowing each college to compare its efficiency with the frontier *for its own group* markedly improves the overall measured performance of institutions. Indeed, in this case efficiencies are concentrated between 0.6 and 0.9, suggesting a good efficiency level of the market. The latent class specification accounts for a structural difference between two subsectors of HE market. Looking at mean output in the two classes, the difference may be primarily due to different scales of operation. Table 6 helps in explaining this point: large universities (with outflows of more than 2000 undergraduates and 1000 postgraduates, and with research income in excess of $100m) predominate in one group (group 1), while the remaining universities are concentrated in the other

Table 6.   Mean levels of output, LC model.

	Bachelor	Postgraduate	Research ($million)
Group 1	2234	1306	127
Group 2	848	356	10.9

group. Thus, estimated cost structures are not radically different across institutions within each group. What is different is the scale of operation, and it makes sense to compare each institution's efficiency with its counterparts in the submarket.

The last case to be analyzed is that of the RPM specification. Here the distribution indicates a concentration of efficiencies above 0.8, though there remain significant numbers of institutions with lower levels of measured efficiency. It is worth noting here that the comparison of each college's cost efficiency is not conducted with its competitors in a submarket, but within the whole market.

5.3. *HE subsectors, rankings, and efficiency scores*

It is interesting to analyze whether the efficiency scores reflect subsectors of the HE market. For this purpose, we consider four different types of colleges: (1) private institutions with medical schools; (2) private institutions without medical schools; (3) public institutions with medical schools; and (4) public institutions without medical schools. The aim of such categorization is to investigate whether different institutional characteristics have an impact on the average efficiency. Table 7 reports, in the first row, some descriptive statistics for the entire sample, while the subsequent rows illustrate corresponding descriptive statistics for the subgroups. This analysis focuses only on efficiency scores calculated from the RPM, and these scores refer only to year 2005.

The picture that emerges is that the presence of a medical school is associated with higher efficiency scores. Meanwhile, private ownership of the college is related to lower scores. The differences emerging from this analysis confirm the presence of heterogeneity in performance across subsectors of the HE market.

G. Johnes et al. (2005), when studying the cost structure and efficiency profile of the HE sector in England, pointed out that measured efficiencies were highest in the top five universities and lowest in the colleges of HE. Analogously, here we try to find out whether an analogous pattern exists in the US HE market. We followed two strategies for this purpose: (1) a look at the efficiency scores for 'high intensive research' universities – as defined by the Carnegie classification; and (2) an analysis of efficiency scores for the best universities as classified by the US News and World Report ranking (2009 edition). The last row of Table 7 gives evidence that supports our hypothesis that research universities are relatively efficient. These universities are typically characterized not only by higher average efficiency, but also by a very small standard deviation – in other words, performance (efficiency) is very similar across institutions that are members of this elite group. The distribution of efficiency scores, presented in Figure 6, gives intuitive

Table 7. Efficiency scores (random parameters model) – descriptive statistics.

College type	Mean	Standard deviation	Min	Max	#
Overall	0.5776	0.2556	0.0013	0.9988	898
Medic private	0.7728	0.2505	0.0330	0.9872	60
Medic public	0.8384	0.1839	0.0043	0.9988	93
No medic public	0.5997	0.2151	0.0013	0.9891	425
No medic private	0.4359	0.2304	0.0022	0.9796	320
Research universities	0.9132	0.0480	0.7213	0.9988	95

Notes: All the efficiency scores refer to year 2005. Research universities included in the last row are those classified as 'very high research' universities by the Carnegie classification.
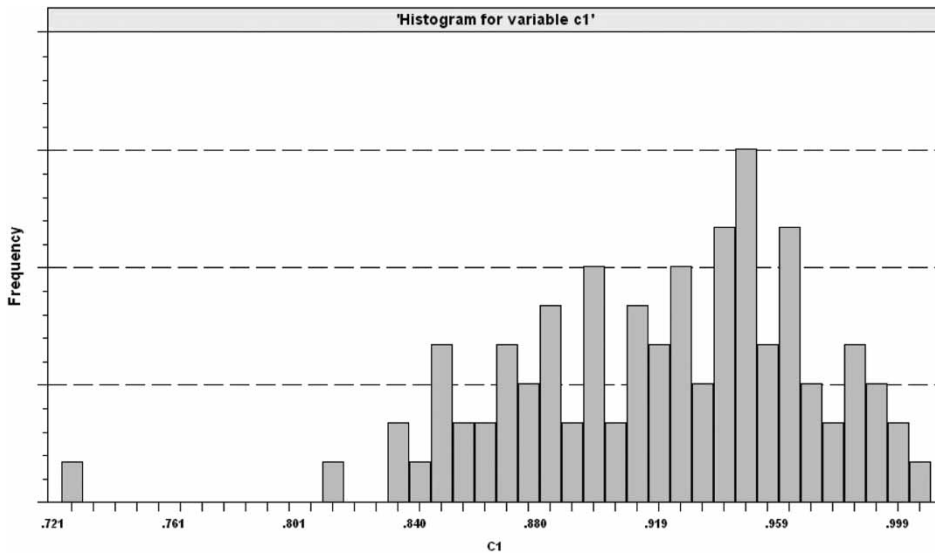
Figure 6. Distribution of efficiency scores for 'research intensive' universities – random parameter frontier model.
Notes: All the efficiency scores refer to year 2003. Research universities are classified as 'very high research' universities by the Carnegie classification.

evidence of the structural difference that characterizes this group of universities with respect to the general distributions of scores (see Figures 3–5).

Table 8 reports the efficiency scores for the top 20 universities listed in the US News and World Report ranking (2009 edition). Here, the efficiency scores are those for all the three years considered (2003–5). Again, what emerges is a very high efficiency score, well above the average of the sector.

Several factors could explain this finding. One possible interpretation is that efficiency, productivity and research and teaching excellence are highly correlated measures of performance. Costs in these, as in other, institutions are determined in part by research activity, and the weight associated with such activity is captured by the estimated parameters on variables that measure research. If these parameters are upwardly biased, then so will be the estimated efficiency scores for research-intensive institutions. While we have no reason to suppose that the parameters are, in fact, biased, the reader should bear in mind that frontier models of the kind estimated here, in attempting to distinguish statistically between the drivers of costs and the prevalence of inefficiency, are asking a lot of the data.

5.4. *Scale and scope economies*

Ray returns to scale, derived using the RPM, are reported, for the average institution, in Table 9.

The general picture that emerges is that there are no more scale economies to be exploited. There are, however, unexploited economies of scope – especially at the lower levels of output. The interpretation of this story is that the 'optimal' size of institutions is already almost reached, but gains could be realized by increasing the involvement of more institutions in both postgraduate education and research.

Table 8. The efficiency of the best colleges – as classified by the US News and World Report 2009 Edition.

College	2003	2004	2005	Mean
Harvard University	0.9933	0.9722	0.9024	0.9560
Princeton University	0.9180	0.9416	0.9340	0.9312
Yale University	0.9846	0.9652	0.9446	0.9648
MIT	0.9959	0.9774	0.9413	0.9715
Stanford University	0.9590	0.8901	0.9571	0.9354
California Institute of Technology	0.8571	0.8545	0.8642	0.8586
University of Pennsylvania	0.9979	0.9821	0.9512	0.9771
Columbia University	0.9961	0.9067	0.9109	0.9379
Duke University	0.9710	0.9726	0.9833	0.9756
University of Chicago	0.9225	0.8474	0.8747	0.8815
Dartmouth College	0.9414	0.9561	0.9224	0.9400
Northwestern University	0.9900	0.9584	0.9585	0.9689
Washington University St. Louis	0.9594	0.9600	0.9673	0.9622
Cornell University	0.9879	0.9870	0.9489	0.9746
John Hopkins University	0.9531	0.9970	0.9613	0.9705
Brown University	0.9340	0.9130	0.8417	0.8962
Rice University	0.8935	0.8742	0.8307	0.8661
Emory University	0.9103	0.9056	0.9514	0.9225
University of Notre Dame	0.9297	0.9187	0.9060	0.9181
Vanderbilt University	0.9883	0.9472	0.9220	0.9525
University of California Berkeley	0.9733	0.9811	0.9264	0.9603

Notes: The efficiency scores have been derived using the random parameter modeling.

Table 9. Ray economies of scale and scope economies, random parameter stochastic frontier model.

Ray economies of scale	
1	0.9642
0.75	0.9736
1.25	0.9545
Scope economies	
1	0.1389
0.75	0.4688
1.25	–0.0575

Notes: The mean values considered are the means of the output vector.

6. Concluding remarks

The HE system of the USA has been the subject of much research over the years, but the present paper represents the first attempt to apply recently developed econometric methods for this country as a means of evaluating costs, returns to scale, returns to scope and efficiency in the context of a framework which allows for the heterogeneity

of institutions. It is also the first paper to explore the usefulness of latent class models in this context. Our results confirm our view that American institutions of HE are indeed heterogeneous. Those that have high profile as research institutions tend also to operate at high levels of efficiency. There is, however, a large tail of less efficient institutions.

From the perspective of cost efficiency, there is evidence to suggest that there are too many HEIs in the USA. To some extent, this may be due to geography – to cater for the needs of students who are unwilling or unable to travel far for tertiary education, there need to be widespread geographical coverage. Nonetheless, global costs could be reduced by increased concentration of provision of undergraduate education. This has clear policy implications for the delivery of HE, particularly in the context of state universities. Meanwhile, scope economies in some institutions could be improved if postgraduate provision were extended to a greater number of institutions.

It is instructive to compare these results with those obtained for the HE systems of some other countries. In Germany and Spain, studies have found evidence to suggest that product-specific economies of scale remain unexploited for all outputs of HEIs (G. Johnes and Schwarzenberger 2011; G. Johnes and Salas-Velasco 2007); in the case of Italy, meanwhile, Agasisti and Johnes (2010) have found the opposite. Perhaps the most intriguing comparison, though, is with England, where, in sharp contrast to the USA, G. Johnes and Johnes (2009) found increasing product-specific returns to scale for postgraduate education and decreasing returns to undergraduate provision. These findings suggest that there is much that can be learned from the comparison of systems across countries, and we suggest that further work on international data should be a priority for future research.

Notes

1. Multiproduct organizations have been the subject of much research since the seminal contribution of Baumol, Panzar, and Willig (1982). This topic is particularly interesting for policy and practice; such organizations enjoy different resource endowments and so specialize in the production of different outputs. The structure of costs in these organizations likely differs across firms, and this heterogeneity gives rise to some interesting issues surrounding the evaluation of overall organizational performance.
2. The academic literature also pointed out that universities could contribute to the economic growth of the territories in which they operate; see Huggins and Johnston (2009).
3. Nevertheless, given the difficulty of identifying proper measures for the 'third mission' work of this kind, estimated cost functions usually include only measures of teaching and research; indeed only a very few attempts have been made to include proxy for the 'third mission' (an exception is G. Johnes et al. 2005; J. Johnes, Johnes, and Thanassoulis 2008).
4. G. Johnes et al. (2005) also include a quality measure in their work. Collinearity problems often preclude this.
5. The definition is slightly different, albeit equivalent, for public universities on the one hand and private universities on the other. This is because accounting principles for the former follow the guidelines of the Governmental Accounting Standards Board while those for the latter follow the guidelines of the Financial Accounting Standards Board.

References

Agasisti, T., and G. Johnes. 2010. Heterogeneity and the evaluation of efficiency: the case of Italian universities. *Applied Economics* 42: 1365–75.

Aigner, D.J., C.A. Lovell, and P. Schmidt. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6: 21–37.

Baumol, W.J., J.C. Panzar, and R.D. Willig. 1982. *Contestable markets and the theory of industry structure*. San Diego: Harcourt Brace Jovanovich.

Bonaccorsi, A., and C. Daraio. 2007. *Universities and strategic knowledge creation*. Cheltenam: Edward Elgar.

Brinkman, P.T., and L.L. Leslie. 1986. Economies of scale in higher education: sixty years of research. *Review of Higher Education* 10: 1–28.

Butler, D. 2007. Academics strike back at spurious rankings. *Nature* 447: 514–15.

Cohn, E., and S.T. Cooper. 2004. Multi-product cost functions for universities: economies of scale and scope. In *International Handbook on the Economics of Education*, ed. G. Johnes and J. Johnes, 579–612. Cheltenham, UK: Edward Elgar.

Cohn, E., S.L.W. Rhine, and M.C. Santos. 1989. Institutions of higher education as multi-product firms: economies of scale and scope. *Review of Economics and Statistics* 71: 284–90.

De Fraja, G., and P. Valbonesi. 2008. The design of university system. Centre for Economic Policy Research, Discussion Paper Series, No. 7038.

De Groot, H., W.W. McMahon, and J.F. Volkwein. 1991. The cost structure of American research universities. *The Review of Economics and Statistics* 73, no. 3: 424–31.

Del Rey, E. 2001. Teaching versus research: a model of state university competition. *Journal of Urban Economics* 49: 356–73.

Dill, D.D., and M. Soo. 2005. Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education* 49, no. 4: 495–533.

EUA (European University Association). 2011. *Global university rankings and their impact*. Brussels, Belgium: EUA.

Greene, W. 2005. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126: 269–303.

Gutiérrez-Romero, R., D. Haubrich, and I. McLean. 2008. The limits of performance assessments of public bodies: external constraints in English local government. *Environment and Planning C: Government and Policy* 26, no. 4: 767–87.

Harter, J.F.R., J.A. Wade, and T.G. Watkins. 2005. An examination of costs at four-year public colleges and universities between 1989 and 1998. *The Review of Higher Education* 28, no. 3: 369–91.

Hazelkorn, E. 2011. *Rankings and the reshaping of higher education: The battle for world-class excellence*. New York, NY: Palgrave Macmillan.

Hoxby, C. 1999. How the changing market structure of US higher education explains college tuition. NBER working paper 6323.

Huggins, R., and A. Johnston. 2009. The economic and innovation contribution of universities: a regional perspective. *Environment and Planning C: Government and Policy* 27, no. 6: 1088–1106.

Izadi, H., G. Johnes, R. Oskrochi, and R. Crouchley. 2002. Stochastic frontier estimation of a CES cost function: the case of higher education in Britain. *Economics of Education Review* 21: 63–71.

Jeremic, V., M. Bulajic, M. Martic, and Z. Radojicic. 2012. A fresh approach to evaluating the academic ranking of world universities. *Scientometrics* 87, no. 3: 587–96.

Johnes, G., and J. Johnes. 1993. Measuring the research performance of UK economics departments. *Oxford Economic Papers* 45: 332–47.

Johnes, G., and J. Johnes. 2009. Higher education institutions' costs and efficiency: taking the decomposition a further step. *Economics of Education Review* 28: 107–13.

Johnes, G., J. Johnes, E. Thanassoulis, P. Lenton, and A. Emrouznejad. 2005. *An exploratory analysis of the cost structure of higher education in England*. London: Department for Education and Skills Research Report 641.

Johnes, G., and M. Salas-Velasco. 2007. The determinants of costs and efficiencies where producers are heterogeneous: the case of Spanish universities. *Economics Bulletin* 4, no. 15: 1–9.

Johnes, G., and A. Schwarzenberger. 2011. Differences in cost structure and the evaluation of efficiency: the case of German universities. *Education Economics* 19: 487–500.

Johnes, J., G. Johnes, and E. Thanassoulis. 2008. An analysis of costs in institutions of higher education in England. *Studies in Higher Education* 33, no. 5: 527–49.

Koshal, R.K., and M. Koshal. 1999. Economies of scale and scope in higher education: a case of comprehensive universities. *Education of Economics Review* 18: 269–77.

Laband, D.N., and B.F. Lentz. 2003. New estimates of economies of scale and scope in higher education. *Southern Economic Journal* 70, no. 1: 172–83.

Laband, D.N., and B.F. Lentz. 2004. Do costs differ between for-profit and not-for-profit producers of higher education? *Research in Higher Education* 45, no. 4: 429–41.

Orea, L., and S.C. Kumbhakar. 2004. Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics* 29: 169–83.

Sav, T.G. 2004. Higher education costs and scale and scope economies. *Applied Economics* 36: 607–14.

Sullivan, T.A., C. Mackie, W.F. Massy, and E. Sinha, eds. 2012. *Improving measurement of productivity in higher education*. Washington, DC: National Academy of Sciences.

Tsionas, E.G. 2002. Stochastic frontier models with random coefficients. *Journal of Applied Econometrics* 17: 127–47.

US News and World Report, Best Colleges, 2009 Edition, USA.

Appendix 1. Mathematical definitions of AIC, MC, scale and scope economies

According to Baumol, Panzar, and Willig (1982), we define the average incremental cost (AIC) associated with product k as:

$$AIC(y_k) = \left[C(y_K) - C(y_{K-k})\right]/y_k, \tag{A1}$$

where $C(y_K)$ is the cost of producing the observed output vector, and $C(y_{K-k})$ is the cost associated with producing the observed values of all outputs other than the kth output, and where the output of type k is zero. Defining $C_k(y)$ as the marginal cost of the kth output, we can then define product-specific returns to scale associated with the kth output as

$$S_k(y) = AIC(y_k)/C_k(y). \tag{A2}$$

This definition is therefore analogous to the ratio of average to marginal costs that is often used as a measure of scale economies in single product contexts. A value of $S_k(y)$ that exceeds unity reflects product-specific returns to scale that are increasing, and vice versa.

Ray returns to scale (S_R), which capture scale effects associated with a simultaneous and proportional change in all outputs, may be calculated as

$$S_R = \frac{C(y)}{\sum_k y_k C_k(y)}. \tag{A3}$$

A value of S_R exceeding unity indicates that a simultaneous proportional increase in the production of all output types results in economies of scale, while a value less than one indicates decreasing returns to scale.

Global economies of scope are calculated using the formula

$$S_G = \left[\sum_k C(y_k) - C(y)\right]/C(y), \tag{A4}$$

where $C(y_k)$ is the cost of producing only the observed value of the kth output, with zero output of all other types. This formula therefore compares, in the numerator, the cost of producing the observed output vector in a single institution with that of producing the same output in several different, single-product, institutions. If S_G is positive, then it is cheaper to produce jointly than not, and so economies of scope are said to exist. Conversely, $S_G < 0$ implies diseconomies of scope.